CASUALTY ACTUARIAL SOCIETY FORUM

Summer 1999

Including the Dynamic Financial Analysis Discussion Papers

CASUALTY ACTUARIAL SOCIETY
ORGANIZED 1914

Printed for the Society by
Colortone Press
Landover, Maryland

NOTICE

The Casualty Actuarial Society is not responsible for statements or opinions expressed in the papers in this publication. These papers have not been reviewed by the CAS Committee on Review of Papers.

The Casualty Actuarial Society *Forum*

Summer 1999 Edition

Including the Dynamic Financial Analysis Discussion Papers

To CAS Members:

This is the Summer 1999 Edition of the Casualty Actuarial Society *Forum*. It contains eight Dynamic Financial Analysis Discussion Papers and one additional paper.

The Casualty Actuarial Society *Forum* is a non-refereed journal printed by the Casualty Actuarial Society. The viewpoints published herein do not necessarily reflect those of the Casualty Actuarial Society.

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,

Janet G. Lockwood

Janet G. Lockwood, CAS *Forum* Chairperson

The 1999 CAS DFA Discussion Papers
Presented at the
1999 Seminar on Dynamic Financial Analysis
July 19-20, 1999
Drake Hotel
Chicago, Illinois

The Summer 1999 Edition of the CAS *Forum* is a cooperative effort of the CAS *Forum* Committee and the CAS Committee on Dynamic Financial Analysis. The call for papers was initiated in 1998 by the Valuation and Financial Analysis Committee.

The Committee on Dynamic Financial Analysis is pleased to present eight papers prepared in response to its 1999 DFA Call for Discussion Papers, entitled "Parameterizing Models." These papers include papers that will be discussed by the authors at the 1999 CAS Seminar on Dynamic Financial Analysis, July 19-20, in Chicago, Illinois.

Table of Contents

Parameterizing Interest Rate Models

Kevin C. Ahlgrim, ASA,
Stephen P. D'Arcy, FCAS, MAAA, and
Richard W. Gorvett, FCAS, MAAA

Biographical Information

Kevin C. Ahlgrim is a graduate student in the Department of Finance at the University of Illinois at Urbana-Champaign. His research interests include fixed income derivatives and financial risk management of life insurers. Kevin has taught a course on financial markets and has assisted in developing a course for delivery over the internet.

Kevin received his B.S. in Actuarial Science from the University of Illinois. He is an Associate of the Society of Actuaries and a member of the American Academy of Actuaries. He has worked as a consulting actuary at Aon in Chicago and as a pricing actuary at CIGNA in Hartford.

Stephen P. D'Arcy is a Professor of Finance at the University of Illinois at Urbana-Champaign. He is a Fellow of the Casualty Actuarial Society, a member of the American Academy of Actuaries, and is currently President of the American Risk and Insurance Association. He received his B.A. in Applied Mathematics from Harvard College and his Ph.D. in Finance from the University of Illinois.

Prior to his academic career, Steve was the Actuary at CUMIS Insurance Society. His research interests include dynamic financial analysis, financial pricing models for property-liability insurers, catastrophe insurance futures, pension funding and insurance regulation.

Richard W. Gorvett is currently an Assistant Professor at The College of Insurance in New York City. In August, 1999, he will become an Assistant Professor in the Actuarial Science Program at the University of Illinois at Urbana-Champaign. He is a Fellow of the Casualty Actuarial Society, a Member of the American Academy of Actuaries, and an Associate in Risk Management. He received a B.S. in Mathematics from the University of Illinois at Chicago, an MBA in Finance and Econometrics from the University of Chicago, and a Ph.D. in Finance from the University of Illinois at Urbana-Champaign, where his dissertation was entitled "Dynamic Financial Analysis of Property-Liability Insurance Companies." He has co-authored a paper entitled "A Comparison of Property-Liability Insurance Financial Pricing Models" for the *Proceedings of the Casualty Actuarial Society* and two papers on Dynamic Financial Analysis for Call Paper Programs. He is continuing research into the application of financial theory to insurance. Rick has taught courses in property-liability insurance, corporate risk management, investments, corporate finance, and financial risk management. Prior to entering academia, he worked as an actuary and consultant with several firms in the Chicago area.

CAS Committee on Valuation & Financial Analysis and
Dynamic Financial Analysis Task Force on Variables
Call Paper Program

PARAMETERIZING INTEREST RATE MODELS

ABSTRACT

Actuaries are now being called upon to incorporate interest rate models in a variety of applications, including dynamic financial analysis (DFA), ratemaking, and valuation. Although there are many articles and texts on interest rate models, most of these presume an understanding of financial terminology and mathematical techniques that makes it difficult to begin to learn this material. This paper provides an overview, at a level aimed at actuaries, of some common interest rate models used by financial economists. The purpose of this paper is to explain the basics of interest rate modeling by demonstrating the different models both graphically and empirically, and by showing how changing the various model parameters affects the results. Several of the more popular interest rate models are simulated, and the results are compared with historical interest rate movements.

Introduction

The volatility of interest rates has become an important feature of the modern financial environment. Changes in interest rates can impact the way in which companies compete and can even impact the ultimate survival of the firm. Financial intermediaries, such as banks and insurance companies, may be especially exposed to interest rate fluctuations because both their assets and liabilities are correlated with interest rate movements. Mismatches of interest rate sensitivities (or durations) of assets and liabilities can have a magnified effect on surplus. A popular example of the potential vulnerability of financial intermediaries is based on the experience of the savings and loan industry in the 1980s. Rapidly increasing interest rates quickly turned profits into billions of dollars of losses and numerous insolvencies. The assets of S&Ls were primarily long-term, fixed-rate mortgages; their liabilities were mostly short-term demand deposits. When the interest rates paid on those short-term deposits increased, the normal differential between the interest rate they were receiving on their assets and that which they were paying on their liabilities disappeared or even reversed. Given such potential effects of interest rate volatility, it has become important to develop models of interest rate changes so that risk management tools can be used to insulate the firm from financial disaster.

Traditionally, insurance companies have not incorporated interest rate models into the product development and pricing processes. Pricing actuaries typically used "conservative," fixed interest rates when developing products. By crediting policyholders with a low interest rate, or ignoring investment income when setting property-liability insurance rates, insurers had some assurance that they could ultimately earn the assumed rate of return used in pricing. Any excess interest earnings served as a cushion to protect surplus against adverse experience, as well as being a source of insurer profits.

The assumption of fixed interest rates was an acceptable practice during periods when interest rates were low and relatively stable. In fact, such an environment existed in the U.S. into the 1970s. The fixed interest rate assumption used by most insurers seemed innocuous.

However, in 1979, the U.S. Federal Reserve altered its policy from one that targeted interest rates to a policy that now targets inflation via the money supply. As a result, interest rates became significantly more volatile. During the transition of the early 1980s, interest rates spiked upward to unprecedented levels. It was clear that the interest rate environment had shifted dramatically.

The change in the Fed policy affected insurers in several ways. First, the underlying value of insurance products changed due to the change in interest rate volatility. Insurance products typically include embedded options that give specific rights to policyholder and, in some cases, to the insurer. An example of these options is the right to renew the policy on terms set at the beginning of the coverage period. The value of these embedded options is highly sensitive to the underlying interest rate assumption and, more importantly, to the volatility of future interest rates. Insurers that used a fixed interest rate assumption ignored the interest-sensitive option values in their policies.

It has long been recognized that life insurers are exposed to interest rate risk. The life insurance industry experienced heavy disintermediation when interest rates increased in the 1980s. Before the rapid increase in rates, life insurers believed that high interest rate scenarios were in their favor because they implied additional income. However, they failed to understand the risks in their liabilities. The policy loan feature of ordinary life insurance policies capped the interest rate that could be charged to the policyholder. Once interest rates exceeded that cap, policyholders were able to borrow at the policy loan rate, and then turn around and invest the proceeds at higher yields. The result was an outflow of cash from the life insurance industry that caused many insurers to sell bonds at depressed prices due to the high yield environment.

It is also becoming evident that property-liability insurers are exposed to interest rate risk on both sides of the balance sheet. Fixed income assets of property-liability insurers have the same exposure to interest rate risk that life insurer assets have, with market values declining as interest rates increase. On the other hand, the liabilities of property-liability insurers are not fixed values. Since inflation is correlated with interest rates, and future claim payments on loss reserves will increase with inflation, the statutory values of liabilities will tend to increase as interest rates increase. Thus, an increase in interest rates leads to a decline in asset value and an increase in the value of liabilities, creating a magnified effect on the surplus of property-liability insurers.

In most DFA models for property-liability insurers, interest rates are the driving factor in the model, affecting investment income, loss severity, asset returns, and target underwriting profit margins (see, for example, D'Arcy, Gorvett, et al, 1997 and 1998). DFA models are being used for analyzing insurer solvency, in valuing insurers in mergers and acquisitions, and as a business planning tool. The results from DFA applications are heavily dependent upon the particular interest rate model used, as well as the parameters chosen for the models.

These examples help illustrate how critical the underlying interest rate assumption is to the evaluation of insurance company assets and liabilities. Insurers must incorporate the new interest rate paradigm into their pricing and asset/liability management (ALM) processes by using assumptions that reflect the stochastic nature of interest rates. Fortunately, within the field of finance, extensive effort has been devoted to developing stochastic interest rate models.

Financial researchers have long been concerned with the dynamics of interest rates. Models have been formulated using two approaches: (1) a general equilibrium framework, where interest rate changes are derived from economic agents who maximize expected utility; and (2) the no-arbitrage approach, which assumes that financial markets have no arbitrage opportunities. Examples of the general equilibrium approach include the models of Vasicek (1977), Dothan (1978), Cox, Ingersoll, and Ross (CIR) (1985), Brennan and Schwartz (1979) and Longstaff and Schwartz (1992). Two models based on arbitrage arguments are Ho and Lee (1986) and Heath, Jarrow, and Morton (HJM) (1992).

The choice of interest rate model is not a trivial decision. The form of the model used in the pricing or ALM process depends on the characteristics of the insurance products being reviewed. Choosing a model is always a tradeoff between perfectly describing the actual interest rate process and having a tractable model that can be used to value a variety of financial instruments. One consideration in selecting an interest rate assumption is to compare modeled prices of financial assets with market prices, if a market exists. When using a model for a specific application, one should compare market prices of assets that are similar in terms of interest rate sensitivity. Another consideration is choosing which interest rate to model. The spot rate is today's interest rate for a specific maturity. A forward rate is an interest rate that is applicable to future periods[1]. After deciding on which interest rates to model, one must determine how many parameters to include. Using more parameters obviously increases the complexity of a model, so one must consider whether the added complexity yields sufficient benefits. Finally, choosing the values of the parameters in an interest rate model can be the most important, as well as the most challenging, factor in implementation.

This paper aims to illustrate how various models operate and to show how well the models fit historical data. Through descriptions and illustrations of the models, it is hoped that this paper will increase the comfort level of casualty actuaries with these new tools and encourage them to begin to apply them in pricing and asset/liability management functions.

The estimates used in this presentation are based on previous work in the area. Chan, Karolyi, Longstaff, and Sanders (1992) (hereafter CKLS), empirically estimate and compare several popular interest rate processes used in the literature. Their most important finding is that the interest rate volatility is sensitive to the level of the interest rate. Also, Amin and Morton (1994) estimate parameters for six forms of the HJM model. They find that two-parameter models fit market price data better, but that the resulting estimates are less stable.

---

[1] Example: The expected forward rate from year one to year two can be implied from the current spot rates based on the following formula: $(1+i_1)(1+f) = (1+i_2)^2$, where f is the forward rate and i_t is the t-year spot rate. If the one-year spot rate is 3% and the two-year spot rate is 4% (expressed as an annual rate), this implies that the forward rate is $1.04^2/1.03 - 1 = 5.01\%$

Introduction to Notation

The various interest rate models will be presented here in the mathematics of continuous time. The finance literature is based on continuous time because functions of continuous time processes (e.g., options that are dependent on the interest rate) have desirable features including continuity and differentiability. This allows many of these functions to have closed form solutions without the need for numerical procedures. The mathematics behind discrete time processes is not always as elegant. Later in this paper, we discuss how to translate the continuous time processes into discrete time for use in insurance applications (see the "Simulations" section).

The interest rate models that are presented in this paper are either models of the short-term rate or the forward rate. The short-term rate (also called the short rate or instantaneous rate) is the (annualized) rate of return expected over the next instant. For example, the return (r) over the next instant (dt) on an initial wealth level (W) earns

$$dW = rWdt$$

The time t prices of bonds (P) that pay \$1 at time T are determined by expectations of investors regarding the evolution of the short rate from time t until maturity.

$$P(t,T) = E\left[\exp\left(-\int_t^T r(u)du\right)\right]$$

This formula shows that the price of a bond is simply the discounted value *over every instant* from time t until maturity at T. Instead of modeling interest rates explicitly, many financial economists (e.g., Vasicek (1977), Dothan (1978), Cox-Ingersoll-Ross (1985) and Brennan-Schwartz (1979)) model the changes in the short-term rate using the following generic, stochastic form:

$$dr_t = a(r_t,t)dt + \sigma(r_t,t)dB_t$$

To understand the changes in interest rates, consider individually the two terms on the right-hand side of the equation. The first term represents the predictable, deterministic portion of changes in the interest rate. Thus, $a(r_t,t)$ is the expected change in the short-term rate and is called the instantaneous drift. The second term represents the uncertainty in interest rate changes; B_t represents a standard Brownian motion so that dB_t is essentially a random draw from the standard normal distribution, which is then scaled by the magnitude $\sigma(r_t,t)$. The second term in the stochastic equation thus denotes the volatility of interest rate changes. Most interest rate models begin with this form but differ in their specifications of the terms $a(r_t,t)$ and $\sigma(r_t,t)$.

Instead of modeling the short-term rate, other authors (Ho-Lee (1986) and Heath-Jarrow-Morton (1992)) use a process for forward rates. The instantaneous forward rate (f) represents the interest rate available now for an investment to be made at a future time. It is implicitly defined by a difference in bond prices, which reflects the expected instantaneous interest rate T-t periods in the future:

$$\frac{P(t,T+dt)}{P(t,T)} = \exp(-f(t,T)dt)$$

By rearranging and integrating, we can obtain the bond price in terms of the existing instantaneous forward rates:

$$P(t,T) = \exp\left(-\int_t^T f(t,u)du\right)$$

One can interpret this formula in the same manner as in footnote 1. We are "constructing" a spot rate which applies from time t to T by including consecutive instantaneous forward rates. Ho-Lee and HJM model the entire term structure by using a process for forward rates of all maturities:

$$df(t,T) = a(t,T,f(t,T))dt + \sigma(t,T,f(t,T))dB_t$$

Here, the terms $\alpha(t,T,f(t,T))$ and $\sigma(t,T,f(t,T))$ are the drift and the volatility, respectively, of the forward rate and are analogous to the short-rate drift and volatility discussed above.

Having defined the notation and general stochastic process used to model interest rates, we turn to describing desirable features of an interest rate model and then present alternative models that have been used in the literature.

Characteristics of Interest Rate Movements

Before presenting the interest rate models, we discuss some general features of interest rate movements. Our attempt is to provide some intuitive form for an interest rate model.

1. The volatility of yields at different maturities varies. In particular, long-term rates do not vary as much as shorter term rates.
2. Interest rates are mean-reverting. Interest rate increases tend to be followed by rate decreases; conversely, when rates drop, they tend to be followed by rate increases.
3. Rates of different maturities are positively correlated. Rates for maturities that are closer together have higher correlations than maturities that are farther apart.
4. Interest rates should not be allowed to become negative.
5. Based on the results reported in CKLS, the volatility of interest rates should be proportional to the level of the rate.

No known model captures all of the features mentioned above. Therefore, one of the first steps in choosing an interest rate model is to understand which of these characteristics are important based on the use of the model. One should resist the urge to rank models based on the number of listed conditions that are satisfied. Instead, it is imperative that the modeler understand the limitations of alternative models and their impact on the desired application.

Equilibrium vs. Arbitrage

The first distinction of interest rate models is between those that are derived from equilibrium models of the economy, and those that are based on arbitrage arguments. Equilibrium interest rate models are based on the assumption that bond prices, and yields, are determined by the market's assessment of the evolution of the short-term interest rate. In the models discussed here, the short rate is assumed to follow a diffusion (a continuous time stochastic) process. The general form for these models is described in terms of changes in the short rate, as follows:

$$dr_t = \kappa(\theta - r_t)dt + \sigma r_t^{\gamma} dB_t$$

r_t = current level of the instantaneous rate
κ = speed of the mean reversion
θ = rate to which the short rate reverts
σ = volatility of the short rate
γ = proportional conditional volatility exponent
B_t = standard Brownian motion

The first important feature of this type of model is mean reversion of the short-term rate. This feature is appealing since it presumes that when rates become very high or very low, they will tend to revert to "normal" levels. The speed of reversion is determined by the parameter κ. This parameter ultimately affects the shape of the yield curve. If κ is high, the yield curve quickly trends toward the long-run yield rate θ. If κ is low, the yield curve slowly trends toward θ. (See Figure 1 versus Figure 2.)

The difference between the Vasicek, CIR, and Dothan models (see below) primarily revolves around the parameter γ (the exponent). Vasicek assumes it to be 0, CIR assumes it to be 0.5, and Dothan assumes it to be 1.0. The basic question distinguishing the models is whether the conditional volatility of changes in interest rates is proportional to the level of the rate. This subsequently determines the parameter γ. CKLS (1992) have provided empirical estimates of the exponent. Their main finding is that the conditional volatility of interest rates is significantly related to the level of the short rate. In fact, their estimate of γ is around 1.5. Although their work has been the subject of some criticism due to their estimation period, it nonetheless illustrates the strong relationship between the level of interest rates and volatility. Throughout most periods, γ has been estimated between 0.5 and 1.0 (Phoa 1997). The exponent of individual models will be discussed more fully when we look at the individual models in the next section.

Equilibrium models are criticized because they do not fit the existing term structure. Although parameters can be chosen to minimize errors from today's yield curve, the fit will not be perfect. Whereas this is a valid criticism for models being used to value financial assets for trading purposes, it may not be a problem when the models are being used for long-term financial modeling, such as in DFA.

Arbitrage-free models take the entire yield curve as given and model the dynamics of the entire curve. The only constraint of such an approach is that yield curve movements do not produce any arbitrage opportunities. Heath, Jarrow, Morton (1992, hereafter HJM) generalize the arbitrage-free framework by modeling the forward rates derived from the current yield curve. The continuous time model of Ho and Lee (1986, hereafter Ho-Lee) is the simplest case of the HJM framework.

In the next section we look at several of the popular interest rate models used today.

The Vasicek Model

Vasicek formulates the interest rate model in terms of changes in the short-term (or instantaneous) interest rate:

$$dr_t = \kappa(\theta - r_t)dt + \sigma dB_t$$

The price of a bond, $P(t,T)$, is then dependent on the expected path of future interest rates. Vasicek shows that bond prices have the following form:

$$P(t,T) = A(t,T)e^{-r(t)B(t,T)}$$

where $A(t,T)$ and $B(t,T)$ are functions of only κ, θ, and σ and independent of the current spot rate, $r(t)$. Bond yields, $R(t,T)$ are then related to prices by:

$$P(t,T) = e^{-R(t,T)(T-t)}$$

These two equations determine the entire term structure of interest rates. Since bond prices and yields have closed-form solutions, the Vasicek model is very easy to implement in practice, with no need for complicated simulation techniques. Also, there are closed form solutions for certain interest rate-dependent claims such as options.

The Vasicek model assumes that (absolute) changes in the interest rate are normally distributed, due to the inclusion of the Wiener process. From the normality assumption it follows that bond prices are lognormally distributed. One weakness of the model is that normality in interest rate changes results in a (small) positive probability of negative interest rates [2].

Another feature of the Vasicek model is that all bond prices are related to the same factor, the instantaneous interest rate. Consequently, all bond price movements are derived from movements in the same factor. This implies that all bond prices are perfectly correlated. Thus, another shortcoming of the Vasicek model is that the dynamics of the term structure are severely limited.

Note that, from the general case above, the Vasicek model assumes $\gamma=0$. The conditional volatility of interest rate changes is constant and equal to σ. The results of CKLS (1992) illustrate that the assumption of constant volatility is questionable. The link between interest rate volatility and the level of the rate implies that the Vasicek model may provide unrealistic interest rate forecasts. When rates are low, volatility is overstated, and when interest rates are high, volatility is understated.

[2] It may be argued that this is not necessarily an implausible scenario. There have been some periods in the U.S. that real interest rates have been negative.

In summary, the Vasicek model is very tractable and provides convenient closed-form solutions for many interest rate-dependent instruments. However, the model has some serious drawbacks including restricted dynamics of the term structure and constant conditional volatility.

The Vasicek model is illustrated in Figures 1 through 3. Each exhibit illustrates the yield curves based on three different realizations of the modeled instantaneous rate. Figure 1 is based on parameter estimates from CKLS. However, it should be pointed out that, in their tests, they reject the model of Vasicek (1977) because of its homoskedastic feature. Note that when the instantaneous yield is high, the curve is inverted, and when the short rate is low, the curve is normal. In all cases, the long-term yields tend toward the parameter θ, the long-run average. In Figure 2, the mean reversion parameter (κ) has been increased. Note that longer yields revert back to the long-run average more quickly. In Figure 3, the long-run average is decreased and all yield curves tend to the lower long-run curve.

The Cox, Ingersoll, Ross Model

Another model of interest rates was formulated by Cox, Ingersoll, and Ross (1985) (CIR). The model is as follows:

$$dr_t = \kappa(\theta - r_t)dt + \sigma\sqrt{r_t}dB_t$$

The CIR model is also known as the "square-root" process because the volatility is related to the square root of the current level of the interest rate. Unlike the Vasicek model, the CIR model relates the conditional volatility to the level of the short rate. A second improvement of the CIR model over the Vasicek model is that interest rates cannot be negative [3]. Although CKLS find that interest rate volatility is more sensitive to the level of interest rates than proposed by the CIR

[3] Note that negative interest rates are ruled out in the continuous time case. However, it is possible that interest rates become negative if a discrete process is used in simulations.

specification, other researchers defend the model by commenting on the estimation approaches employed by CKLS (see Eom, 1994).

The CIR model can also be used to determine bond prices analytically. CIR show that bond prices are determined by the following (the "hats" simply indicate that the equations for *A(t,T)* and *B(t,T)* are different under CIR vs. the Vasicek model).

$$P(t,T) = \hat{A}(t,T)e^{-r(t)B(t,T)}$$

Thus, the equation above can be used to derive the yield curve for the CIR model. Because the driving factor of bond prices and yields is still the short-term rate, the CIR specification again assumes perfect correlation among all bonds and therefore restricts term structure dynamics.

The resulting yield curves of the CIR model are very similar to the Vasicek curves presented in Figures 1 through 3. The difference between the models relate more to the dynamics of yield curve fluctuations than to the shape of a particular curve given the instantaneous rate.

The Dothan Model

The model of Dothan (1978) increases the volatility exponent to 1.0:

$$dr_t = \sigma r_t dB_t$$

Because of the higher exponent, the model relates the volatility of interest rate movements more strongly to the level of interest rates. Courtadon (1982) extends Dothan's model to include mean reversion in the drift. Dothan's model is more difficult to implement in practice because there are no closed form, analytic solutions as in the Vasicek and CIR models. The user must resort to simulation to implement the model. Given the lack of closed form solutions and the inability of general equilibrium models to match the existing yield curve, the Dothan model has not been a popular model for use in evaluating interest rate securities.

Multi-Factor Models

To alleviate the problem of correlated bond prices, a model can incorporate two or more stochastic factors. In the two-factor model as described in Brennan and Schwartz (1979, 1982), one factor is used to represent the short-term rate while the other factor is the rate θ on a perpetuity (i.e., the long-term rate).

$$dr_t = \kappa(\theta - r_t)dt + \sigma_1 r_t dB_1$$
$$d\theta = \sigma_2 \theta dB_2$$
$$dB_1 dB_2 = \rho dt$$

σ_1, σ_2 = volatility of the short- and long-rate processes, respectively
dB_1, dB_2 = standard Brownian motions
ρ = correlation between short- and long-rate processes

Another popular two-factor model is presented in Longstaff and Schwartz (1992), where the second factor is stochastic volatility of the short-term rate. By explicitly modeling these factors separately, the potential range of yield curve dynamics is enhanced.

Heath, Jarrow, Morton Framework

The restrictions on yield curve movements of the one-factor models make them less exact, which in some cases, such as investment banking, represents a serious drawback. The main limitation is that yields of all maturities are perfectly correlated. However, history shows that different parts of the yield curve can shift in different directions and this can wreak havoc on an insurer's surplus. The interdependence across all maturities is most critical for insurers where assets and liabilities have unequal sensitivities at different points on the yield curve (see Reitano, 1990 and 1992).

Litterman and Scheinkman (1991) show that there are two additional factors, aside from parallel shifts in the yield curve, that have affected bond returns. The first factor, called steepening, reflects the fact that short-term rates may move in the opposite direction of long-term rates. The Brennan and Schwartz (1979) model above addresses the potential for a steepening term structure. The second factor affecting bond returns in Litterman and Scheinkman (1991) is a curvature component. This factor addresses the potential for intermediate yields to be more or less volatile than extreme maturities.

As mentioned above, a criticism of equilibrium models is that they are not arbitrage-free in the sense that the yield curves produced by the models do not match the existing term structure. This makes these models unsatisfactory for pricing option-embedded securities. If the model cannot accurately portray the existing term structure, there is little confidence that it will accurately imitate the dynamics of the curve (Hull, 1993).

Heath, Jarrow, and Morton (1992) use the no-arbitrage argument to develop the process for the *forward* rate implied by the relationship of bond prices.

$$df(t,T) = \mu(t,T,f(t,T))dt + \sigma(t,T,f(t,T))dB_t$$

$f(t,T)$ = instantaneous forward rate at time t with maturity T
$\mu(t,T,f(t,T))$ = drift of the forward rate process
$\sigma(t,T,f(t,T))$ = volatility of the forward rate process
B_t = standard Brownian motion

HJM find that by imposing the no-arbitrage argument to term structure movements, the drift of the forward rate process can be stated in terms of volatilities. Thus, the structure of the volatility becomes the most important element of the HJM model. Different functional forms of the volatility reveal an entire family of HJM models. In particular, a simple functional form is of the following type:

$$\sigma(t,T,f(t,T)) = \sigma_0 f(t,T)^\gamma$$

Amin and Morton (1994) look at a more general form and estimate the parameters of several specifications.

Historical Data

The choice of interest rate model can have an enormous impact on the resulting interest rate risk of any financial instrument. Although determining a perfect model of interest rates is beyond the scope of this paper, understanding the impact of the choice of interest rate model will assist insurers in analyzing the inherent risks of the embedded options in their liabilities and in choosing the appropriate model for their analyses. Any individual who wishes to use a model to simulate interest rate movements must first get a feel for historical changes. This section illustrates the historical movements in Treasury yields over the last 45 years. For ease of presentation, the focus will be on four critical points on the yield curve: (1) the one-year rate, (2) the three-year rate, (3) the five-year rate, and (4) the ten-year rate. Historical rates will then be compared with the theoretical models at these points. The data is taken from the St. Louis Federal Reserve web site.

The time series of the four yields is illustrated in Figure 4. A casual inspection of Figure 4 shows that interest rates increased from 1953 through 1979. Then, interest rates spiked in the early 1980s during the transition of the Federal Reserve policy mentioned above. Finally, since the peak in 1981, yields have exhibited a general downward trend.

Table 1 presents some summary statistics on the levels of yields over the 45 year period. These statistics help illustrate several features of historical interest rates. The first result relates to the shape of the yield curve on any particular date. The yield curve is a graphical representation of the relationship of the yield on bonds and their maturities. Figures 5 through 7 show three yield curves that have been observed historically. Typically, long-term yields are higher than short-term yields. When this occurs, the yield curve is upward sloping. The upward sloping yield curve is common enough that it is characterized as a "normal" curve as depicted in Figure 5. Occasionally, yield curves become inverted – short-term rates exceed long-term rates (see Figure 6). Inverted curves are typically observed in periods of high interest rates and the yield inversion is usually short-lived. Finally, humped yield curves are characterized by increasing yields at the short end of the curve. Eventually, as the term to maturity increases more, the yields begin to fall slightly (see Figure 7). Many humped yield curves occur during the transition from an inverted yield curve to a normal curve.

In Table 1, the yield curve is categorized according to its shape: normal, inverted, humped, or other. These classifications are made strictly on the four yield points, 1 year, 3 year, 5 year and 10 year. The precise definition of yield curve shape, as it is used here, is based on yield curve slope. The slope is the difference between two adjacent yields. A normal curve has positive slope everywhere, while an inverted curve has negative slope everywhere. A humped yield curve initially has positive slope and eventually has negative slope. If the yield curve does not fit into one of these profiles, it is classified as "other." Note that the yield curve classifications are based on end-of-the-month yields, so that a monthly observation is based on only one moment in time. If the yield curve is normal at that time even though it was inverted at all other times during the month, the curve is nonetheless classified as normal.

It should be noted that the magnitude of the slope does not impact our classification of yield curve slope. In particular, we do not use a "flat" yield classification. A flat yield curve exists if the yields on bonds of all maturities are equal. At no time in the 45 year history is the yield curve exactly flat. However, differences in yields of various maturities may be negligible. Rather than define the term negligible, the approach used here amounts to distributing almost-flat yield curves into the other categories.

Several statistics in Table 1 illustrate how often the normal yield curve occurs. First, the yield curve has been upward sloping over two-thirds of the time. Inversions occurred in only 11.6% of the months and a humped curve occurred 13.4% of the time. In addition to the frequencies of the various shapes, other statistical information points to the tendency of rates to be increasing with maturity. The mean of each of the four yields increases with maturity, and the yield percentiles seem to imply that the typical shape of the yield curve is normal, except when yields are high.

The next group of results illustrates the relationship of yield volatility and maturity. Long-term rates have lower standard deviations, lower skewness, a smaller range of outcomes, and higher autocorrelations than short rates. Thus, our earlier conjecture that long-term yields are less volatile than short-term rates seems to have statistical support.

Two other results are worth pointing out. First, short-term rates appear more positively skewed than longer yields. This could mean that changes in the long-term rate are more symmetric, or it could indicate that large, positive changes in the interest rates are more common in the short-term rate. The second point is that correlations between yields decrease as the rates are further apart.

Instead of looking at yield levels, Figures 8 through 11 look at changes in interest rates. Figure 8 looks at the monthly time series of absolute changes in the one-year yield rate. The Fed policy transition period stands out as an extremely volatile period for short-term rates. To gain further perspective on the transition period, Figure 9 looks at these changes on a relative basis. The extreme volatility of the early 1980s loses some of its distinction when viewed on a relative basis. The implication of Figures 8 and 9 provide some intuitive support for the CKLS result that interest rate volatility is related to the level of the interest rate. When interest rates were high, the percentage changes in yields were about the same as the percentage changes when rates were low. Figures 10 and 11 present a similar story for the 10-year yield.

To get a feel for the volatility of interest rate movements, we computed the standard deviations of the one-year and ten-year changes in yield. For the one-year yield, the standard deviation of absolute changes in the monthly rate over the entire period is 0.47 and the standard deviation of relative changes is 0.07. As expected, the volatility of changes in the ten-year yield is significantly less. The standard deviation of absolute changes in the monthly ten-year yield rate over the entire period is 0.29 and of percentage changes, the standard deviation is 0.03.

Simulations

The interest rate models presented in this paper have been introduced in a continuous time framework. Although some continuous time models may lead to closed form solutions for simple cash flows such as non-callable bonds, insurance liabilities are more complicated. To use the model's dynamics in insurance applications, such as in DFA, one must use discrete time intervals for the interest rate process. This section discusses how to translate the continuous time process into a discrete process and then illustrates the interest rate models presented in this paper through simulations.

As an example of discretization, consider the Vasicek model. Other models follow directly from the Vasicek results. Recall the Vasicek model:

$$dr_t = \kappa(\theta - r_t)dt + \sigma dB_t$$

By using short time intervals, the discrete process approximates the continuous process. More precise estimates will be obtained through the use of short time intervals (hours or minutes) which is most appropriate for trading activities. However, with insurance applications, long-term modeling is required and the use of longer intervals (such as monthly) is more appropriate.

All models presented here include a standard Brownian motion. Random changes in the Brownian motion are based on draws from the standard normal distribution scaled by the time interval. There are two popular approaches for generating standard normal distribution random variables. The first method is to take the average of twelve uniform random variables on the interval [0,1]. The second method is to translate two uniform random variables (u_1, u_2) according to the following:

$$\varepsilon = \sqrt{-2\ln(u_1)} \times \cos(2\pi u_2)$$

The monthly interest rate process then becomes (ε is the standard normal random variable):

$$\Delta r_t = \kappa(\theta - r_t) \times \frac{1}{12} + \sigma\varepsilon\sqrt{\frac{1}{12}}$$

We use this discrete approach to perform monthly simulations of several interest rate models. Our goal is to get a feel for how the models operate and to compare the resulting simulated yield distributions with the historical distribution. The Vasicek and CIR models are the most straightforward to simulate because yields have closed form solutions that depend on only the short rate. Also, the simulations are stable due to the mean reversion drift term. We simulate the yield curve for 10,000 months using the parameter estimates of CKLS:

Parameter	*Vasicek*	*CIR*
Mean reversion strength (κ)	0.1779	0.2339
Long-term rate (θ)	0.0866	0.0808
Volatility (σ)	0.0200	0.0854

The results of the Vasicek simulations are shown in Table 2. The shape of the yield curve is more frequently inverted than in the historical experience. In fact, the statistics show that the "average" yield curve is actually slightly inverted, but close to being flat. An inspection of the percentile statistics reveals that at low percentiles (when the one-year yield is low), the yield curve appears to be upward sloping. As the short rate increases, the curve is inverted. Another note from the shape frequencies illustrates the restrictions of the Vasicek model on the shape of the yield curve. The yield curve is normal, inverted, or humped. No other shape is seen under the Vasicek model. The standard deviation and percentile statistics show that the long yields are less volatile in the Vasicek model. All yields are perfectly correlated, as expected based on the fact that all yields are derived from the same instantaneous (short) rate. As explained in the presentation of the model, interest rates can become negative with the Vasicek model. In fact, the first percentile is negative.

Compared to the historical rates, the Vasicek model is negatively skewed and less peaked. This can be seen in the skewness and excess kurtosis statistics as well as by looking at the distributions of the one-year and ten-year yields. The historical distributions are shown in Figures 12 and 13 while the Vasicek simulation distributions are shown in Figures 14 and 15.

The CIR simulation results are presented in Table 3, and the distributions of one- and ten-year yields are illustrated in Figures 16 and 17. As in the Vasicek case, the CIR model is more frequently inverted than in historical data (47.6% inversions in the CIR simulation vs. 11.6% historically). The average yield curve is inverted but is close to being flat. The percentiles reveal a pattern similar to the Vasicek results. When the short rate is low, the curve appears normal. As the one-year yield increases, the yield curve inverts. One difference from the Vasicek results is that the median yield curve is almost perfectly flat. The yield curve shape is never other than normal, inverted, or humped.

The volatility of the ten-year yield is lower than the one-year yield volatility as measured by the standard deviation and interquartile range. Also, note that interest rates in the CIR model remain positive. The correlations among yields of all maturities are all 1.0. Finally, there is positive skewness for all rates, and the value is closer to the historical statistics than the Vasicek model. The distribution of longer maturities appears more peaked relative to historical numbers (see the excess kurtosis numbers and Figures 13 vs. 17).

Given the popularity of arbitrage free models, we present some short simulations of 100 months to see how these models function. Because the Ho-Lee model is the constant volatility case of the HJM model, we present a simulation on the more general HJM framework. Recall that the drift in an HJM framework is a function of the volatilities. Thus, unlike the Vasicek and CIR models, the drift is positive and the interest rate is not mean-reverting. Using long simulations to generate smooth distributions of yields is not possible because the curve will (on average) continue to increase. Rates quickly begin to drift to "unrealistic" levels. The arbitrage-free models are used to assure that the interest rate process does not generate arbitrage opportunities in the short term. As the interest rates are observed, the model is recalibrated and another simulation is performed. Thus, the simulation performed here uses only 100 months. In that simulation, the ending yield curve is near 13%, demonstrating the drift in these types of models.

Another difficulty when comparing HJM models to others is in calculating the shape statistics given the drifting problem. The shape of the curve becomes too dependent on the initial curve given the short simulation period. If the curve starts out as normal, most of the subsequent curves remain normal. Similarly, just the opposite occurs when the yield curve is initially inverted. In the simulations presented, the initial curve was based on year-end 1998 yields.

Results of the HJM simulation are presented in Table 4 and in Figures 18 and 19. The important feature of these results is that yields of different maturities are not driven by the same factor. Therefore, statistics such as skewness, excess kurtosis, and correlations are not exactly the same for all yields (although they are close). Contrast these differences with the results of Vasicek and CIR models, where these statistics are identical for all maturities.

Caveats

The results illustrated here use the entire historical period of April 1953 to July 1998 as a benchmark for comparing alternative models. This choice was based on obtaining a larger amount of data (compared with other studies) to generate smoother yield distributions, as well as to provide some perspective on interest rates over longer periods. However, the change in Fed policy in 1979 presents an important question regarding whether comparisons among interest rate models are robust to the Fed's shift in focus. To look at these effects, a similar analysis could be performed across different subperiods. One possible breakdown would look at results under the two different Federal Reserve policies. Yield statistics can be generated under the "interest rate target policy" and also under the "inflation target policy." Another subperiod analysis could attempt to isolate the transition period and compare the pre- and post-transition periods to determine if the new Fed policy has affected the underlying interest rate dynamics. It should be pointed out that other factors may be contributing to the dynamics of the curve across any subperiod analysis. For example, the post-transition economy has been very strong with only one short recessionary period. Using only post-transition statistics may not completely embody the true potential for interest rate movements. The main point is that when using yield curve statistics to parameterize an interest rate model, one should be aware of any underlying factors that may be affecting the dynamics of yields and incorporate judgment in choosing specific models.

As mentioned above, the parameters used in the simulations were based on estimates reported by CKLS (1992) and Amin and Morton (1994). The CKLS study looks at the period from June 1964 through December 1989. Amin and Morton look at the period from 1987 through 1992. Estimation over different time periods will more than likely generate different parameters. Thus, one must keep in mind the interest rate environment when estimating parameters from past data for use in future periods. Care should be taken to ensure that the potential interest rate dynamics are consistent with the parameter assumptions.

Conclusion

Interest rate volatility now requires that actuaries incorporate stochastic interest rate assumptions into the pricing, forecasting, and valuation processes. The goal of this paper has been to provide a simplified introduction to and illustration of these models. The focus has been on comparing the results of simulations based on a variety of stochastic interest rate models with historical interest rate statistics. It is hoped that this work helps casualty actuaries begin the process of incorporating these modeling skills into their actuarial toolkits.

REFERENCES

Amin, K. and A. Morton, 1994, "Implied volatility functions in arbitrage-free term structure models," *Journal of Financial Economics* 35, 141-180.

Brennan, M.J. and E.S. Schwartz, 1979, "A Continuous Time Approach to Pricing Bonds," *Journal of Banking and Finance*, 3: 133-55.

Brennan, M.J. and E.S. Schwartz, 1982, "An Equilibrium Model of Bond Pricing and a Test of Market Efficiency," *Journal of Financial and Quantitative Analysis*, 17 (September): 301-329.

Chan, K.C., G.A. Karolyi, F.A. Longstaff, and A.B. Sanders, 1992, "An Empirical Comparison of Alternative Models of the Short-Term Interest Rate," *Journal of Finance*, 47: 1209-1227.

Courtadon, G., 1982, "The Pricing of Options on Default-Free Bonds," *Journal of Financial and Quantitative Analysis*, 17: 75-101.

Cox, J.C., J.E. Ingersoll, and S.A.Ross, 1985, "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53: 385-407.

D'Arcy, S. P., R. W. Gorvett, J. A. Herbers, T. E. Hettinger, S. G. Lehmann, and M. J. Miller, 1997, "Building a Public Access PC-Based DFA Model," *Casualty Actuarial Society Forum*, Fall 1997, Vol. 2, pp. 1-40.

D'Arcy, S. P., R. W. Gorvett, T. E. Hettinger, and Robert J. Walling, III, 1998, "Using the Public Access DFA Model: A Case Study," *Casualty Actuarial Society Forum*, Summer 1998, pp. 53-118.

Dothan, U.L., 1978, "On the term structure of interest rates," *Journal of Financial Economics* 6, 59-69.

Eom, Y.H., 1994, "In Defense of the Cox, Ingersoll, and Ross Model: Some Empirical Evidence," Working Paper, New York University.

Heath, D., R. Jarrow, and A. Morton, 1992, "Bond Pricing and the Term Structure of Interest Rates: A New Methodology," *Econometrica*, 60: 77-105.

Ho, T. and S. Lee, 1986, "Term structure movements and pricing interest rate contingent claims," *Journal of Finance* 41, 1011-1029.

Hull, J.C., 1993, *Options, Futures, and Other Derivative Securities*, 2nd ed. Englewood Cliffs, N.J.: Prentice-Hall.

Litterman, R. and J. Scheinkman, 1991, "Common Factors Affecting Bond Returns," *Journal of Fixed Income*, 1:54-61.

Longstaff, F. and E. Schwartz, 1992, Interest rate volatility and the term structure: a two-factor general equilibrium model, *Journal of Finance*, 47:1259-1282.

Phoa, W., 1997, "Principles of Pricing and Valuation Models," IBC Conference Mitigating Risk Using Derivatives, June, 1997.

Reitano, R.R., 1990, "Non-parallel Yield Curve Shifts and Durational Leverage," *Journal of Portfolio Management*, 16: 62-67.

Reitano, R.R., 1992, "Nonparallel Yield Curve Shifts and Convexity," *Transactions of the Society of Actuaries*, 44: 479-507.

Vasicek, O., 1977, "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics*, 5: 177-188.

TABLE 1

Historical Yield Statistics
Entire Period (4/53 - 7/98)

Yield Curve Shape

Normal	68.8%
Inverted	11.6%
Humped	13.4%
Other	6.3%

Yield Statistics

	1 Yr	3 Yr	5 Yr	10 Yr
Mean	6.08	6.47	6.64	6.81
Std Dev	3.01	2.88	2.84	2.81
Skewness	0.97	0.84	0.77	0.68
Excess Kurtosis	1.10	0.69	0.48	0.16

Percentiles

	1 Yr	3 Yr	5 Yr	10 Yr
1%	1.07	1.59	1.94	2.38
5%	2.05	2.52	2.72	2.90
10%	2.94	3.38	3.47	3.48
25%	3.81	4.17	4.24	4.25
50%	5.61	6.20	6.44	6.68
75%	7.71	8.01	8.04	8.20
90%	9.97	10.47	10.63	10.78
95%	12.08	12.48	12.59	12.56
99%	15.17	14.69	14.59	14.29

Correlations

		1 Yr	3 Yr	5 Yr	10 Yr
1 Yr		1.000	0.985	0.969	0.944
3 Yr			1.000	0.997	0.984
5 Yr				1.000	0.995
10 Yr					1.000
Auto					
	1	0.988	0.991	0.993	0.995
	2	0.967	0.976	0.980	0.986
	3	0.948	0.963	0.970	0.979
	4	0.932	0.951	0.960	0.972
	5	0.918	0.940	0.951	0.964

TABLE 2

Vasicek Simulation Statistics
(10,000 Simulations)

Yield Curve Shape

Normal	41.6%
Inverted	54.8%
Humped	3.6%
Other	0.0%

Yield Statistics

	1 Yr	3 Yr	5 Yr	10 Yr
Mean	8.81	8.75	8.68	8.52
Std Dev	3.83	3.24	2.77	1.95
Skewness	-0.16	-0.16	-0.16	-0.16
Excess Kurtosis	-0.19	-0.19	-0.19	-0.19

Percentiles

	1 Yr	3 Yr	5 Yr	10 Yr
1%	-0.38	0.97	2.04	3.84
5%	2.33	3.27	4.00	5.22
10%	3.69	4.42	4.98	5.92
25%	6.26	6.60	6.84	7.23
50%	8.94	8.86	8.77	8.59
75%	11.62	11.13	10.72	9.96
90%	13.60	12.80	12.14	10.97
95%	14.69	13.73	12.94	11.53
99%	17.22	15.87	14.76	12.82

Correlations

		1 Yr	3 Yr	5 Yr	10 Yr
1 Yr		1.000	1.000	1.000	1.000
3 Yr			1.000	1.000	1.000
5 Yr				1.000	1.000
10 Yr					1.000
Auto					
	1	0.991	0.991	0.991	0.991
	2	0.982	0.982	0.982	0.982
	3	0.973	0.973	0.973	0.973
	4	0.965	0.965	0.965	0.965
	5	0.956	0.956	0.956	0.956

Note: Model parameters from CKLS estimates: $\kappa = 0.1779$, $\theta = 0.0866$, $\sigma = 0.0200$

TABLE 3

CIR Simulation Statistics
(10,000 Simulations)

Yield Curve Shape

Normal	47.7%
Inverted	47.6%
Humped	4.7%
Other	0.0%

Yield Statistics

	1 Yr	3 Yr	5 Yr	10 Yr
Mean	8.08	8.04	7.98	7.86
Std Dev	2.89	2.31	1.88	1.20
Skewness	0.92	0.92	0.92	0.92
Excess Kurtosis	1.49	1.49	1.49	1.49

Percentiles

	1 Yr	3 Yr	5 Yr	10 Yr
1%	2.92	3.90	4.62	5.71
5%	3.95	4.73	5.29	6.14
10%	4.73	5.35	5.80	6.46
25%	6.14	6.48	6.71	7.05
50%	7.71	7.73	7.73	7.70
75%	9.57	9.23	8.95	8.48
90%	11.80	11.01	10.40	9.41
95%	13.42	12.31	11.45	10.09
99%	17.19	15.33	13.90	11.66

Correlations

		1 Yr	3 Yr	5 Yr	10 Yr
1 Yr		1.000	1.000	1.000	1.000
3 Yr			1.000	1.000	1.000
5 Yr				1.000	1.000
10 Yr					1.000
Auto					
	1	0.976	0.976	0.976	0.976
	2	0.955	0.955	0.955	0.955
	3	0.934	0.934	0.934	0.934
	4	0.914	0.914	0.914	0.914
	5	0.894	0.894	0.894	0.894

Note: Model parameters from CKLS estimates: $\kappa = 0.2339$, $\theta = 0.0808$, $\sigma = 0.0854$

TABLE 4

HJM Simulation Statistics
(100 Simulations)

Yield Statistics

	1 Yr	3 Yr	5 Yr	10 Yr
Mean	7.39	7.51	7.60	7.80
Std Dev	2.26	2.27	2.31	2.44
Skewness	0.51	0.53	0.54	0.54
Excess Kurtosis	-0.88	-0.85	-0.85	-0.86

Percentiles

	1 Yr	3 Yr	5 Yr	10 Yr
1%	4.45	4.48	4.52	4.59
5%	4.79	4.85	4.90	4.99
10%	5.00	5.10	5.13	5.21
25%	5.25	5.45	5.53	5.63
50%	7.48	7.58	7.65	7.83
75%	8.65	8.75	8.85	9.10
90%	11.02	11.16	11.30	11.68
95%	11.57	11.74	11.92	12.38
99%	12.09	12.26	12.44	12.89

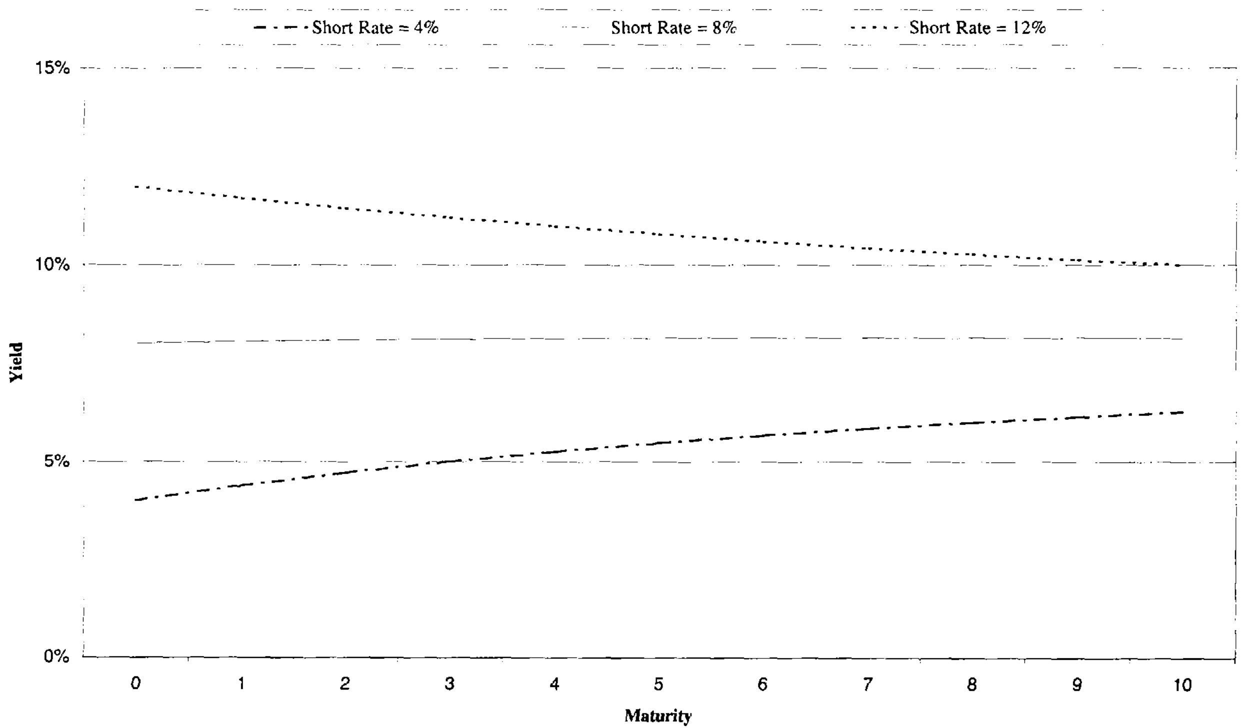
Correlations

	1 Yr	3 Yr	5 Yr	10 Yr
1 Yr	1.000	0.999	0.999	0.999
3 Yr		1.000	1.000	1.000
5 Yr			1.000	1.000
10 Yr				1.000
Auto				
1	0.986	0.986	0.987	0.987
2	0.969	0.969	0.969	0.972
3	0.954	0.953	0.954	0.957
4	0.938	0.938	0.939	0.943
5	0.925	0.923	0.925	0.929

Note: Model parameters from Amin and Morton: σ =0.0485, γ =0.5

FIGURE 1

Vasicek Model Yield Curves
CKLS Parameters

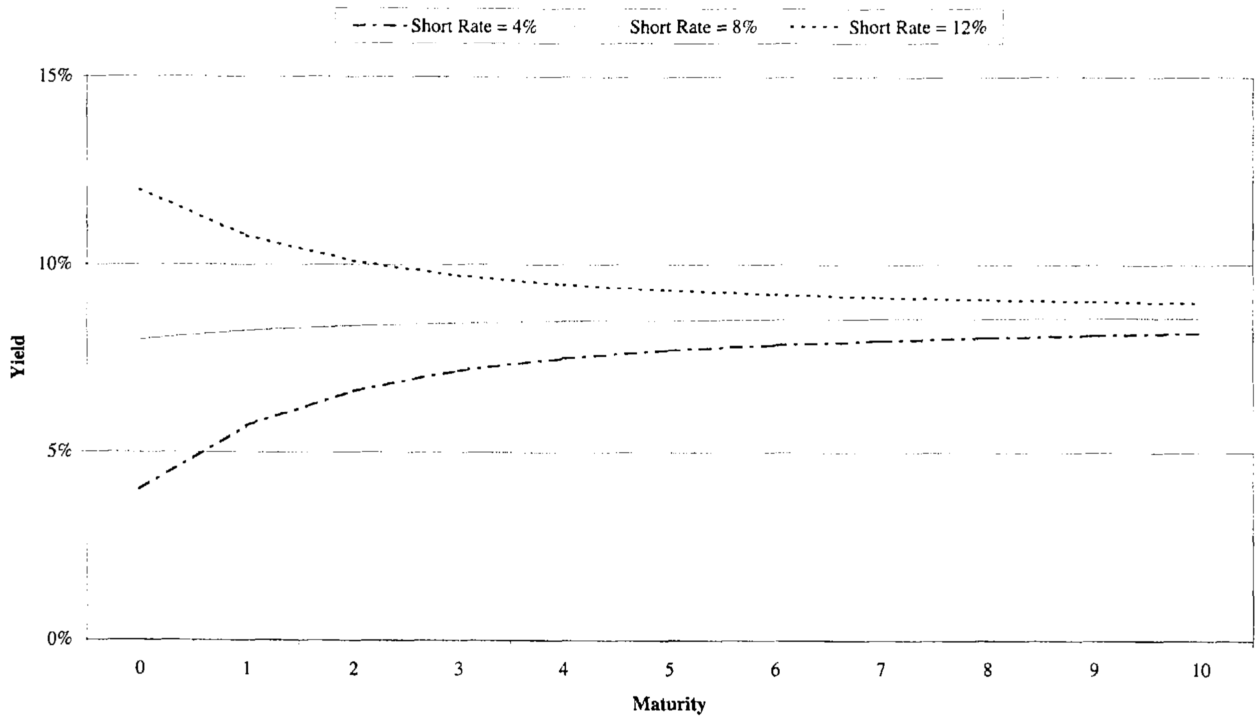


Parameters: $\kappa = 0.1779$, $\theta = 0.0866$, $\sigma = 0.0200$

FIGURE 2

Vasicek Model Yield Curves

CKLS Estimates - Change in Mean Reversion

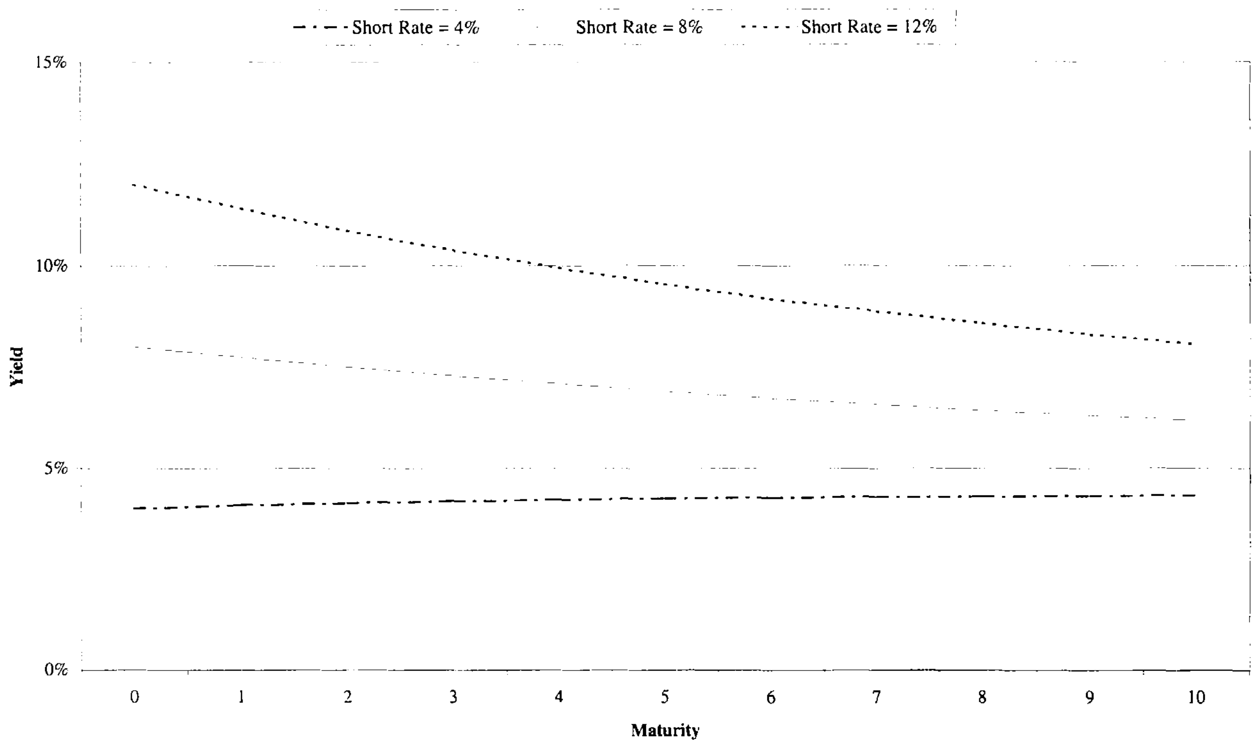


Parameters: κ = 1.0000, θ = 0.0866, σ = 0.0200

FIGURE 3

Vasicek Model Yield Curves

CKLS Parameters - Change in Long-Term Rate



Parameters: $\kappa = 0.1779$, $\theta = 0.0500$, $\sigma = 0.0200$

FIGURE 4
Time Series of Yields

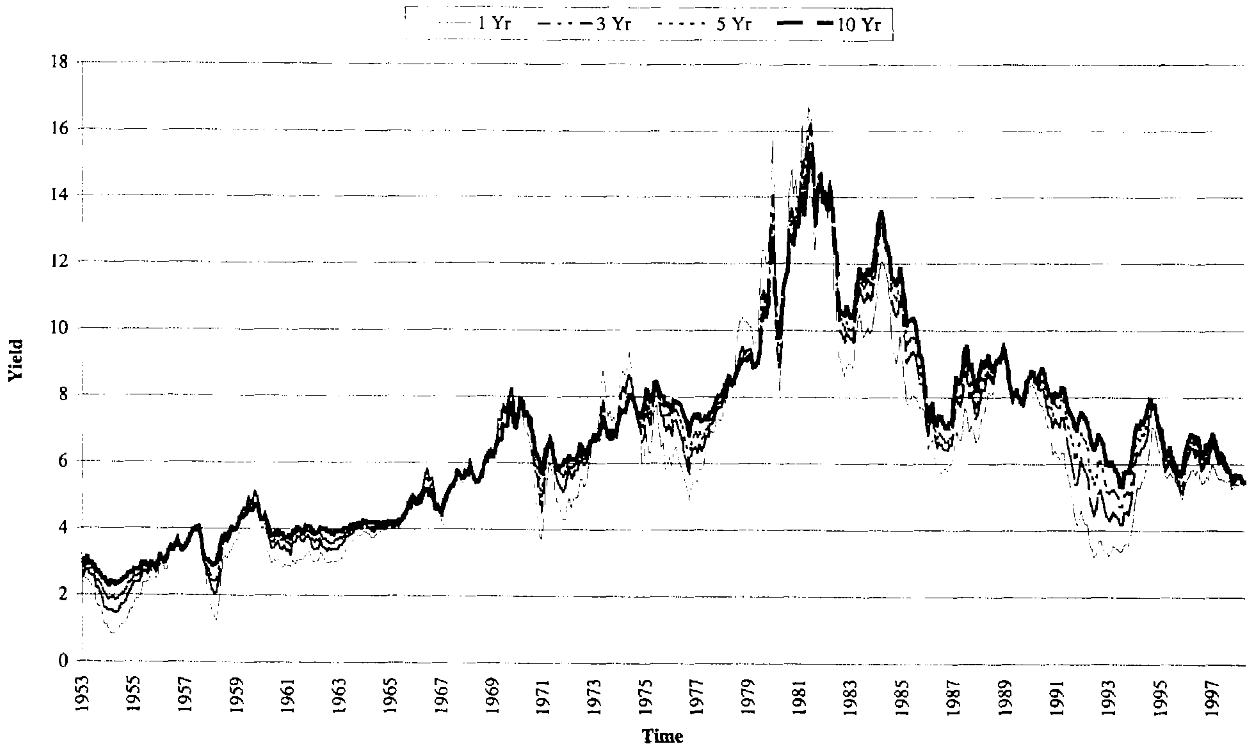


FIGURE 5
Normal Yield Curve
June 1991

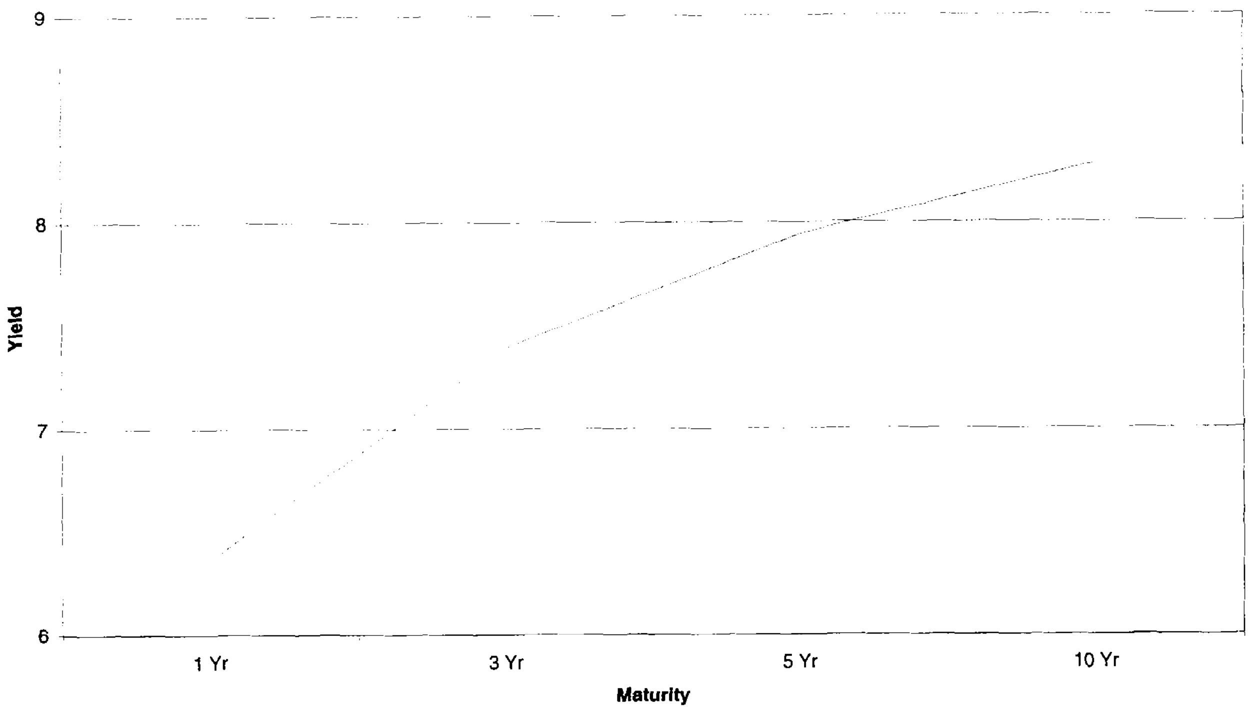


FIGURE 6
Inverted Yield Curve
August 1973

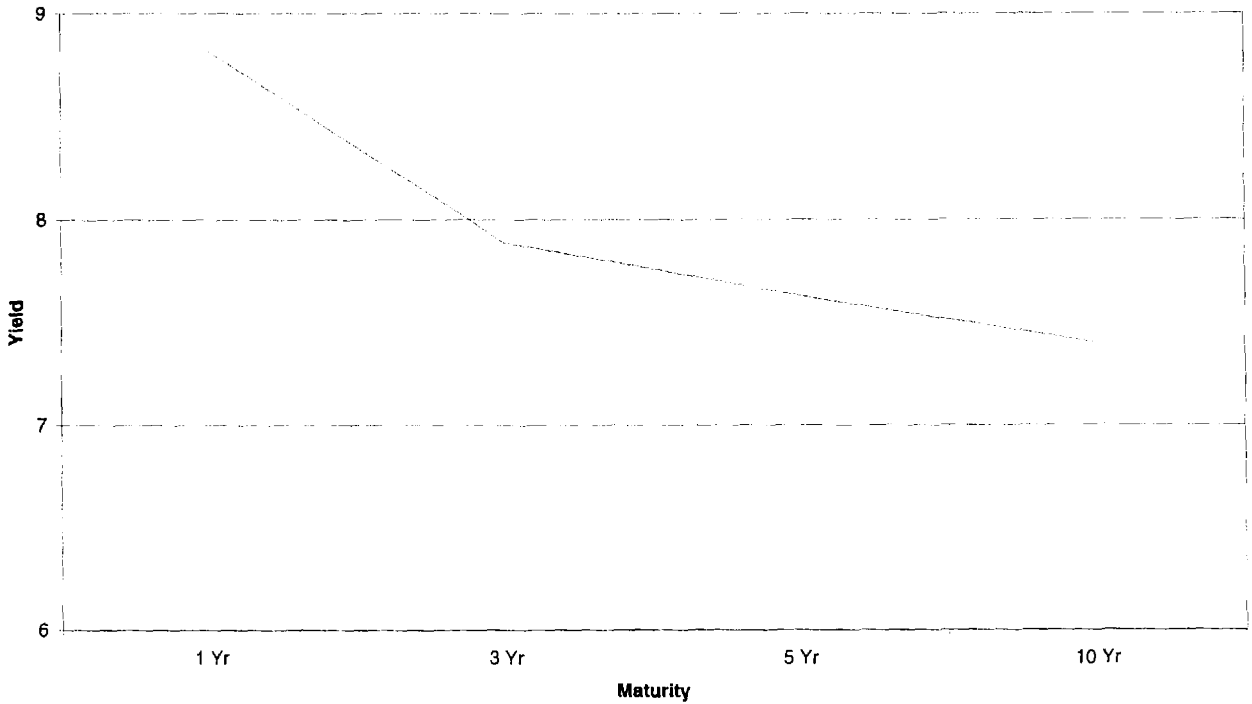


FIGURE 7
Humped Yield Curve
July 1970

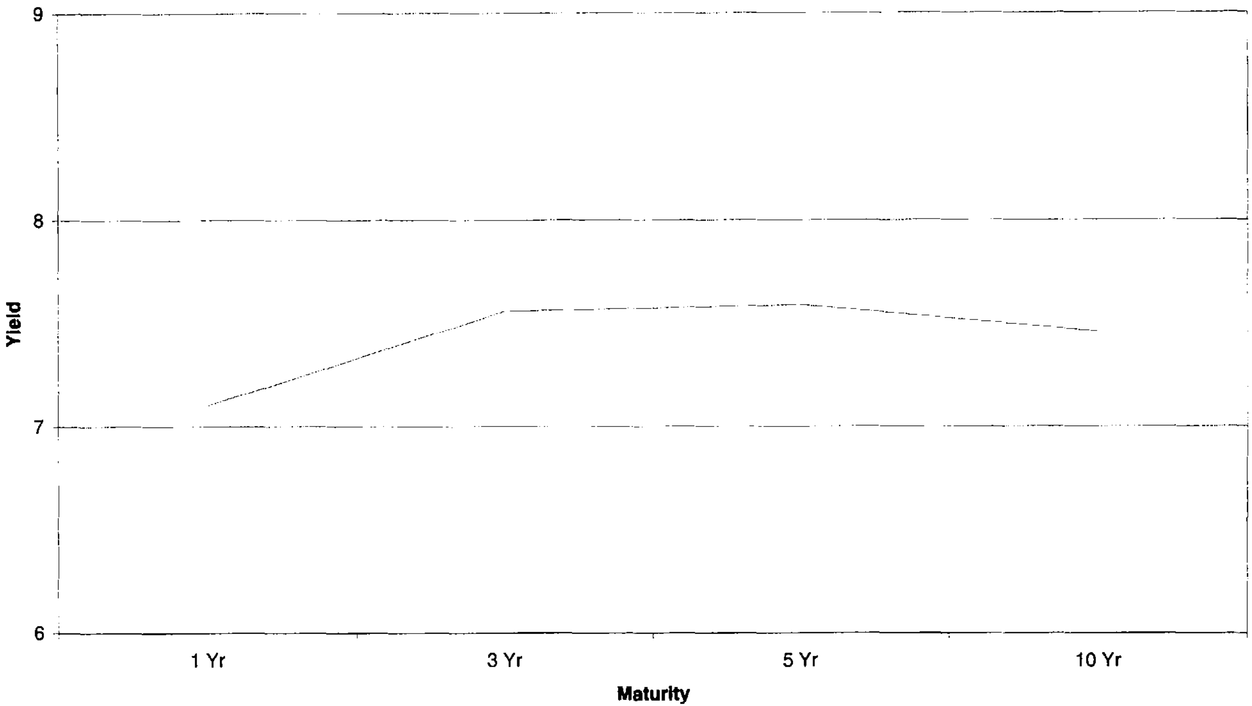


FIGURE 8

Time Series of Monthly Absolute Change in 1 Year Yield

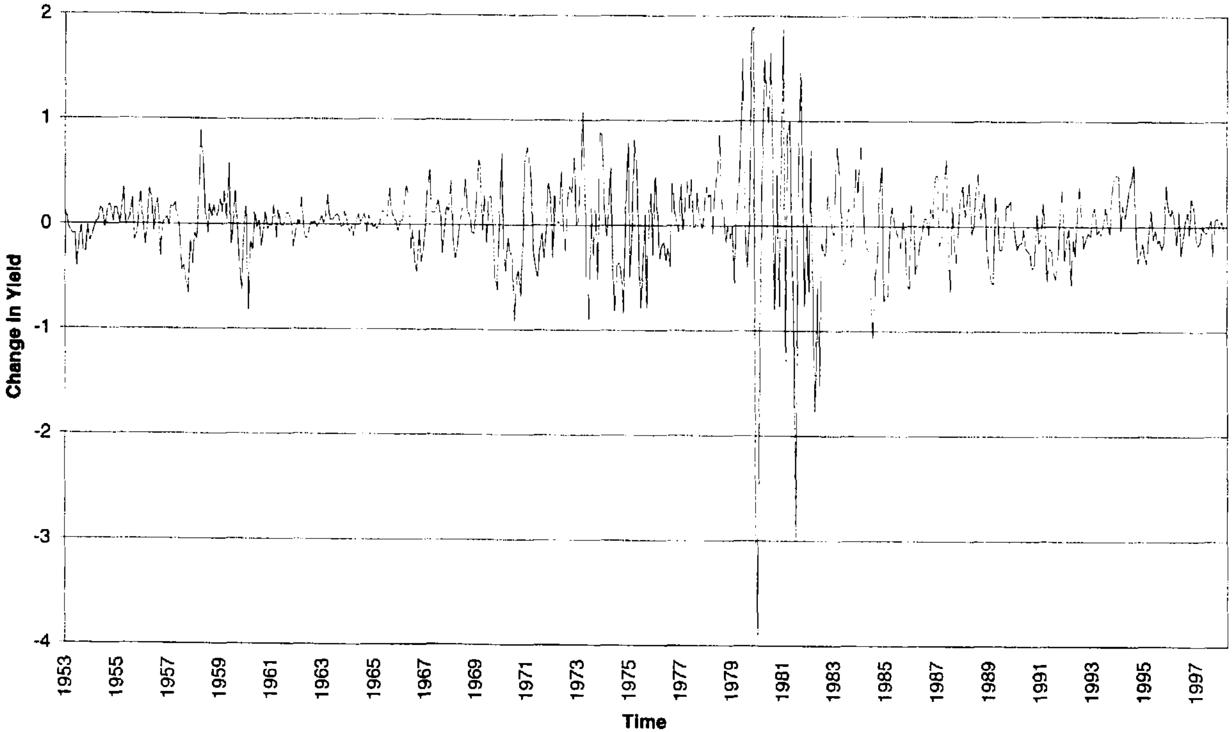


FIGURE 9

Time Series of Monthly Percentage Change in 1-Year Yields

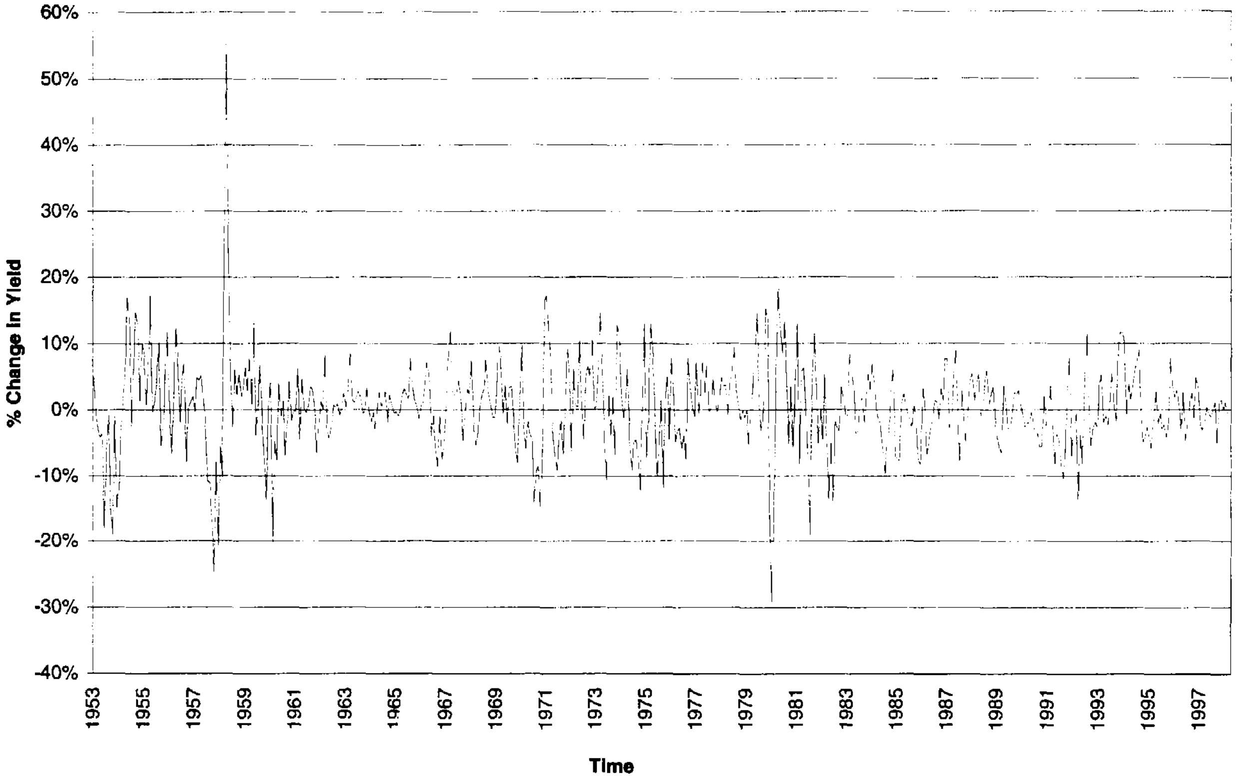


FIGURE 10

Time Series of Monthly Absolute Change in 10-Year Yield

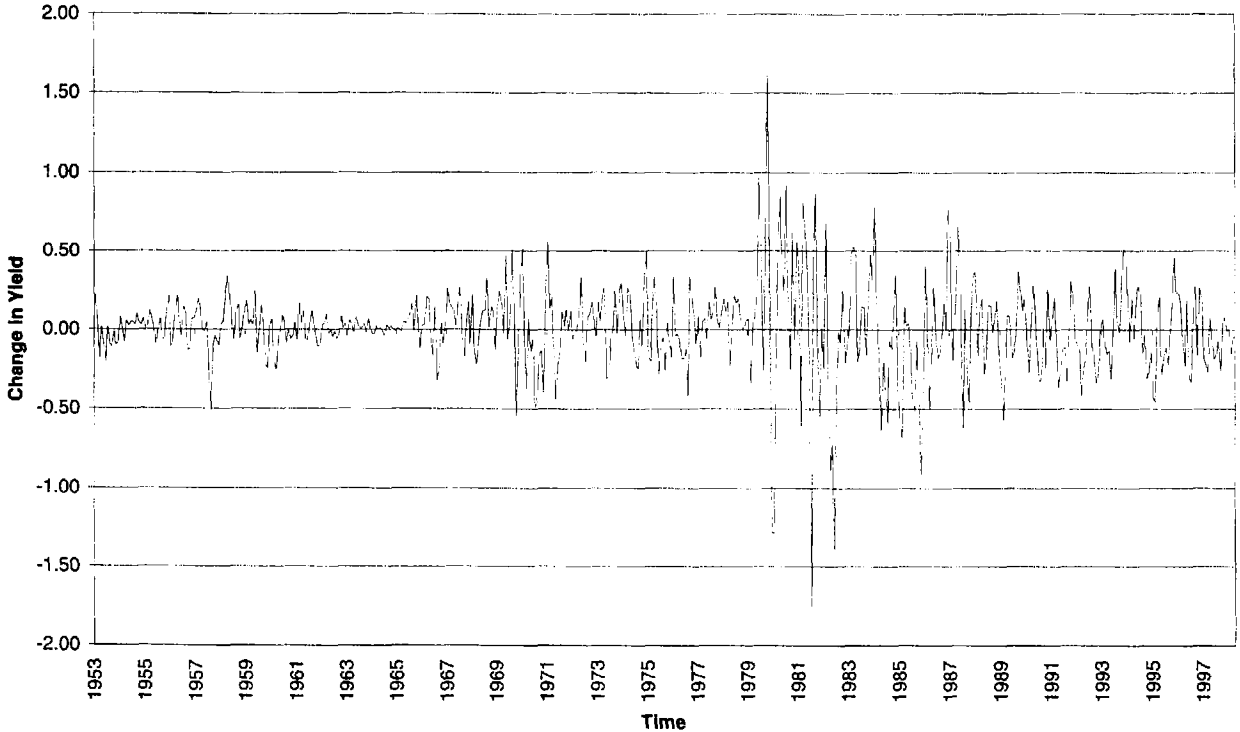


FIGURE 11

Time Series of Monthly Percentage Change in 10-Year Yield (Historical)

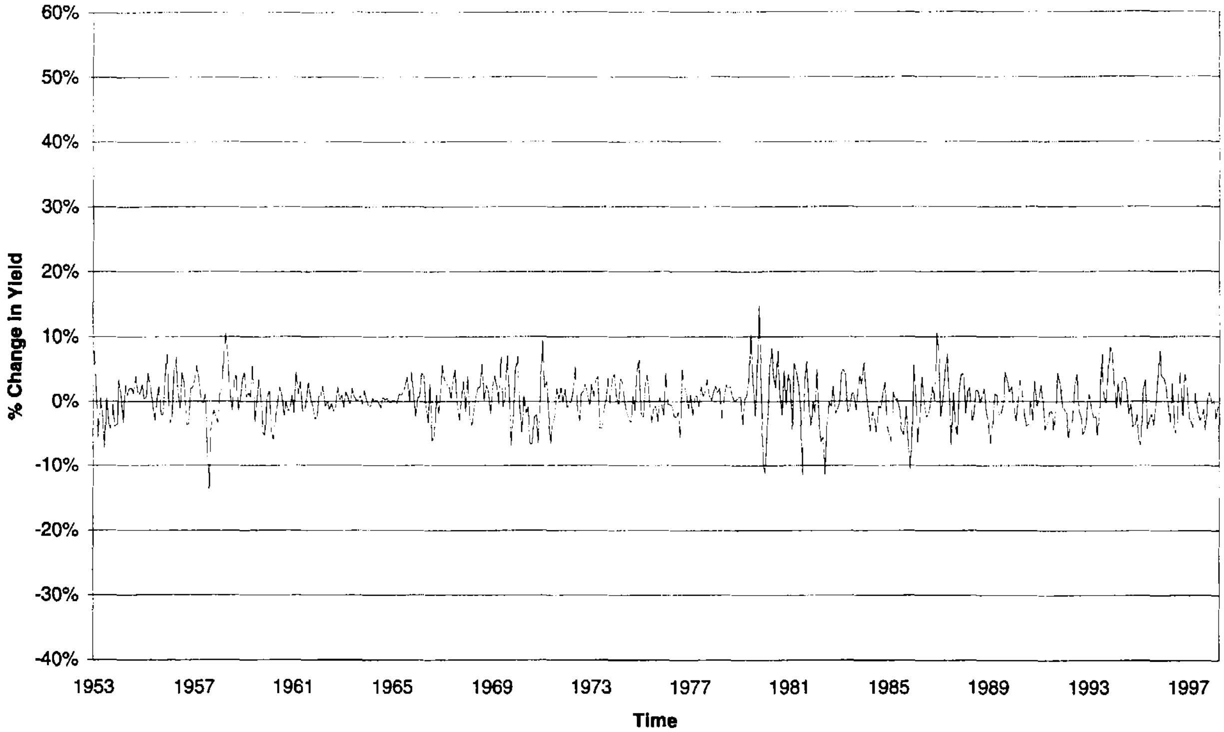


FIGURE 12
Historical 1 Year Yield Distribution

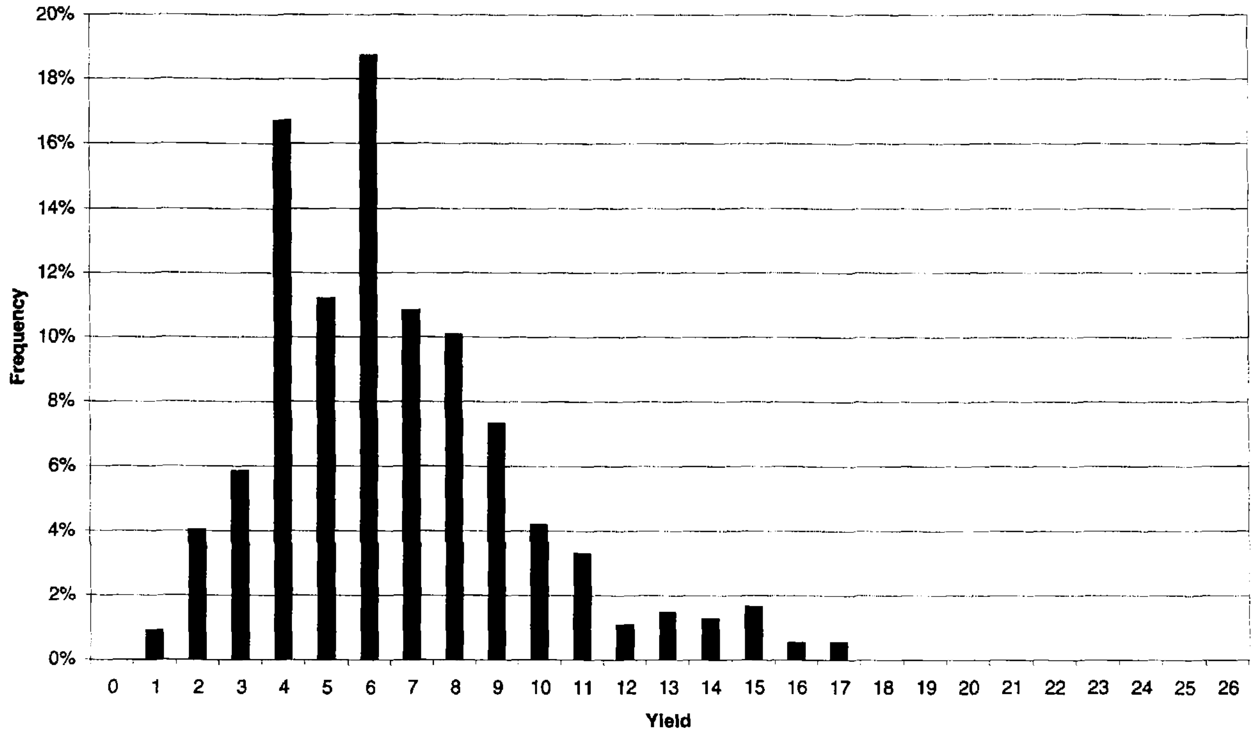


FIGURE 13
Historical 10 Year Yield Distribution

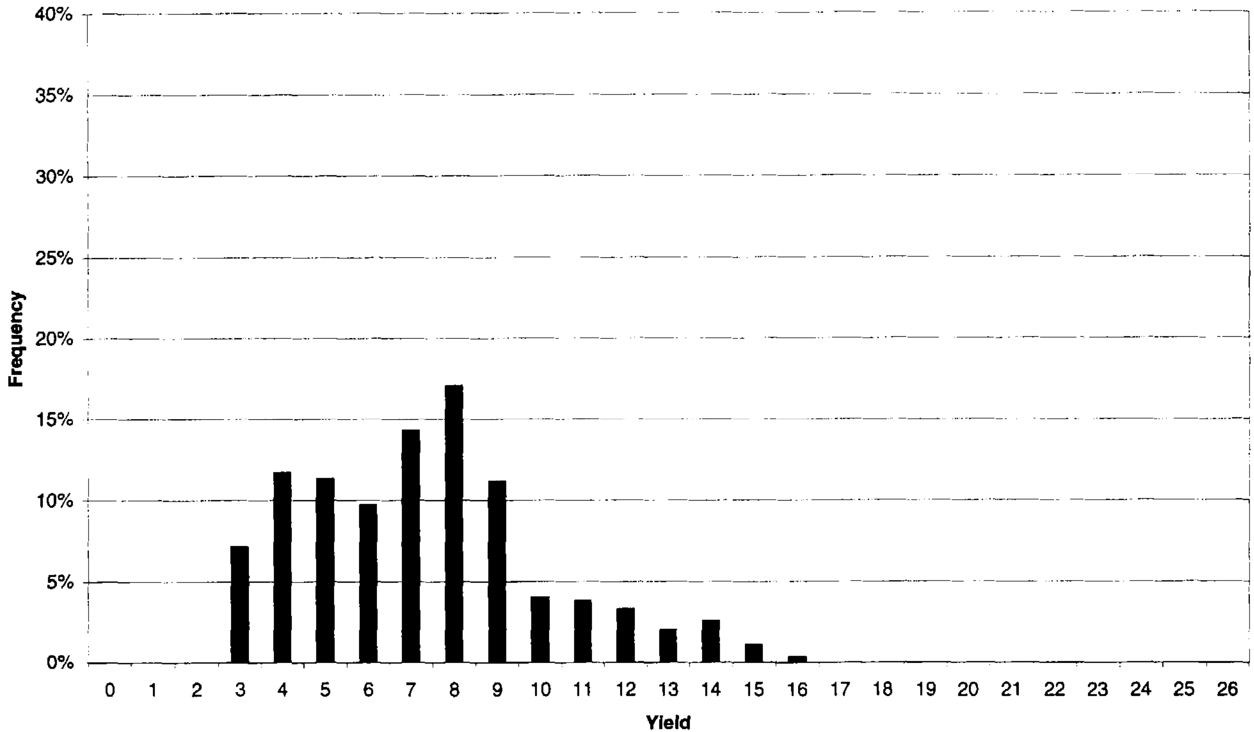
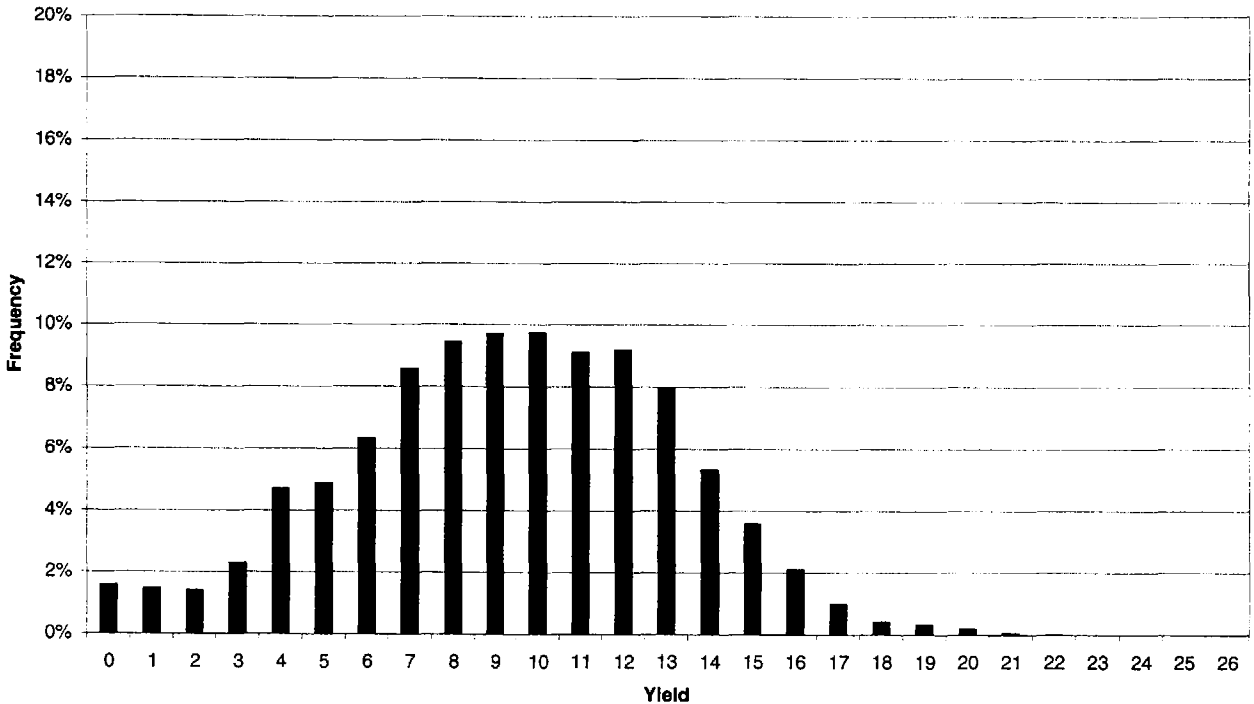
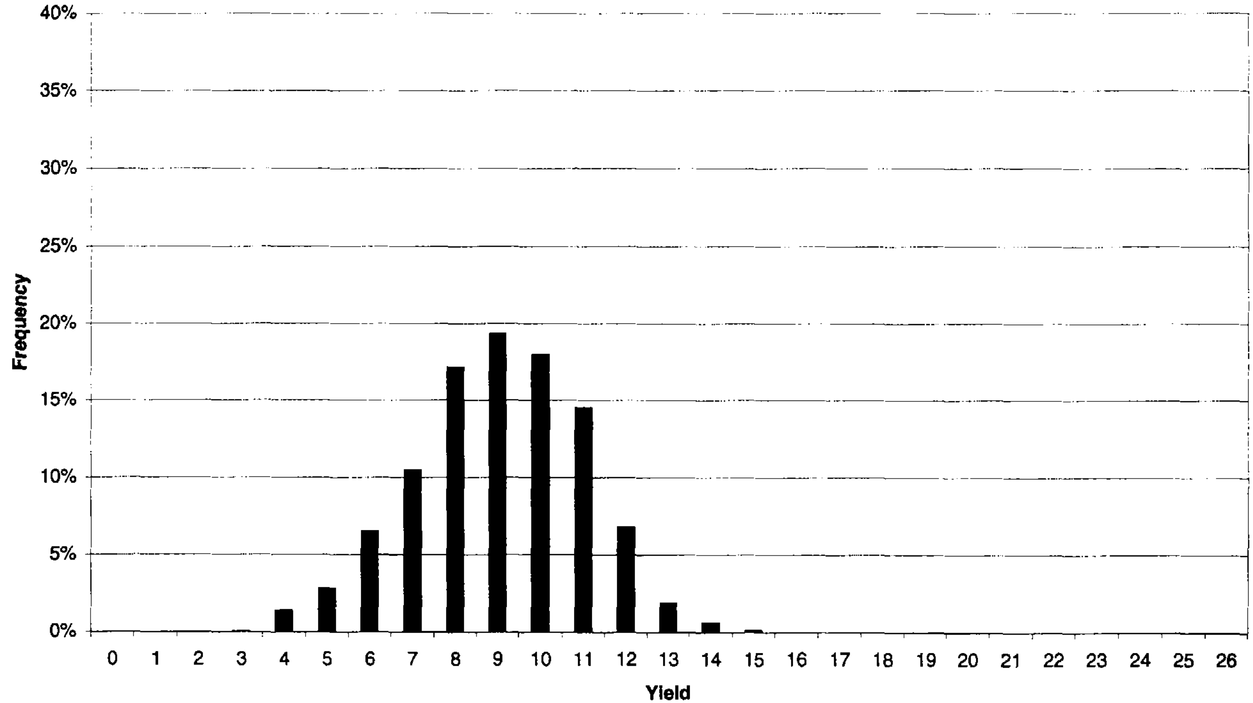


FIGURE 14
Vasicek Simulation
1 Year Yield Distribution



Note: Model parameters from CKLS estimates: $\kappa = 0.1779$, $\theta = 0.0866$, $\sigma = 0.0200$

FIGURE 15
Vasicek Simulation
10 Year Yield Distribution

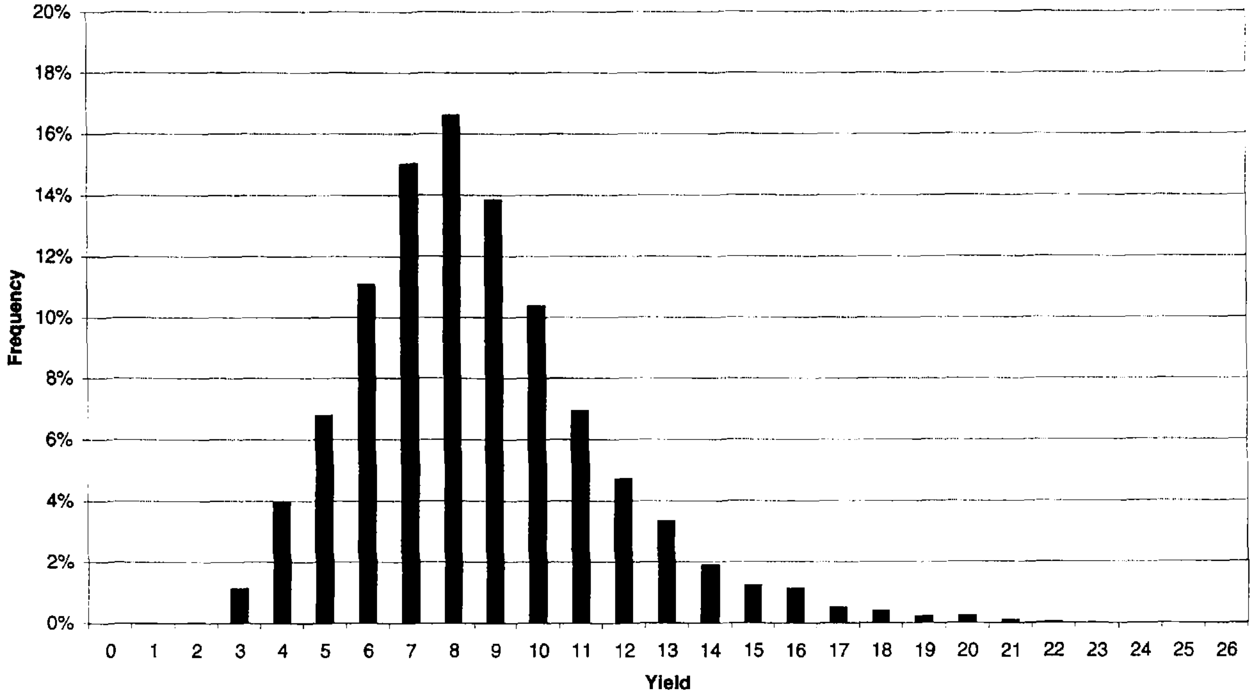


Note: Model parameters from CKLS estimates: $\kappa = 0.1779$, $\theta = 0.0866$, $\sigma = 0.0200$

FIGURE 16

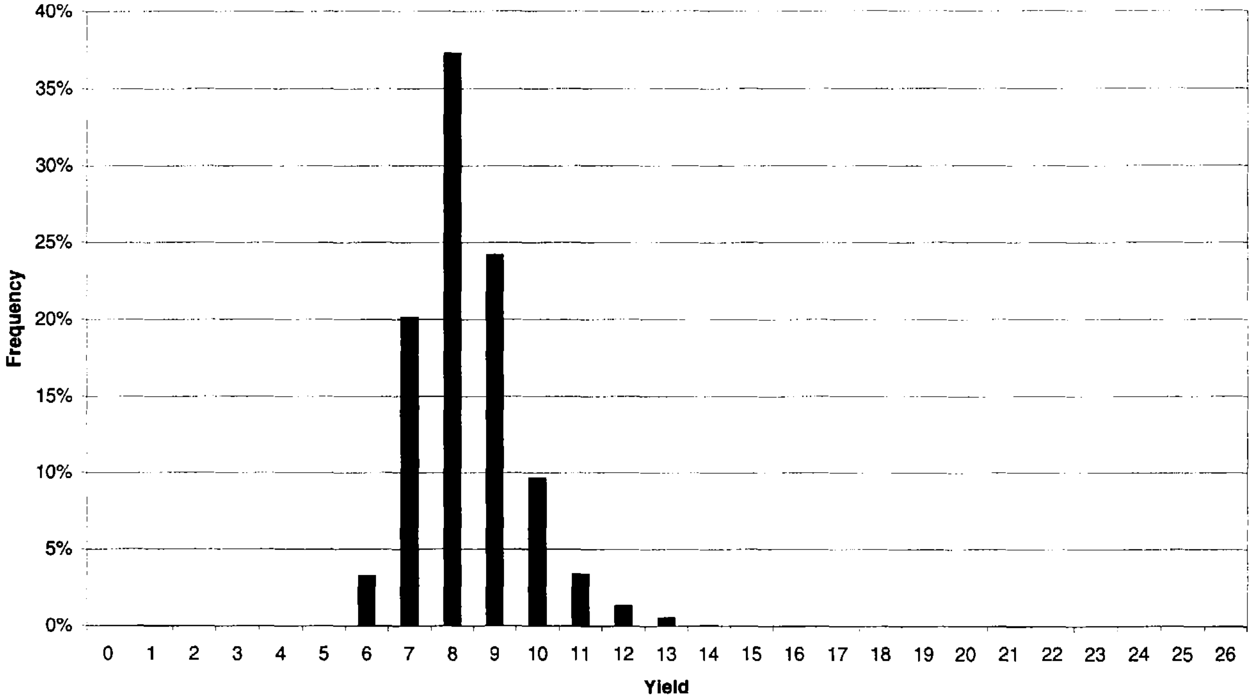
CIR Simulation

1 Year Yield Distribution



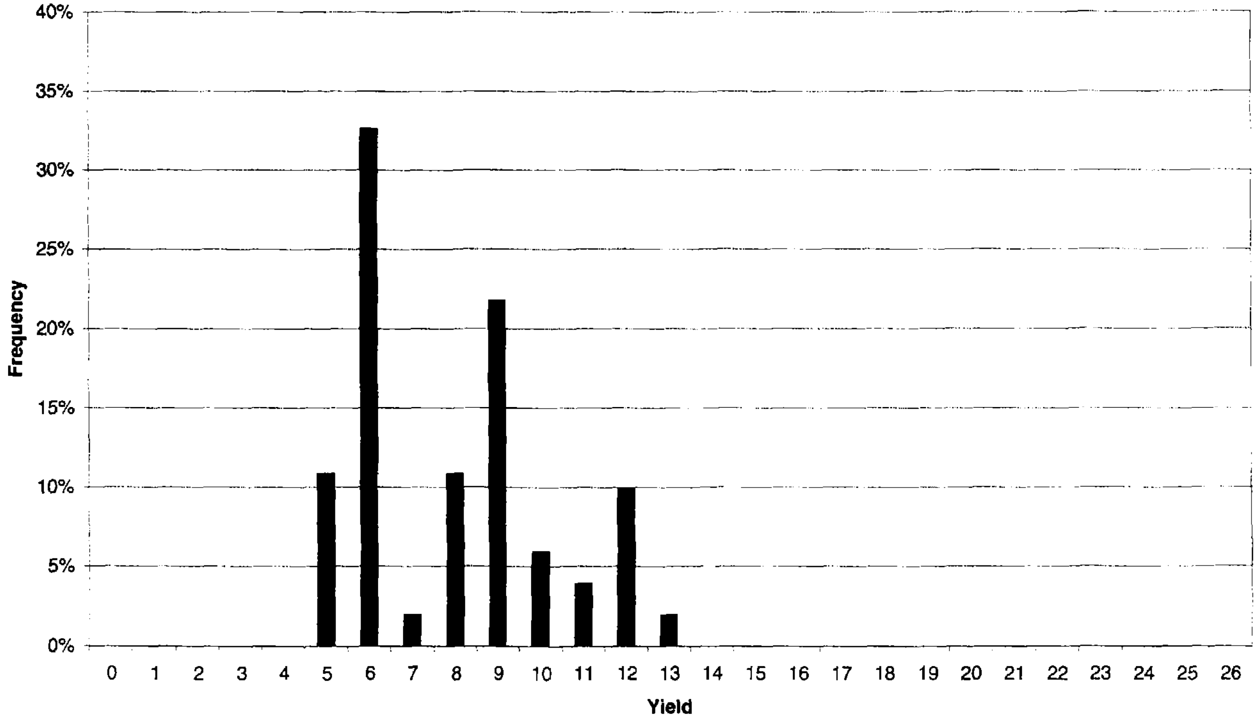
Note: Model parameters from CKLS estimates: $\kappa = 0.2339$, $\theta = 0.0808$, $\sigma = 0.0854$

FIGURE 17
CIR Simulation
10 Year Yield Distribution



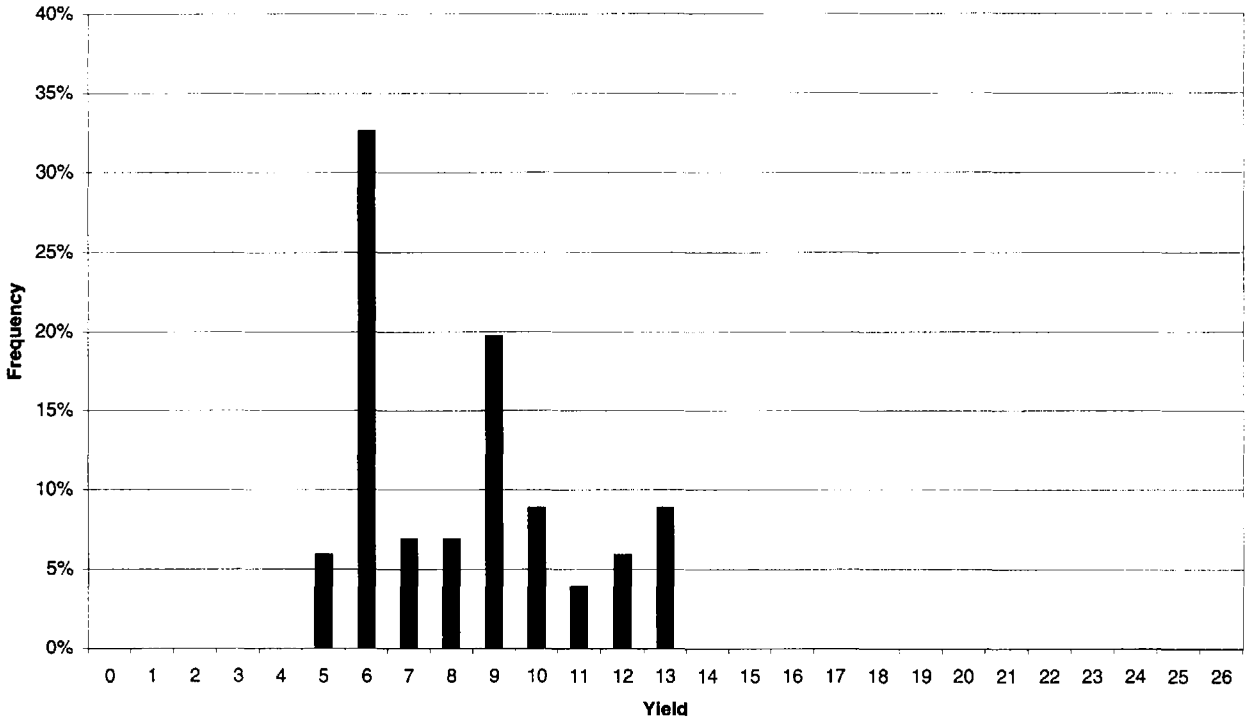
Note: Model parameters from CKLS estimates: $\kappa = 0.2339$, $\theta = 0.0808$, $\sigma = 0.0854$

FIGURE 18
HJM Simulation
1 Year Yield Distribution



Note: Model parameters from Amin and Morton: σ=0.0485, γ=0.5

FIGURE 19
HJM Simulation
10 Year Yield Distribution



Note: Model parameters from Amin and Morton: σ=0.0485, γ=0.5

A Comprehensive System for Selecting and Evaluating DFA Model Parameters

Adam J. Berger, and
Chris K. Madsen, ASA, MAAA

A Comprehensive System for Selecting and Evaluating DFA Model Parameters

Adam J. Berger, Ph.D.
Lattice Financial, LLC
55 Princeton-Hightstown Road, Suite 201
Princeton Junction, NJ 08550
ABerger@LatticeFinancial.com
(609) 716-4672

Chris K. Madsen, ASA, CFA, MAAA
American Re-Insurance Company
555 College Road East
Princeton, NJ 08543-5241
CMadsen@AmRe.com
(609) 243-4311

Abstract

Stochastic scenario generators for assets and liabilities are critical components of a robust DFA model. Vital to any stochastic scenario generation system is the selection of the underlying parameters. The process of parameter estimation is second only to model structure in the quest for generating reasonable results. If the model is simple, we can use standard statistical methods such as maximum likelihood to estimate parameters. However, for very complex models, we need to establish criteria for evaluation and find the parameters that are best with respect to those criteria.

In this paper, we discuss a parameter estimation system called American Re-Insurance Company's Constraint Evaluator System. This system allows modelers to define a multitude of targets and to assign a weight to each target to create a comprehensive objective function. Each target represents a quality that the model should possess with an assigned level of significance (weight). The targets are based on historical analysis or on some rational vision for future relationships. We discuss the analysis involved in setting appropriate targets including the monitoring of relationships between variables in a multi-period environment.

Our goal is to minimize the deviation between the user-defined targets and the model output. This is a non-convex optimization problem, which we use a combination of techniques to solve. Finally, we study the robustness of our parameter estimates as it relates to the number of scenarios and the observed model outputs.

1. Introduction

Stochastic scenario generators for assets and liabilities are important components of a robust DFA system. These generators will forecast asset and liability distributions over time as part of the development of income statement and balance sheet projections. These forecasts are developed as a collection of individual scenarios. Each scenario represents one possible future, and by looking at many scenarios, distributions can be calculated at any point in time. Examples of such systems can be found in Berger and Mulvey (1998), Dempster and Thorlacius (1998), Wilkie (1986), and Mulvey and Thorlacius (1998).

In developing this scenario-based approach, modelers try to understand fundamental economic and asset market structures. For example, when inflation is increasing, how will the stock and bond markets react? By understanding fundamental relationships, more realistic scenarios can be generated. These relationships can be modeled with mathematical equations, thus grounding the model in some amount of economic theory. The danger, however, is that the resulting scenarios don't exhibit characteristics seen in the market historically. For instance, we would not want a model that produces scenarios with negative interest rates.

After the underlying economic relationships are determined and modeled, we control the scenario output by the selection of model parameters, called *calibrating the model*, or *calibration*. Model parameters could include mean reversion level for interest rates, volatility for stock returns, and expected inflation growth. For simple models, standard statistical methods such as maximum likelihood estimation are appropriate. For complex models, we need to employ more sophisticated methods to determine the best parameters.

The calibration method described in this paper allows the user to specify characteristics the scenarios should have, referred to as targets. Each target represents a quality that the scenarios should exhibit, such as a range of bond returns over time, and an accompanying level of significance (weight). The targets can be based on historical analysis or some rational vision of future relationships. We then utilize an optimization procedure to determine best parameter settings to meet the targets.

This paper focuses on an economic scenario generator and the calibration process employed by American Re-Insurance Company headquartered in Princeton, NJ. In the next section, we briefly describe the entire DFA system, of which the scenario generation is one important piece. Section 3 focuses on the economic modeling system, the different types of economic models, and characteristics of a good model. In Section 4, we discuss how to set targets for the calibration process. Section 5 presents an example of the calibration process, utilizing software developed by Lattice Financial. Some final thoughts are in Section 6.

2. A Dynamic Financial Analysis System

American Re-Insurance Company's Risk Management System (ARMS) is an integrated compilation of models. The system is applied to determine internal capital allocation for the Company. The system is also used to assist both Munich Re[1] and American Re-Insurance Company clients in evaluating and setting up efficient re-insurance structures. The structure of the system is laid out in Figure 1.

[1] American Re-Insurance Company is a member of the Munich Re Group

ARMS Structure

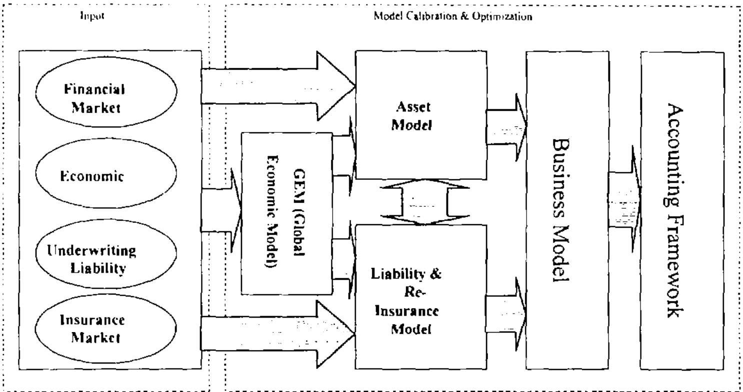


Figure 1. American Re-Insurance Company's Risk Management System (ARMS) is an integrated compilation of models. Historical data from financial and economic markets, underwriting decision processes, and insurance market trends are inputs to the system (left). Output includes balance sheet and income statements, optimal investment mixes and reinsurance structures.

The **Global Economic Model**[2] generates plausible time series outcomes of future economies based on user specifications and parameter settings. The user specifications are inputs reflecting the current economic environment and expectations for long-term median trends. The parameter settings are referred to as calibration parameters and those are set via the **Constraint Evaluator System**.

Each of the economic time series scenarios are fed to the **Asset Model** as well as the **Liability and Re-insurance Model**. These two models project different asset and liability classes along each economic scenario. It is important to recognize that the economic scenario generator lays the foundation for the calibration of the liability and asset models. Although the liability losses are based on fitted frequency and severity distributions (see Hogg & Klugman (1984), Panjer & Willmot (1992)), our analysis of loss data shows dependency on inflation for many lines of business. Therefore, inflation scenarios from the economic model define the trend in the prospective severity distributions over time. Similarly, the prospective premium is trended with inflation. Any discounting for future pricing purposes is based on output from the economic model. We consider thousands of scenarios for many years in the future and thus develop distributions for our underlying asset and liability returns in a multi-period environment.

The **Business Model** considers the underlying strategy of the business managers. It models the decisions we make as the business moves forward through time. For example, how will the

[2] Global Economic Model (GEM) is under development. At the time of writing and for the foreseeable future, the only country modeled is the United States.

business grow if gross margins are reduced by 10% next year? This also includes any change in asset allocation or in re-insurance structure.

The **Accounting Framework** refers not only to accounting but also to tax implications. There are several advantages to separating this functionality. They include the facilitation of operating in a multi-country (and therefor multi-regulatory) environment.

Wrapped around all this functionality is a non-convex optimization engine – the driving force behind the **Constraint Evaluator System**. Since each of these models must be calibrated in one form or another, access to a non-convex optimization system minimizes traditional trial and error attempts to ensure the reasonability of results. Ideally, we want to back-test the models with historical data and ensure optimal performance before we start modeling prospectively.

To better understand the calibration process, we will focus on calibrating the economic scenario generator. A description of the generator in the context of previous modeling efforts is in the next section.

3. Scenario Generator

3.1 What Makes A Good Scenario Generator?

Unfortunately, there are no agreed upon standards for scenario generation techniques. For some, the model must be a series of mathematical equations that are solved analytically (e.g., Black-Scholes option pricing model). Others have a more empirical approach, preferring to forecast future returns directly on current and past conditions (e.g., vector auto-regressive and kernel regression approaches).

The Global Economic Model (GEM) scenario generator strikes a balance between the two. Relationships among economic variables are modeled with explicit stochastic difference equations and the equation parameters are based on historical data via the calibration process[3]. The set of equations is too complex to have a closed form solution. Thus, Monte Carlo simulation is utilized to generate a multitude of paths (scenarios).

American Re-insurance defined the following criteria for the GEM system:

- Must be logically defensible – relationships among the economic variables must be consistent with economic theory and be statistically defensible given historical data.
- Must produce the proper relationships over time - movements in the economic variables must be reasonable across long time horizons and across different time steps. That is, the statistical properties of the factors must be consistent whether the model is run monthly, quarterly, or annually.

A good model must be able to capture risk both within and across time. This can be accomplished with a multi-period model. As a counter-example, the traditional Markowitz model is a one-period asset allocation model based on statistical observations of means, variances and correlations and as such, the Markowitz model does not address risk over time. One of the key

[3] We could calibrate for pricing purposes, but in our experience this does not generate reasonable results for future economies.

statistics for risk over time is serial correlation (sometimes referred to as auto-correlation) which is any time series correlation with itself lagged one (or more) time periods.

The Markowitz model also does not create a direct link between underlying economic variables and the asset model. Thus, the Markowitz model cannot consistently create an asset liability framework as there is no direct link between assets and liabilities. A more preferable approach is to build an underlying economic framework and then evaluate both assets and liabilities based on that framework. As an example, an increasing inflation environment will affect both equity markets and certain insurance liabilities.

Many interest rate models do not build a term structure per se, but rather build short-term rates and short-term forward rates. The forward rates imply a term structure at a given point in the future, and the term structure implied based on forward rates today can be viewed as the market's expectation of the future yield curve. However, this is not necessarily a good predictor or even estimator of future yield curves.

Brennan and Schwartz (1979) propose using stochastic differential equations to price bonds. They start with a model for short-term interest rates and long-term interest rates with some inter-dependencies. Based on these two models, they apply Ito's Lemma to derive the necessary structure of the stochastic equations to create a no-arbitrage condition. This is a pricing application.

While the approach we propose is similar in some regards, we do not solve algebraically to create no-arbitrage stochastic equations. Rather, we monitor the modeled results for reasonability and arbitrage opportunities. Clearly any model that creates persistent and significant arbitrage opportunities must be questioned.

Though the yield curve today is a poor predictor of future rates, it is reasonable to assume that the short-term rate will co-move to some extent with the long-term rate, as the long-term rate holds information about the future expected values of the short-term rate. Brennan-Schwarz captures this through a joint Gauss-Markov process and this reflects both the pure expectations hypothesis and the liquidity premium hypothesis. GEM utilizes a similar methodology - though employing it with forward rates rather than with yields or spot rates.

The Wilkie interest rate model breaks interest rates into two components, specifically a real interest rate, which tends to be fairly stable, and inflation, which can be quite volatile at times. Wilkie notes that equity dividend yields and inflation tend to be highly correlated. He views inflation as driving interest rates rather than the opposite. Note, that Brennan-Schwartz does not consider inflation or other indicators in a larger economic context.

Heath-Jarrow-Morton (Heath et al., 1990) has received much attention during the past few years. The HJM model is a more recent extension of the arbitrage-free pricing model. HJM cleverly extends the single factor (short interest rate only) to a multi-factor environment (two or three) but the complexity increases dramatically. In addition, just because the market expects a given term structure in the future does not by any means suggest that this is at all a reasonable estimator of the future. The market changes its expectations almost instantaneously and continuously. The HJM model is based on forward rates from which spot rates and yields can be derived. There are some advantages to basing a stochastic model (pricing or strategic) on forward rates. Namely, if a reasonable forward rate curve is modeled, it is likely that spot rates and yields look reasonable as well. The reverse is not true (Tilley, 1992).

The GEM system incorporates ideas from all of the above. In addition, we have complemented with our own analysis as shown in the pages that follow.

3.2 Types of Models

We distinguish between two types of asset modeling approaches. Pricing models are entirely based on the notion that any risk-free profit (above the risk-free rate – this is known as arbitrage) will be exploited in the market place until it no longer exists. The very nature of this action eliminates the risk-free profit. Pricing models generally work in the risk-neutral world, which is particularly useful for pricing liquid contingent options that can be replicated through other vehicles that are also liquid (and can be shorted). But the risk-neutral approach falls short when trying to determine reasonable returns for asset classes and interest rates in general over multiple time horizons. Specifically, the inherent assumptions that all asset classes return the risk-free rate[4] is not satisfactory for a risk management system where one at least should have the option to specify different risk premia for different asset classes. There are also practical implications in terms of "exploding" lattice models, which require a geometrically increasing number of branches with increasing number of time periods.

Strategic models consider an almost infinite series of possibilities. The more scenarios one creates through the Monte Carlo simulation, the more possibilities one can explore. These scenarios depict plausible paths for the future. Some paths have high equity returns, some have low returns. Some have rising interest rates. Some have falling interest rates. On average, the asset class returns reflect the risk-premiums associated with the economic environments under which they are modeled. There is no reason that this should be the risk-free rate – just like in the real world.

Pricing models give a pricing snap-shot at a point in time of certain contingent claims. Strategic models provide a view over time that can be used to design strategies that manage risk and return. The GEM system utilizes Monte Carlo simulations.

3.3 Global Economic Model

The Global Economic Model (GEM) is based on a series of stochastic difference equations. The difference equations have an underlying structure as graphed below (Figure 2). We adopt this structure as a way to capture the complex relationships that the real world offers.

The structure demonstrates how the model is developed within each time period. Although the time increments in the model are flexible, the default is monthly. Each month the system simulates values for each item in accordance with this structure.

We use stochastic differential equations to build our underlying framework. The examples in Figure 3 below show the most basic form of Brownian motion. The "dZ" is a Wiener process, which is generated from a standard normal distribution. "l_t" represents the long interest rate (for example, the one-period 30 year forward rate) and "l_μ" is the long-term equilibrium for l_t. "α_l" and "σ_l" are calibration parameters. They control the movement and overall volatility of the stochastic process. α_l is often referred to as the "mean reversion parameter", while σ_l is the

[4] Arguably one could replace the risk-neutral probabilities with "real-world" probabilities to generate "real-world" scenarios

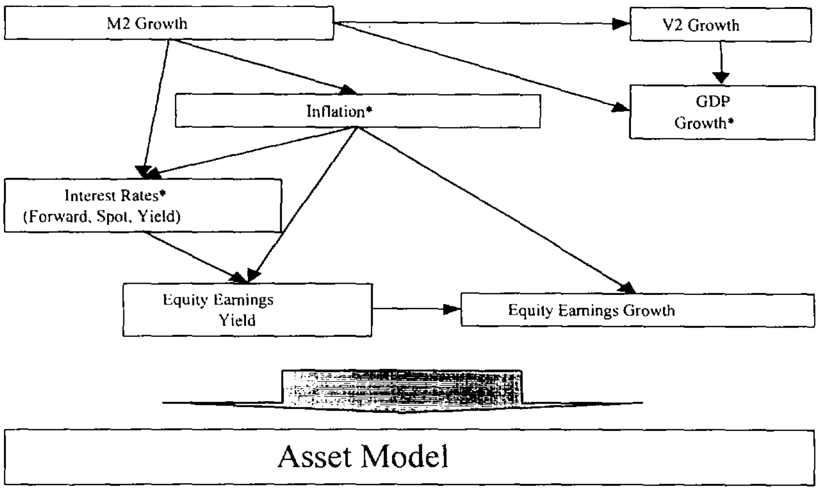


Figure 2. The economic framework underlying GEM. Each variable is modeled via stochastic difference equations. Asterisks indicate variables that could have links to other countries.

volatility parameter. Note, that we are modeling the difference from one time period to the next. This captures the basic notion that economic time series tend to exhibit significant serial correlation over time, while any change in the series tend to be more independently distributed. Similar observations apply to the short rate process shown below the long rate process.

Figure 3:

Long interest rates

$$dl_t = a_l(l_\mu - l_t)dt + l_t\sigma_l dZ_l$$

Short interest rates

$$dr_t = a_r(r_\mu - r_t)dt + r_t\sigma_r dZ_r$$

We normally start the process with the economic environment today (for prospective simulations). Specifically, we get l_0 and the rest of the starting yield curve from publicly available data. When calibrating the model (back-testing), however, we start the model in the economic environment that matches the data starting point.

Generally, we will define a stable long-term economic environment that looks very much like the current environment except for a change in the short interest rate to create a more normal looking yield curve. The normal yield curve spread is assumed to be 150 basis point (bp), and short real yields are assumed to be 200 bp. Based on this information, we develop our base line simulation ("base"). We calibrate to fit our targets to the base.

Once the base has been fitted, we change the long-term median assumptions. Clients will often want to explore the risk they are facing if the median environment differs from the one assumed. What happens if interest rates are most likely to increase over the next ten years? What if they are most likely to fall? We can explore all of these options separately or together, and we must ensure that the model holds up to these stress tests and still generates acceptable results (Mulvey and Madsen, 1999).

4. Setting Targets

Targets are properties we would like the generated scenarios to possess. The statistic is the actual value calculated from the scenarios. To fix the idea, a target could be the average value (across scenarios) for the annualized standard deviation of stock prices, such as 20%. The statistic would be the calculated average standard deviation of stock prices from the generated scenarios, which we would hope would be close to 20%. Our goal could be to have the statistic as close as possible to the target. Alternatively, we can specify a range of acceptable values and penalize statistics outside the target range.

Some targets we specify are:

- Arithmetic means
- Compound means
- Standard deviations
- Skewness and kurtosis (though we generally place less weight on these)
- Tails of non-normal distributions
- Minimum and maximum observations
- Correlations
- Serial correlations
- Yield curve statistics

David Becker of Lincoln National studied US interest rates (Becker, 1995). He used the period 1955 --1994 and made a number of interesting observations. Based on his observations, he developed a number of "stylized facts" that an interest rate model should possess:

- Rates are non-negative
- Rates do not go to zero nor do they go low and stay low
- Rates do not go to infinity nor do they go high and stay high
- Rates neither increase nor decrease rapidly with significant frequency
- Rates have the appearance of a random walk
- Rates have the appearance of mean reversion, i.e. when rates fall they rebound to "normal" levels, and similarly when rates rise
- Rates tend to cluster in trading ranges, or narrow bands, before breaking out to a higher or lower range
- Periodic movements in rates are not independent, but are correlated to a limited number of prior period movements
- Levels of serial correlation tend to decrease with maturity
- Short term and long term rates are highly correlated, but not perfectly correlated
- Generally, rates tend to rise and fall together. Thus, shifts in term structure are largely "parallel"

- Higher absolute interest rate levels are associated with higher absolute interest rate volatility
- Rate volatility declines with maturity
- Yield curve inversions:
  - Frequency: Less than 16% absolute
  - Infrequent and of limited duration
  - Occur during severe economic stress, geopolitical and/or policy volatility
- Yield spreads decrease with maturity, i.e. 1 year – 3 month spread > 3 year – 1 year spread and so on
- Correlation between increase in CPI and Treasuries declines as maturity increases
- *In general, as rates rise spreads narrow such that the yield curve flattens; and as rates fall, spreads widen such that the yield curve steepens*

We designed our model targets to capture these stylized facts as well as other calibration targets.

Cash tends to have a high serial correlation as does inflation, whereas stocks tend to have slightly negative serial correlation. Even these general observations, however, change over time as is illustrated by the example below.

Example of Target:

The correlation between long-term yields and inflation has ranged from –35% to 70% (Figure 4).

10 Year Correlation Between LT Yields and Inflation

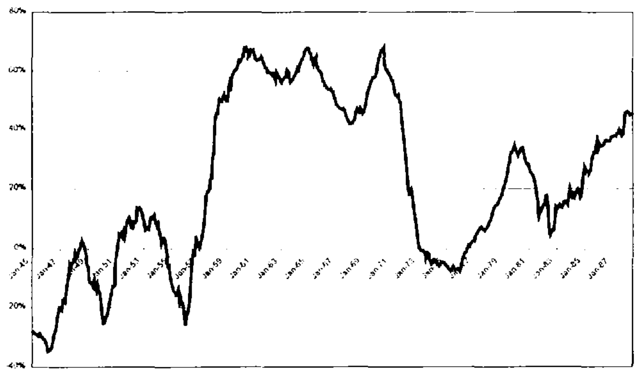


Figure 4. Historical correlation between long-term government bond yields and inflation.

How do we set a reasonable target based on this information? Our target becomes a distribution with an expected value of 30%-40%. We still create some paths with correlation of –40%, but they occur less frequently than paths with 30% correlation between the two variables (Figure 5).

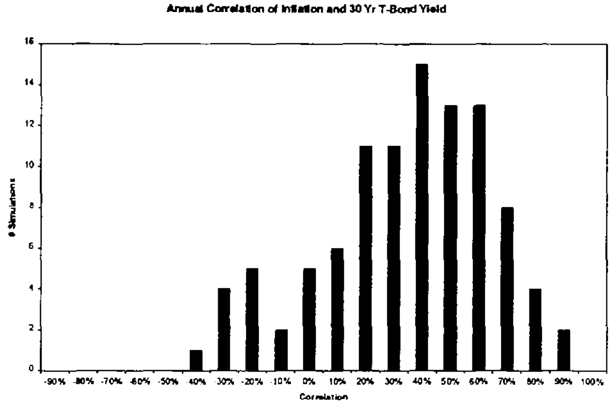


Figure 5. Simulated correlation between long-term government bond yields and inflation (distribution looks "choppy" as only 100 simulations were run).

5. Calibration Methodology

Calibration targets can be monthly, annual or any other time period. A penalty is assigned for each deviation from a target. The goal is to calibrate the model to minimize the assigned penalties. AmRe's Constraint Evaluator System is used in this process. The Constraint Evaluator System utilizes a non-convex optimizer developed by Lattice Financial. See Berger et al. (1998) for an algorithm overview and Berger (1999) for technical information.

Model parameters are set to initial values using linear multiple regression. We take historical data, set up the difference equation, perform the regression, and utilize the results as the starting point for the analysis.

Calibration Example #1:

$$\Delta f_t = A \cdot \Delta I_t + B \cdot \Delta Y_t + C \cdot \Delta u_t + D \cdot \sqrt{t} \cdot \Delta Z$$ [5]

Here f represents the 3-month one-period forward rate, I represents inflation, Y represents the 30 year one-period forward rate, u represents the inflation adjusted mean reversion process, and dZ is Wiener term and t is time unit. A, B, C, and D are the calibration parameters for this difference equation. A controls the effect inflation has on the 3 month forward rate and B controls the relationship with the long end of the forward rate curve. C controls the rate of reversion, while D reflects the volatility added to the stochastic process.

Regressing this on monthly historical data from 1974 through 1998 (Figure 6), we get {A, B, C} = {0.015, 1.3, -0.015}. All parameters have significant t-statistics with 90%

[5] The difference equation offered here is actually a two-part log-linear process.

confidence, and the R^2 is 58%. D is added after reviewing the residual standard error, which is 0.004. The ratio of the residual standard error to the mean is 0.06. Since this is a log-linear process, D is 1.06.

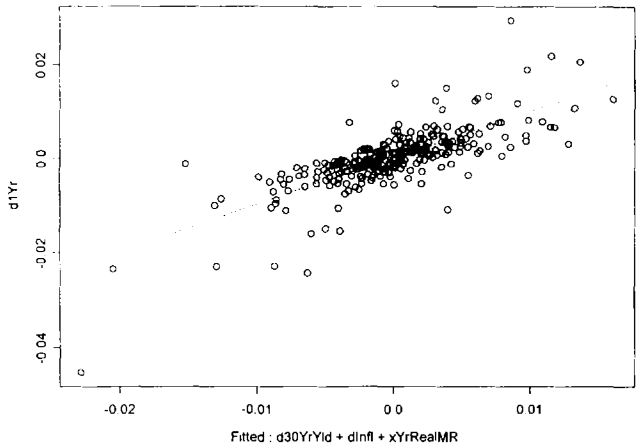


Figure 6. Linear regression results (using historical data) of regressing the change in 3 month yield versus the change in 30 year yield, change in monthly inflation rate, and real inflation-adjusted mean reversion

Now we need to incorporate the regressed results with our simulation goals and simulated data. We have a number of criteria that we monitor with respect to the generated time series. Is serial correlation high enough for shorter term yields? Are we generating a reasonable number of recessions? Are recessions characterized by both inverted yield curves and drops in real GDP? The list goes on to include basic statistics of the modeled indicators.

We code our targets and perform the following optimization described below. Notice that each time series depends on the calibration vector. Specifically, changing the values of {A, B, C, D} will give us different time series, as the difference equations change. We use our regression as a starting point and we want the calibration vector that comes closest to our targets. We run the following optimization:

$$Minimize \sum_{i=1}^{Scenarios} \sum_{j=1}^{Targets} w_j \cdot \left(Statistic_{i,j} - Target_j\right)^2$$

In this case, the result is {0.75, 0.5, -0.04, 1.05}, and we utilize these new values to generate the economic scenarios. The main vector changes were:

- shift weight from the 30 year rate to inflation to increase the correlation between inflation and the 3 month treasury bill

- increase the level of inflation-adjusted mean reversion to avoid "run-away" scenarios (tails were overstated using regression scenarios)
- decrease volatility slightly

If we had wanted to maintain a closer correspondence with the historical regression parameters, we could have penalized deviations from our initial calculated values. In this example, we were more concerned with matching our other calibration targets.

Calibration Example #2:

The optimization (minimize penalties by changing the calibration parameter set – see equation above) can be reviewed from other perspectives as well. We take a closer look at inflation in the calibration. The starting vectors (except for monetary growth, which is at the top of the structure – Figure 2) are all based on linear regressions using historical data. In this particular case, we can see from the chart below (Figure 7) that the volatility of inflation associated with our starting calibration parameters is understated compared with the historical data.

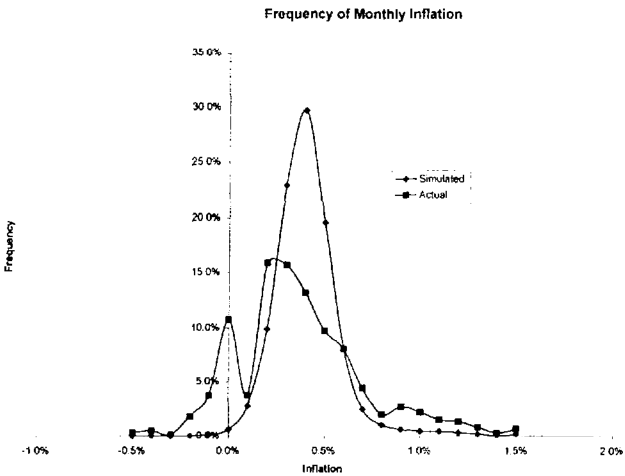


Figure 7. Distribution of monthly inflation levels in generated scenarios based on regression vector has a higher mean and tighter range compared with historical observations from 1974 through 1998.

The differences between simulated and historical results are due to a number of factors. There are sources of variation that are not represented in the regressed data. In addition, estimating the error term from the regression in terms of difference equations is often tricky. Further, statistics such as serial correlation is not monitored through regression, and the relationship may not be perfectly linear. In the graph above, we note that the tails based on historical inflation are much wider.

To address this discrepancy, we specify the volatility of inflation as a calibration target. The historical volatility is 0.33% (3.2% annually) and the volatility from the simulated scenarios above is much less. We specify the historical volatility of 0.33% as a target for the optimization. After optimizing, the resulting inflation levels are shown in Figure 8. The distribution is now much closer to the historical observations. Note that we were able to accomplish this by specifying only one parameter of the distribution (volatility). If this still did not produce the desired results, or if we wish to match more closely, we could specify quantiles on the distribution as targets.

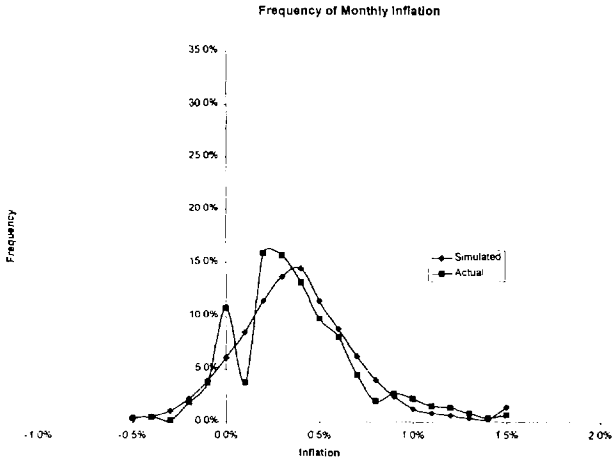


Figure 8. Distribution of monthly inflation levels in generated scenarios based on optimized vector now matches historical observations.

The optimizer helps us fit our model to the available data. Thus, we are able to maintain our economic framework, which is consistently applied to our loss simulation and our asset simulations. We are simultaneously capturing data we would otherwise only be able to capture with more limited methodologies.

In practice, we work with up to 245 calibration parameters for the US model though approximately 50 parameters capture the main process. Optimizing on all these at once has not been practical. Rather, we work our way down the structure shown in Figure 2. We initially calibrate the parameters associated with monetary growth and velocity. Then we calibrate inflation and so on.

6. Conclusion

In this paper, we have discussed the scenario generation component of a dynamic financial analysis system. The goal is to produce coherent and comprehensive scenarios for use in modeling an insurance company's financial position over time. American Re-Insurance's GEM system is an example of a generator grounded in economic theory, but one which produces scenarios consistent with historical observations. The calibration process is the mechanism for achieving this: Model parameters are chosen so that the generated scenarios have statistics consistent with user-specified targets. Lattice Financial's optimization software automates the process of determining the best model parameters to meet the desired targets.

References

Becker, D., "The Frequency of Inversions of the Yield Curve and Historical Data on the Volatility and Level of Interest Rates", *Risks and Rewards*, p. 6-9, 1991.

Becker, D., "Stylized Historical Facts Regarding Treasury Interest Rates from 1955 to 1994", Technical Report, Lincoln National, 1995.

Berger, A.J., "Lattice Financial's Dynamic Optimization Tools - Technical Overview," Lattice Financial Technical Report, 1999.

Berger, A.J. and J.M. Mulvey, "An Asset and Liability Management System for Individual Investors," in *Worldwide Asset and Liability Modeling*, (eds., W.T. Ziemba and J.M. Mulvey), Cambridge University Press, 1998.

Berger, A.J., J.M. Mulvey, K. Nish, and R. Rush, "A Portfolio Management System for Catastrophe Property Liabilities," in *Casualty Actuarial Society Forum*, Summer 1998.

Brennan, M.J. and E.S.Schwartz, "An Equilibrium Model of Bond Pricing and a Test of Market Efficiency," *Journal of Financial and Quantitative Analysis*, 17, 3, September 1982, p. 301 – 329.

Davlin, M.F., M.S. Tenney, "The Empirical and Theoretical Foundations for Interest Rates Models", SOA Seminar, 1996.

Daykin, C.D., T. Pentikainen, and M. Pesonen, *Practical Risk Theory for Actuaries* (Monographs on Statistics and Applied Probability), Chapman & Hall, 1994.

Dempster, M.A.H. and E. Thorlacius, "Stochastic Simulation of International Economic Variables and Asset Returns: The Falcon Asset Model," Swiss Reinsurance Company Technical Report, 1998.

Heath, D., R.A. Jarrow, and A. Morton, "Contingent Claim Valuation with a Random Evolution of Interest Rates," *The Review of Futures Markets* 9 (1) 1990, p. 54-76.

Hogg, S.A. and S.A. Klugman, *Loss Distributions*, Wiley, 1984.

Madsen, C.K., "American Re-Insurance Company's Global Economic Model", American Re-Insurance Company, 1999.

Madsen, C.K., J. Pardoe and A. Rippert, "Dynamic Financial Analysis", *Eagle News and Views*, December, 1998.

Markowitz, H.M., *Portfolio Selection*, John Wiley & Sons, 1951.

Mulvey, J.M. and C.K. Madsen, "Strategic Decision Making for Pension Plans Under Stress," Princeton University SOR Report, 1999.

Mulvey, J.M., C.K. Madsen and F. Morin, "Linking Strategic and Tactical Planning Systems for Dynamic Financial Analysis", *Casualty Actuarial Society Forum*, Summer 1998.

Mulvey, J.M. and E. Thorlacius, "The Towers Perrin Global Capital Market Scenario Generation System," in *Worldwide Asset and Liability Modeling*, (eds., W.T. Ziemba and J.M. Mulvey), Cambridge University Press, 1998.

Panjer, H.H. and G.E. Willmot, *Insurance Risk Models*, 1992.

Tilley, J.A., "Building Stochastic Interest Rates Generators", *SOA Transactions*, Volume XLIV, 1992.

Wilkie, A.D., "A Stochastic Investment Model for Actuarial Use", *Transactions of the Faculty of Actuaries*, 39, p. 391-403.

Estimating Uncertainty in Cash Flow Projections

Roger M. Hayne, FCAS, MAAA

ESTIMATING UNCERTAINTY IN CASH FLOW PROJECTIONS

by

Roger M. Hayne

Abstract

In order to be complete dynamic financial analysis (DFA) models should deal with both the amount and timing of future loss and loss adjustment expense payments. Even more than asset cash flows, these future payments are very uncertain.

This paper begins by estimating both process and parameter uncertainty in reserves for annuity-type benefits such as available in some automobile no-fault states or in workers compensation. Arguably, such reserves have underlying distributions (inherent in the mortality models) that may be more easily understood and treated than many other casualty coverages. We explore the estimation of both process and parameter uncertainty for this example. In the process we derive formulae that can be used to model uncertainty in other applications, once the various parameters are estimated. Many of the estimation methods covered should generalize to non-annuity applications.

There is also a companion of this paper, titled "Modeling Parameter Uncertainty in Cash Flow Projections" that provides motivation for the estimates contained in this paper. In that paper we discuss approaches to modeling future cash flows and argue for separation of parameter and process uncertainty as well as describing methods to model them both.

Biography

Roger is a Fellow of the Casualty Actuarial Society, a Member of the American Academy of Actuaries, and Consulting Actuary in the Pasadena, California office of Milliman & Robertson, Inc. with over twenty-one years of casualty actuarial consulting experience. Roger is a frequent speaker on reserve and DFA related topics and has authored several papers dealing with considerations and estimates of uncertainty in reserve projections. Roger is currently the chair of the CAS Research Policy and Management Committee and has served as chair of both the CAS Committee on Theory of Risk and the CAS/AAA Joint Committee on the Casualty Loss Reserve Seminar.

ESTIMATING UNCERTAINTY IN CASH FLOW PROJECTIONS

1. Introduction

There have been a number of papers and articles dealing with uncertainty in loss reserve estimates. However, dynamic financial analysis for risk bearing entities requires more than simply the distribution of reserves. Also of critical importance is the timing of those future payments and their distribution.

A simple example may clarify the point. Suppose two insurers, *Short Tail Insurance Company* and *Long Tail Insurance Company* are identical in all aspects except for the timing of future payments. Both companies are in runoff, both have $1 million in assets invested in the bank yielding 3% interest, and both will settle all losses in a single payment according to the following distribution:

Table 1: Hypothetical Distribution of Payments

Probability		Amount
20%	$	500,000
20%		750,000
20%		1,000,000
20%		1,250,000
20%		1,500,000

The only difference is that *Long Tail* will not pay this amount for 10 years, while *Short Tail* must pay it at the end of this year. Even though both insurers have the same assets and face the same distribution of reserves, *Short Tail* would face insolvency 40% of the time while *Long Tail* will only be insolvent 20% of the time (since 1,000,000 x 1.03 = 1,030,000 and 1,000,000 x 1.03^{10}=1,343,916). Though timing may not be everything, it is substantial.

Thus knowing the distribution of the reserves is necessary to model the financial condition of a risk bearing entity, but it is not sufficient. Rather, to appropriately model the future cash flows we need to know the distribution of payments in each future year.

In addition, economic conditions and unanticipated changes in cost inflation often impact reserves and contribute to the variability in both reserves and future payments as well as on assets. Thus, in dynamic financial analysis (DFA) applications where economic assumptions may be used as a "linkage" between asset and liability models, it will probably be necessary to separate the contributions of these economic factors from others in modeling liabilities.

In this paper we will begin with an example of how estimates of the means and variances of payment distributions by year can be made. This first example will focus on claims involving lifetime payments, such as for certain workers compensation claims or unlimited no-fault medical claims. Unlike many casualty claims, the fact that payments are contingent on survival actually provides us with an underlying probability structure for the payments on individual claims and makes discussion of many of the topics we will address more accessible. However, unlike many life coverages, the future payments are contingent not only on the claimant's survival, but on uncertain future costs.

We will then consider how to carry these concepts over to other coverages. These concepts also can be useful in constructing models for use in dynamic financial analysis.

2. *A Relatively Simple Example*

Suppose our insurer only has a fixed book of life pension workers' compensation indemnity claims and does not need to fund for the medical portion of these losses.

Further, to keep this first example relatively simple, we also assume:

2.1 We have mortality tables that appropriately reflect survival probabilities for these claimants.

2.2 There is no escalation of benefits for individual claimants due to inflation or some other index.

2.3 Future annual payments for each claimant are fixed and known.

2.4 We are not currently interested in the time value of money (i.e. no discounting).

2.5 The various claimants are statistically independent.

Here the expected future payments for any individual claim can easily be calculated using a life annuity. Not only can we use the mortality tables to obtain expected costs, but we can also use them to review the expected distribution of payments for our population in any particular future year.

To see this we let:

a_{xt} denote the payment for claimant x in year t in current dollars,

p_{xt} denote the probability that claimant x lives for t years and then dies or otherwise exits the claim population.

It is easy to see the distribution of payments in any future year s is given by:

Table 2: Payment Distribution for a Single Claim

Probability	Amount
$\sum_{t=s}^{\infty} p_{xt}$	a_{xs}
$1-\sum_{t=s}^{\infty} p_{xt}$	0

From this it is easy to see the payments in year s, have expected value

(2.1) $$E(X_s) = a_{xs}\sum_{t=s}^{\infty} p_{xt}$$

and variance

(2.2) $$\begin{aligned} Var(X_s) &= E(X_s^2) - E(X_s)^2 \\ &= a_{xs}^2\sum_{t=s}^{\infty} p_{xt} - a_{xs}^2\left(\sum_{t=s}^{\infty} p_{xt}\right)^2 \\ &= a_{xs}^2\left(\sum_{t=s}^{\infty} p_{xt}\right)\left(1-\sum_{t=s}^{\infty} p_{xt}\right) \end{aligned}$$

This is the result we would expect from the binomial distribution for the payments in year s.

In addition, from our assumptions we see that the future payments for this claimant will have a discrete distribution with payments totaling $\sum_{s=1}^{t} a_{xs}$, occurring with probability p_{xt}. Thus the total expected future payment for this claimant is given by:

$$(2.3) \qquad E(X) = \sum_{t=1}^{\infty} p_{xt} \sum_{s=1}^{t} a_{xs} = \sum_{s=1}^{\infty} a_{xs} \sum_{t=s}^{\infty} p_{xt}$$

The second is simply the total expected payments in each future year.

Similarly we can also calculate the variance.

$$(2.4) \qquad Var(X) = \sum_{s=1}^{\infty} \sum_{r=1}^{\infty} a_{xs} a_{xr} \left(\sum_{t=\max(r,s)}^{\infty} p_{xt} \right) \left(1 - \left(\sum_{t=\min(r,s)}^{\infty} p_{xt} \right) \right)$$

Although this formula may not be immediately obvious it is not difficult to derive. We show the derivation in Appendix A.

Thus for a single claimant we can easily obtain the distribution of future payments, its mean and variance as well as the distribution of payments in any future year. We can still explicitly determine the distributions for multiple claimants, however, the calculations become more complex (such calculations may be necessary if, for example, reinsurance attaches on a per incident not per claimant level). For example, for two independent claimants, x and y, the payments in year s have the following discrete distribution:

Table 3: Payment Distribution for Two Claims

Probability	Amount
$\left(\sum_{t=s}^{\infty} p_{xt} \right) \left(\sum_{t=s}^{\infty} p_{yt} \right)$	$a_{xs} + a_{ys}$
$\left(\sum_{t=s}^{\infty} p_{xt} \right) \left(1 - \sum_{t=s}^{\infty} p_{yt} \right)$	a_{xs}
$\left(1 - \sum_{t=s}^{\infty} p_{xt} \right) \left(\sum_{t=s}^{\infty} p_{yt} \right)$	a_{ys}
$\left(1 - \sum_{t=s}^{\infty} p_{xt} \right) \left(1 - \sum_{t=s}^{\infty} p_{yt} \right)$	0

We could derive a similar table for the distribution of total future payments for two claimants. Rather than having simply four separate points, the resulting table would have *nxm* points where *n* denotes the number of future years having non-zero probabilities for claimant *x* and *m* the number for claimant *y*. Although we can exactly calculate the resulting distributions for many claimants, the resulting exponential growth in size makes such calculations prohibitive.

On a practical level, however, the problem of combining two distributions is simply one of calculating the aggregate loss distribution for two distributions. Heckman & Meyers[1] provide one means of performing these calculations, Robertson[2] gives another.

We can also approximate the aggregate distribution of the discrete distributions iteratively. We first calculate the aggregate distribution of two distributions exactly, resulting in *mxn* cells. We then compress this large distribution to, say, *m* cells and repeat the process with the next distribution. Straightforward combination of cells will usually result in a reduction in the variance in the final distribution while maintaining the mean. The following is an example of this approach.

Consider the two distributions:

Table 4: Distributions for Convolution Example

Variable 1		Variable 2	
Probability	Amount	Probability	Amount
0.60	100	0.20	250
0.40	300	0.80	500

The resulting aggregate distribution is:

Table 5: Distribution of the Sum of Variables

Probability	Amount
0.12	350
0.08	550
0.48	600
0.32	800

A possible compression of this aggregate distribution is:

Table 6: Collapsed Distribution of Sum

Probability	Amount
0.20	430
0.80	680

Here 0.20=0.12+.0.08, 430=(0.12x350+0.08x550)/0.20, and so forth. Note the expected value of 630 is preserved in the compressed distribution but the variance of the exact distribution is 22,240 while that of the compressed distribution is 10,000. There is some flexibility in this method, however, in that the algorithm used to combine the cells could take into account the purpose of the modeling. For example, if the interest is in probabilities of high loss amounts, then we could maintain more detail in the "tail" of the distribution by combining more cells with smaller loss amounts with less combination of higher loss cells. In the above example, the following is another compression:

Table 7: Alternative Collapsed Distribution

Probability	Amount
0.68	550
0.32	800

The mean is again preserved but the variance is now 13,600, closer to that of the exact distribution.

Another possible approximation would be to assume that the aggregate distribution follows a smooth distribution with a limited number of parameters. We could then "back into" the aggregate distribution making use of moments of the true aggregate distribution. For this, however, we need to be able to calculate those moments. For our simple example, however, the calculations follow very simply from (2.3) and (2.4) if we assume that individual claims are independent from one another. Given the fact that the distributions are based on survival probabilities, and our assumption that the probabilities themselves are correct, this is probably not too restrictive in practice.

In this case, letting T denote the random variable corresponding to the aggregate distribution, we see that, assuming we have N claims, the expected aggregate loss is given by:

$$
\begin{aligned}
\mathrm{E}(T) &= \mathrm{E}\left(\sum_{i=1}^{N} X_i\right) \\
&= \sum_{i=1}^{N} \mathrm{E}(X_i) \qquad (2.5) \\
&= \sum_{i=1}^{N} \sum_{s=1}^{\infty} a_{x,s} \sum_{t=s}^{\infty} p_{x,t}
\end{aligned}
$$

Similarly, because we assumed the claims are independent, we can calculate the variance for the aggregate distribution as:

$$
\begin{aligned}
\mathrm{Var}(T) &= \mathrm{Var}\left(\sum_{i=1}^{N} X_i\right) \\
&= \sum_{i=1}^{N} \mathrm{Var}(X_i) \qquad (2.6) \\
&= \sum_{i=1}^{N} \sum_{s=1}^{\infty} \sum_{r=1}^{\infty} a_{x,s} a_{x,r} \left(\sum_{t=\max(r,s)}^{\infty} p_{x,t}\right)\left(1-\left(\sum_{t=\min(r,s)}^{\infty} p_{x,t}\right)\right)
\end{aligned}
$$

Similar calculations based on (2.1) and (2.2) will give us the mean and variance of the total expected annual payments:

$$
\begin{aligned}
\mathrm{E}(T_s) &= \mathrm{E}\left(\sum_{i=1}^{N} X_{si}\right) \\
&= \sum_{i=1}^{N} \mathrm{E}(X_{si}) \qquad (2.7) \\
&= \sum_{i=1}^{N} a_{x,s} \sum_{t=s}^{r} p_{x,t}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var}(T_s) &= \mathrm{Var}\left(\sum_{i=1}^{N} X_{si}\right) \\
&= \sum_{i=1}^{N} \mathrm{Var}(X_{si}) \qquad (2.8) \\
&= \sum_{i=1}^{N} a_{x,s}^{2} \left(\sum_{t=s}^{\infty} p_{x,t}\right)\left(1-\sum_{t=s}^{\infty} p_{x,t}\right)
\end{aligned}
$$

We note we can calculate the exact distribution for payments in any particular year as with the aggregate distribution for the total. However, in this case, there will "only" be 2^N cells in the distribution. Again, we could use a compression algorithm to obtain approximate distributions.

3. Introducing Some Uncertainty

The problem thus far considers only random fluctuations due to the fact that the exact time of exit from the claimant population is unknown. We have assumed that all other aspects of the problem are known. In short, we have only discussed process uncertainty thus far, i.e., that uncertainty remaining in the situation even if the process itself is known with certainty.

In the real world models used are generally approximations of the underlying process, subject to uncertainty either in their parameters or even whether or not they are appropriate. In this section we begin to introduce uncertainty into the assumptions from section 2.

The first restriction we will relax will be the assumption that underlying survival probabilities for individual claimants are known. In reality payments will often be contingent on the survival of an individual who is already injured and whose injuries may significantly impair chances for continued survival. Thus it may not be appropriate to use standard mortality tables to determine the survival probabilities. It is possible that the tables that are used will be modified or based in some way on populations of injured claimants and thus subject to estimation error.

In addition, it is possible that a claimant will sufficiently recover from his or her injuries so as not to require additional payments from the insurer. Thus exit from the population could occur for reasons other than death. We may need additional modeling to study the effects of such recoveries on exits from the population by claimants.

Since most such analyses focus on the mortality in a year, we let

q_{xt} denote the probability that claimant x will die in year t, given survival through year $t-1$.

These are the standard mortality probabilities. In terms of the p_{xt} variables defined above we have (possibly mixing notation somewhat):

$$p_{xt} = q_{x+t}\prod_{i=0}^{t-1}(1-q_{x+i})$$

(3.1)
$$= (1-(1-q_{x+t}))\prod_{i=0}^{t-1}(1-q_{x+i})$$
$$= \prod_{i=0}^{t-1}(1-q_{x+i}) - \prod_{i=0}^{t}(1-q_{x+i})$$

Very conveniently, these collapse in the sum to yield:

$$\sum_{t=m}^{\infty} p_{xt} = \sum_{t=m}^{\infty}\left(\prod_{i=0}^{t-1}(1-q_{x+i}) - \prod_{i=0}^{t}(1-q_{x+i})\right)$$

(3.2)
$$= \prod_{i=0}^{m-1}(1-q_{x+i}) - \prod_{i=0}^{\infty}(1-q_{x+i})$$
$$= \prod_{i=0}^{m-1}(1-q_{x+i})$$

In addition to allowing uncertainty in the survival probabilities we will also allow the annual benefits to change over time with economic conditions and allow for discounting of the reserves, as would be the case for the medical portion of workers' compensation or certain automobile no-fault benefits. We will allow the combined economic effect of inflation and discounting to be uncertain. Finally we will allow for some uncertainty in the annual payment estimates for individual claimants. Specifically we will relax our various assumptions to the following:

3.1 The relative survival probabilities among various claimants are known, however, the absolute probabilities are based on an analysis of n exposures. Analytically, we assume that there is a random variable y and constants q_{xt}^{*}, such that for all x and t values:

(3.3)
$$1-q_{xt} = (1-q_{xt}^{*})y$$

3.2 The a_{xt} values are stated in current dollars. There is escalation in those amounts between time *t-1* and time *t* in the amount of $1+f_t$. This escalation will be the same for all claimants but may vary from year to year. The $1+f_t$ amounts are not known with certainty.

3.3 The present value of 1 at time *t-1* is $1+v_t$ at time *t*. The $1+v_t$ amounts are not known with certainty.

3.4 There is a random variable *u* and constants a^*_{xt} such that for all claimants *x* and time *t*, the following holds:

$$a_{xt} = a^*_{xt}u \tag{3.4}$$

3.5 The various claimants are statistically independent.

3.6 There are random variables w_t and constants f_t^*, and v_t^* such that, for all *t* values:

$$\frac{1+f_t}{1+v_t} = \frac{1+f_t^*}{1+v_t^*}w_t \tag{3.5}$$

The variable *y* in 3.1 could be considered as a global load, reflecting the uncertainty in estimating the overall closure rate from experience. We recognize that this does not consider the uncertainty regarding the relative closure probabilities. For example, it is likely that younger claimants will experience a greater reduction in survival chances due to the injury causing the claim than older claimants will. Thus, except in the simplest situations, the variable *y* probably should not be considered as a mortality load, but rather a global uncertainty parameter.

We can estimate the degree of uncertainty arising from the sample size of *n* life-years used to estimate the survival or closure probabilities. For this we use sample theory and an application of Bayes' Theorem. In fact, if we assume:

1. The random variable *y* has a binomial distribution with expected value θ.
2. The random variable θ itself has a uniform distribution between 0 and 1 (i.e. we have no prior knowledge of the appropriate value of θ).
3. Our sample size is *n*.
4. We observe *z* claims remaining open after one year from our sample.

If we make the more general assumption in 2 above that θ has a beta distribution with parameters α and β it turns out that θ given the observations has a beta distribution with parameters $z+\alpha$ and $n-z+\beta$. We show this in Appendix B. In particular, then,

$$
\begin{aligned}
\mathrm{E}(\theta^r|z) &= \int_0^1 \frac{\Gamma(\alpha+\beta+n)}{\Gamma(z+\alpha)\Gamma(n-z+\beta)}\theta^{z+r+\alpha-1}(1-\theta)^{n-z+\beta-1} \\
&= \frac{\Gamma(\alpha+\beta+n)}{\Gamma(z+\alpha)\Gamma(n-z+\beta)}\int_0^1 \theta^{z+r+\alpha-1}(1-\theta)^{n-z+\beta-1} \\
&= \frac{\Gamma(\alpha+\beta+n)}{\Gamma(z+\alpha)\Gamma(n-z+\beta)}\frac{\Gamma(z+r+\alpha)\Gamma(n-z+\beta)}{\Gamma(\alpha+r+\beta+n)} \\
&= \frac{\Gamma(\alpha+\beta+n)\Gamma(z+r+\alpha)}{\Gamma(z+\alpha)\Gamma(\alpha+r+\beta+n)}
\end{aligned} \tag{3.6}
$$

Thus, in particular,

$$
\begin{aligned}
\mathrm{E}(\theta|z) &= \frac{\Gamma(\alpha+\beta+n)\Gamma(z+1+\alpha)}{\Gamma(z+\alpha)\Gamma(\alpha+1+\beta+n)} \\
&= \frac{\Gamma(\alpha+\beta+n)\Gamma(z+\alpha)(z+\alpha)}{\Gamma(z+\alpha)\Gamma(\alpha+\beta+n)(\alpha+\beta+n)} \\
&= \frac{z+\alpha}{\alpha+\beta+n}
\end{aligned} \tag{3.7}
$$

Thus we have:

$$
\begin{aligned}
\mathrm{E}\left(\left(\frac{\theta}{\mathrm{E}(\theta)}\right)^r\right) &= \frac{\mathrm{E}(\theta^r|z)}{\mathrm{E}(\theta|z)^r} \\
&= \left(\frac{\alpha+\beta+n}{z+\alpha}\right)^r \frac{\Gamma(\alpha+\beta+n)\Gamma(z+r+\alpha)}{\Gamma(z+\alpha)\Gamma(\alpha+r+\beta+n)}
\end{aligned} \tag{3.8}
$$

Now, the special case we will consider is no preference in the prior distribution for θ. This is simply a special case of the beta distribution with $\alpha = \beta = 1$. In this case we have:

$$E(y^r) = E\left(\left(\frac{\theta}{E(\theta)}\right)^r\right)$$

(3.9)
$$= \left(\frac{n+2}{z+1}\right)^r \frac{\Gamma(n+2)\Gamma(z+r+1)}{\Gamma(z+1)\Gamma(r+n+2)}$$

$$= \prod_{i=0}^{r-1} \frac{(n+2)(z+i+1)}{(z+1)(n+i+2)}$$

The last equation follows from the recursive properties of the gamma function and makes calculation easier in practice. In terms of the survival probabilities we have:

$$E_\theta\left(\sum_{t=m}^{\infty} p_{xt}\right) = E_\theta\left(\prod_{i=0}^{m-1}(1-q_{x+i})\right)$$

$$= E_\theta\left(\prod_{i=0}^{m-1}(1-q^{*}_{x+i})y\right)$$

$$= E_\theta\left(y^m \prod_{i=0}^{m-1}(1-q^{*}_{x+i})\right)$$

(3.10)
$$= E_\theta(y^m)\prod_{i=0}^{m-1}(1-q^{*}_{x+i})$$

$$= E_\theta\left(\left(\frac{\theta}{E_\theta(\theta)}\right)^m\right)\prod_{i=0}^{m-1}(1-q^{*}_{x+i})$$

$$= \left(\prod_{i=0}^{r-1} \frac{(n+2)(z+i+1)}{(z+1)(n+i+2)}\right)\prod_{i=0}^{m-1}(1-q^{*}_{x+i})$$

As one would expect, the first term in the last product tends to unity as the sample size n becomes large if

(3.11)
$$\lim_{n\to\infty} \frac{z}{n} = \theta$$

for some value θ. The proof is shown in Appendix B.

Assumptions 3.2 and 3.4 deal with cost escalation and discounting and 3.6 relates the two. We assume that the combined impact of inflation and discounting is uncertain with the variables w_t providing that uncertainty.

Finally we will modify the assumption that all future payments (at current cost levels) are known to one wherein there is "global" uncertainty regarding future payments. This is reflected in the variable u.

For simplicity we will assume that the variables w_t and u all have independent lognormal distributions, and that the distribution for the various w_t have the same means and variances. In particular we will assume that all these variables are independent and:

$$u \sim \text{lognormal}\left(-\tfrac{1}{2}\sigma^2, \sigma^2\right) \text{ and}$$
$$w_t \sim \text{lognormal}\left(-\tfrac{1}{2}\tau^2, \tau^2\right) \text{ for all t.} \tag{3.12}$$

Here and throughout this paper we will use the normal-transformed parameterization of the lognormal distribution. For example, (3.12) assumes that the normal variable ln u has a normal distribution with mean $-\tfrac{1}{2}\sigma^2$ and variance σ^2. More generally when we say

$$x \sim \text{lognormal}\left(\mu, \sigma^2\right) \tag{3.13}$$

we mean that the random variable x has the probability density function

$$f(x) = \frac{\exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right)}{x\sigma\sqrt{2\pi}} \tag{3.14}$$

With this parameterization, then we have:

$$E(X) = \exp\left(\mu + \tfrac{1}{2}\sigma^2\right)$$
$$\text{Var}(X) = \exp\left(2\mu + \sigma^2\right)\left(\exp\left(\sigma^2\right) - 1\right) \tag{3.15}$$
$$c.v.(X) = \sqrt{\frac{\text{Var}(X)}{E(X)^2}} = \sqrt{\exp\left(\sigma^2\right) - 1}$$

This last relationship shows that, with this parameterization, the coefficient of variation (ratio of standard deviation to the mean) depends only on the σ^2 parameter.

It could be argued quite convincingly that u would not be the same for all claimants or for all years. That is clearly a refinement to the methodology we present here. However, to keep the calculations to a manageable level, we have elected to make this simplifying assumption here. However, the assumption of lognormality for the economic variables is probably much more plausible, although the assumption of constant variance may be somewhat restrictive. In both cases, here, we note that the expected values of both distributions are unity, that is both u and the w_t variables are assumed to represent random shocks to our overall expectations.

We are now ready to calculate the mean and variance of the total population reserve. The calculation makes repeated applications of the following relationships that hold for independent conditional distributions:

$$
\begin{aligned}
\mathrm{E}(Z) &= \mathrm{E}_{\xi}\left(\mathrm{E}(Z|\xi)\right) \\
\mathrm{Var}(Z) &= \mathrm{E}_{\xi}\left(\mathrm{Var}(Z|\xi)\right) + \mathrm{Var}_{\xi}\left(\mathrm{E}(Z|\xi)\right)
\end{aligned}
\tag{3.16}
$$

In this case we assume that the distribution of the random variable Z with probability density function $f(z,\xi)$ that depends on a parameter ξ which itself is a random variable with probability density function $g(\xi)$. These assumptions result in the following formulae for the mean and variance of the total distribution:

$$
\begin{aligned}
\mathrm{E}(T) &= \sum_x \sum_{s=1}^{\infty} b^{*}_{xs} \left(\prod_{t=0}^{s-1} (1 - q^{*}_{xt}) \right) \prod_{i=0}^{s-1} \frac{(n+2)(z+i+1)}{(z+1)(n+i+2)} \\
\mathrm{Var}(T) &= \left(\sum_x \sum_{s=1}^{\infty} b^{*}_{xs} \left(\prod_{t=0}^{s-1} (1 - q^{*}_{xt}) \right) \mathrm{E}(y^s) \right)^2 \left(\exp(\sigma^2) - 1 \right) \\
&\quad + \exp(\sigma^2) \sum_{s=1}^{\infty} \left(\sum_x b^{*}_{xs} \left(\prod_{t=0}^{s-1} (1 - q^{*}_{xt}) \right) \right)^2 \left(\mathrm{E}(y^{2s}) \exp(s\tau^2) - \mathrm{E}(y^s)^2 \right) \\
&\quad + \exp(\sigma^2) \sum_x \sum_{s=1}^{\infty} \sum_{r=1}^{\infty} b^{*}_{xs} b^{*}_{xr} \exp\left(\min(r,s)\tau^2\right) \left(\prod_{t=0}^{\max(r,s)-1} (1 - q^{*}_{xt}) \right) \\
&\quad \times \left(\mathrm{E}\left(y^{\max(r,s)}\right) - \mathrm{E}\left(y^{r+s}\right) \left(\prod_{t=0}^{\min(r,s)-1} (1 - q^{*}_{xt}) \right) \right)
\end{aligned}
\tag{3.17}
$$

In these formulae we have taken

(3.18) $$b_{xt}^{*} = a_{xt}^{*} \prod_{s=1}^{t} \frac{1+f_s^{*}}{1+v_s^{*}}$$

These are the present value of future payments without consideration of uncertainty or the probability of payment. As a practical matter, the value of σ^2 is not needed in the detailed calculations. We can calculate the various terms in (3.17) that involve individual claim information separately, and then include the value of σ^2 in a fairly simple calculation.

If, now, we assume that there is no uncertainty in any of the estimates then $\sigma = \tau = 0$ and the expectations of all powers of *y* are 1 (infinite sample size) the first three terms in the variance sum vanish leaving:

(3.19) $$\operatorname{Var}(T|\text{Certainty}) = \sum_{x} \sum_{s=1}^{t} \sum_{r=1}^{t} b_{xs}^{*} b_{xr}^{*} \left(\prod_{l=0}^{\max(r,s)-1} \left(1-q_{xl}^{*}\right) \right) \left(1 - \prod_{l=0}^{\min(r,s)-1} \left(1-q_{xl}^{*}\right) \right)$$

Here, and throughout this paper, we use the term "Certainty" in the formulae to denote the situation where there is no parameter uncertainty. We use this shorthand to help keep the formulae as simple as possible.

Thus incorporating uncertainty regarding the closure rates adds to the expected value of the total. With this we see that E(*T*) will equal the reserve estimates calculated by the model if the survival rate were based on an infinite population, otherwise said, if we are certain about the annual survival rate.

If we define

(3.20) $$p_{xt}^{*} = \prod_{s=0}^{t-1} \left(1-q_{xs}^{*}\right)$$

Then this last formula becomes the standard variance formula.

$$\text{(3.21)}\quad \begin{aligned}\mathrm{Var}(T|\text{Certainty}) &= \sum_x \sum_{s=1}^{\infty}\sum_{r=1}^{\infty} b_{xs}^{*} b_{xr}^{*}\left(\sum_{t=\max(r,s)}^{\infty} p_{xt}^{*}\right)\left(1-\sum_{t=\min(r,s)}^{\infty} p_{xt}^{*}\right)\\
&= \sum_x \sum_{s=1}^{\infty}\sum_{r=1}^{\infty} b_{xs}^{*} b_{xr}^{*}\left(\left(\sum_{t=\max(r,s)}^{\infty} p_{xt}^{*}\right)-\left(\sum_{t=r}^{\infty} p_{xt}^{*}\right)\left(\sum_{t=s}^{\infty} p_{xt}^{*}\right)\right)\\
&= \sum_x \sum_{s=1}^{\infty}\sum_{r=1}^{\infty} b_{xs}^{*} b_{xr}^{*}\left(\sum_{t=\max(r,s)}^{\infty} p_{xt}^{*}\right)-\sum_x \sum_{s=1}^{\infty}\sum_{r=1}^{\infty} b_{xs}^{*} b_{xr}^{*}\left(\sum_{t=r}^{\infty} p_{xt}^{*}\right)\left(\sum_{t=s}^{\infty} p_{xt}^{*}\right)\\
&= \sum_x \left(\sum_{t=1}^{\infty} p_{xt}^{*}\sum_{s=1}^{t}\sum_{r=1}^{t} b_{xs}^{*} b_{xr}^{*}-\left(\sum_{s=1}^{\infty} b_{xs}^{*}\left(\sum_{t=s}^{\infty} p_{xt}^{*}\right)\right)^2\right)\\
&= \sum_x \left(\sum_{t=1}^{\infty} p_{xt}^{*}\left(\sum_{s=1}^{t} b_{xs}^{*}\right)^2-\left(\sum_{t=1}^{\infty} p_{xt}^{*}\left(\sum_{s=1}^{t} b_{xs}^{*}\right)\right)^2\right)\end{aligned}$$

The actual calculations in deriving (3.17) are quite lengthy and are contained in Appendix C. Similarly we have the following formulae for the mean and variance of payments in year *s*, as shown in detail in Appendix D.

$$\mathrm{E}(T_s)=\sum_x b_{xs}^{*}\left(\prod_{t=0}^{s-1}(1-q_{xt}^{*})\right)\prod_{i=0}^{s-1}\frac{(n+2)(z+i+1)}{(z+1)(n+i+2)}$$

$$\text{(3.22)}\quad \begin{aligned}\mathrm{Var}(T_s) &= \left(\sum_x b_{xs}^{*}\left(\prod_{t=0}^{s-1}(1-q_{xt}^{*})\right)\right)^2\left(\exp(\sigma^2+s\tau^2)\mathrm{E}(y^{2s})-\mathrm{E}(y^s)^2\right)\\
&\quad+\exp(\sigma^2+s\tau^2)\sum_x b_{xs}^{*2}\left(\prod_{t=0}^{s-1}(1-q_{xt}^{*})\right)\left(\mathrm{E}(y^s)-\mathrm{E}(y^{2s})\left(\prod_{t=0}^{s-1}(1-q_{xt}^{*})\right)\right)\end{aligned}$$

Although, to maintain some simplicity we have not substituted from formula (3.9) in the variance formula in either (3.17) or (3.22), both formulae, with this substitution, no longer depend on the conditional variables. It can be easily seen that in the case of no uncertainty (i.e. $\sigma=\tau=0, \mathrm{E}(y^s)=1$) the formulae in (3.22) reduce to (2.7) and (2.8). We also see that the expected total reserve in (3.17) is simply the sum of the expected payments by future year from (3.22). However, as we would expect, the variance terms are much less comparable. This is due to the nature of the dependencies we introduced with some of the uncertainty variables.

Thus, for the relatively simple case of known lifetime care claimants we can calculate the mean and variance for both the total reserves and the payments in each future year. We can incorporate at least some parameter uncertainty in these calculations.

In short, these calculations provide a way to estimate the mean and variance of case reserves, including a potential provision for uncertainty in the case estimates as evidenced by the parameter u, but do not consider uncertainty regarding claims that are incurred but not reported. It also does not consider reported claims that are not yet recognized as potential lifetime care claimants or for which there is not sufficient information available to establish estimates of future claim and medical costs.

4. Additional Areas of Uncertainty

We consider three categories of claims.

1. Those having annual cost estimates with case reserves calculated using the annuity model described in sections 2 and 3 above.

2. Claims reported but for which annual cost information is not yet available, and

3. Claims incurred but not reported (true IBNR).

Continuing with our development we have implicitly incorporated additional development in case reserves, along with its corresponding uncertainty, in the estimates in section 3. Thus there is increasing uncertainty as we move through these categories of claims. In the first instance we have information regarding individual claims with uncertainty regarding inflation, investment, exit from the population, and some uncertainty regarding the accuracy of the annual cost estimates. All these elements of uncertainty are present in the second category along with additional uncertainty as to the overall average for the claims themselves. Finally the third category incorporates all this uncertainty as well as uncertainty as to the number of claims to ultimately be reported.

In order to reflect this uncertainty we will use the following notation. Let:

N_R denote the number of claims having annual cost estimates

N_B denote the number of reported claims without specific annual cost estimates

λ denote the expected number of IBNR claims

χ denote a random variable with $E(\chi)=1$ and $Var(\chi)=c$

β denote a random variable with $E(\beta)=1$ and $Var(\beta)=b$

γ denote a random variable with E(γ)=a and Var(γ)=d

ζ denote a random variable with E(ζ)=r and Var(ζ)=z

With this notation, we will use a modification of Algorithm 3.3 from the Heckman & Meyers[1] paper:

1. Select claims with case reserves, $X_1, X_2, \ldots, X_{N_R}$
2. Randomly select a value for χ.
3. Randomly select N from a Poisson distribution with expected value $\lambda\chi$.
4. Randomly select independent claims $X_{N_R+1}, X_{N_R+2}, \ldots, X_{N_R+N_B+N}$ from the same distribution having the mean and variance equal to that of the case reserved claims.
5. Randomly select values for β, ζ, and γ.
6. Calculate the aggregate reserve as

$$(4.1) \qquad T = \beta\left(\sum_{i=1}^{N_R} X_i + \zeta \sum_{i=N_R+1}^{N_R+N_B} X_i + \gamma \sum_{i=N_R+N_B+1}^{N_R+N_B+N} X_i\right).$$

Here χ incorporates uncertainty regarding the claim count estimate, β global uncertainty regarding the overall estimates, ζ additional uncertainty and scaling for known but not-case-reserved claims, and γ additional uncertainty and scaling for IBNR claims. We will assume in the following, that claims other than those with case reserves, except for the scaling values a and r, will have the same mean and variance as those with individual case reserves.

If we consider the case where there are no IBNR claims and that we have case reserve estimates for all claims, (4.1) becomes:

$$(4.2) \qquad T = \beta\sum_{i=1}^{N_R} X_i$$

From this we can calculate the mean as:

$$(4.3)\qquad \begin{aligned} E(T_R) &= E_\beta\left(E\left(\sum_{i=1}^{N_R} X_i|\beta\right)\right) \\ &= E_\beta\left(\beta\sum_{i=1}^{N_R} E(X_i)\right) \\ &= E_\beta(\beta)\sum_{i=1}^{N_R} E(X_i) = \sum_{i=1}^{N_R} E(X_i) = N_R\,E(X) \end{aligned}$$

Since we are assuming all the claims are independent, this last term denotes the expected claim costs with no parameter uncertainty. This can be calculated using (3.17) by letting the sample size tend to infinity. Now if we let all the uncertainty above be expressed in the parameter β, then we have

$$(4.4)\qquad \begin{aligned} \mathrm{Var}(T_R) &= E_\beta\left(\mathrm{Var}(T_R|\beta)\right) + \mathrm{Var}_\beta\left(E(T_R|\beta)\right) \\ &= E_\beta\left(\mathrm{Var}\left(\beta\sum_{i=1}^{N_R} X_i\right)\right) + \mathrm{Var}_\beta\left(E\left(\beta\sum_{i=1}^{N_R} X_i\right)\right) \\ &= E_\beta\left(\beta^2\,\mathrm{Var}\left(\sum_{i=1}^{N_R} X_i\right)\right) + \mathrm{Var}_\beta\left(\beta E\left(\sum_{i=1}^{N_R} X_i\right)\right) \\ &= E_\beta(\beta^2)\,\mathrm{Var}\left(\sum_{i=1}^{N_R} X_i\right) + \mathrm{Var}_\beta(\beta)E\left(\sum_{i=1}^{N_R} X_i\right)^2 \\ &= \left(\mathrm{Var}_\beta(\beta) + E_\beta(\beta)^2\right)\mathrm{Var}\left(\sum_{i=1}^{N_R} X_i\right) + \mathrm{Var}_\beta(\beta)E\left(\sum_{i=1}^{N_R} X_i\right)^2 \\ &= (b+1)\,\mathrm{Var}(T_R|\mathrm{Certainty}) + bE(T_R|\mathrm{Certainty})^2 \end{aligned}$$

Solving for b we obtain:

$$(4.5)\qquad b = \frac{\mathrm{Var}(T_R) - \mathrm{Var}(T_R|\mathrm{Certainty})}{\mathrm{Var}(T_R|\mathrm{Certainty}) + E(T_R|\mathrm{Certainty})^2}$$

We can then use (3.17) or (3.22) to derive a value for b that will explicitly incorporate parameter uncertainty into this algorithm. Assuming, in addition, that estimates for the second and third claim categories depend on case reserves, we are able to quantify a level of global uncertainty inherent in the estimates.

We use calculations similar to those led us to the mean and variance estimates in Appendices C and D to obtain the following:

$$\mathrm{E}(T) = (N_R + rN_B + a\lambda)\mathrm{E}(T_R|\text{Certainty})/N_R$$

$$(4.6)\ \mathrm{Var}(T) = (b+1)\left(1 + \frac{N_B(z+r^2)+(d+a^2)\lambda}{N_R}\right)\mathrm{Var}(T_R|\text{Certainty})$$
$$+\left(\frac{(b+1)\left((d+a^2)(\lambda+c\lambda^2)+\lambda^2 d + zN_B^2\right)}{N_R^{\ 2}} + b\left(1+\frac{rN_B+a\lambda}{N_R}\right)^2\right)\mathrm{E}(T_R|\text{Certainty})^2$$

These are shown in detail in Appendix E.

Thus, under the above assumptions, we can express the mean and variance of the distribution of total claims in terms of the mean and variance of the distribution of case reserved claims, without parameter uncertainty, and the various parameters specified above.

On review of that analysis we see that we did not specifically assume that the calculations were for total reserves. Thus a similar formula holds for payments in a particular year:

$$\mathrm{E}(T_s) = (N_R + rN_B + a\lambda)\mathrm{E}(T_{Rs}|\text{Certainty})/N_R$$

$$(4.7)\ \mathrm{Var}(T_s) = (b+1)\left(1 + \frac{N_B(z+r^2)+(d+a^2)\lambda}{N_R}\right)\mathrm{Var}(T_{Rs}|\text{Certainty})$$
$$+\left(\frac{(b+1)\left((d+a^2)(\lambda+c\lambda^2)+\lambda^2 d + zN_B^2\right)}{N_R^{\ 2}} + b\left(1+\frac{rN_B+a\lambda}{N_R}\right)^2\right)\mathrm{E}(T_{Rs}|\text{Certainty})^2$$

Inherent in these calculations is that we can use the same uncertainty variables for both the aggregate reserves and for the payments in each year.

We note that, although the genesis of (4.6) and (4.7) were based on a book of life-pension claims, there is nothing in the derivation that requires such a book. If we can separate our reserving problem into the three categories above and are willing to make the assumptions indicated above, we can calculate the variance of the aggregate distribution.

5. Estimating the Parameters

We will consider parameter estimation in two phases, we will first address the ***b*** parameter and then the remaining ones. Again, the discussion will begin with the life annuity model and then move to potential for generalization.

5.1 Estimating the b Parameter

We have already hinted at an approach that we could use to estimate the *b* parameter. Using (4.5) all we need are estimates of the variance of reserved claims with and without parameter uncertainty. The estimate without parameter uncertainty follows directly from the annuity calculations as given in (3.19) or (3.21). Using (3.17) and the assumptions going into that estimate we can derive an estimate of the variance for claims having case reserves if we can estimate:

$E(y^s)$	Uncertainty regarding the mortality assumptions
τ^2	Uncertainty regarding (composite) economic estimates
σ^2	Uncertainty regarding the annual cost estimates

5.1.1 Mortality Considerations

There are other practical issues in the use of mortality assumptions, especially in usual applications in property and casualty insurance. In almost every situation property and casualty claimants eligible for lifetime care will be physically impaired in some manner, either by trauma or disease. Often one may expect the impairment to affect the claimant's future survival chances as compared to the general population. In addition, we could expect different injuries to have different effects on survival probabilities.

There has been substantial research on the effect of spinal cord injuries on survival rates. As opposed to head trauma, spinal cord injuries are relatively easy to categorize and are relatively uniform from patient to patient, and generally do not change during a claimant's life. For example, the following table, attributed to the National Spinal Cord Injury Statistical Center, University of Alabama at Birmingham, shows differences in life expectancies for various levels of spinal cord injury[3]:

Table 8: Life Expectancies by Age and Spinal Cord Injury

Current Age	Normal	Ventilator Dependent	High Tetraplegic	Low Tetraplegic	Paraplegic	Motor Function at Any Level
	Life Expectancy					
20	56.3	19.9	32.8	38.6	44.8	49.0
30	46.9	15.9	26.8	30.7	36.7	40.5
40	37.6	12.4	20.9	23.6	28.8	31.7
50	28.6	9.3	15.5	17.0	21.2	23.4

We have not been able to locate similar statistics for traumatic head injuries. Analysis for such injuries are complicated by the fact that head injuries are more difficult to categorize than spinal cord injuries and, in contrast to spinal cord injuries usually identified by the location and degree of lesion in the spinal column. In addition, the level of severity of a head injury can change substantially during the course of treatment.

Other property and casualty claimants could have still different mortality profiles. For example, a back injury, though disabling a person from employment, may have little or no effect on that person's future life expectancy. Conversely, heart conditions or stress related illnesses could have a substantial impact on future survival chances. Compounding difficulties are the effects of medical treatment on the claimant's survival chances, especially in situations where there is no limit on the amount that can be expended for medical treatment. Thus, unlike many situations where mortality is a consideration, the appropriate survival functions are often uncertain.

For this reason, it may be useful to consider construction or modification of mortality tables to reflect the injured population. In this case the table could be based on a fairly small sample, though could still produce reasonable results. In this case formula (3.9) gives an estimate of $E(y^s)$ under the assumption that uncertainty in the mortality table is uniform across all claimants and ages and depends only on the sample size used in estimating the mortality table and the overall average mortality for the population. However, the considerations above would seem to indicate that (3.9) may only produce a lower bound on the level of uncertainty inherent in the selection of mortality assumptions.

5.1.2 Uncertainty in Economic Assumptions

We note in (3.5) and (3.12) we have made the simplifying assumptions that the net discount rates (ratio of annual cost inflation to annual interest rate) are independent from year to year. In addition, we assumed that the distributions of the rates in each year all have the same coefficient of variation.

There has been much attention recently devoted to modeling economic scenarios in conjunction with dynamic financial analysis, for example Daykin et.al.[4] If we were using such models one could estimate the value of τ^2 using the results of those models.

Although the models can be quite complex, actual economic conditions have experienced some rather spectacular swings, even over the past twenty to thirty years. For example, the hospital room component of the U.S. Consumer Price Index for Urban Wage Earners (CPI-W) increased by 15.7% during 1981 and by only 3.5% during 1996. Interest rates also experienced similar swings during that same time with the average 1 year United States Treasury Bill moving from 14.8% in 1981 to 5.5% in 1996.

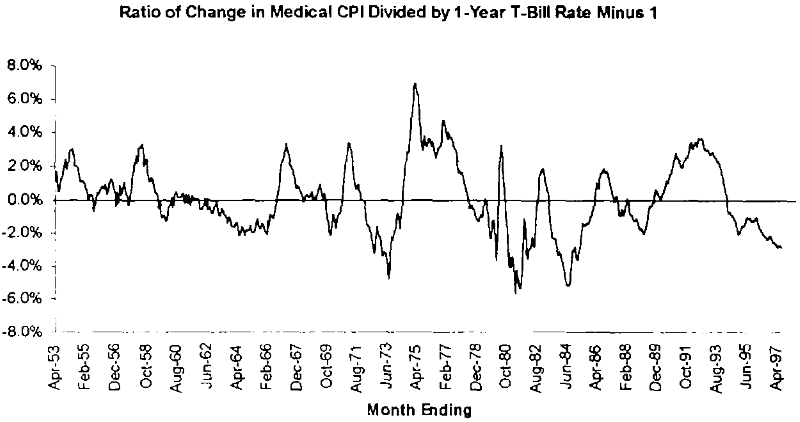
We could also use this historic volatility to estimate τ^2. For this we review the historical volatility in the quantity:

$$(5.1) \qquad \frac{1+f_t^*}{1+v_t^*}-1$$

Here we use f_t^* to denote the annual change in the medical care cost component of the U.S. Consumer Price Index for Urban Wage Earners measured from month t-12 to month t and v_t^* to denote the average yield for 1 Year U.S. Treasury Bills during month t. Of course, if we assume that claim costs would experience a different market basket than medical costs in general then we would re-weight them accordingly.

We also somewhat randomly selected the 1-Year U.S. Treasury Bill rate for this example. Again, unique characteristics of the company's investment portfolio may dictate a different measure for investment return. These values should be illustrative of the degree of variation we could expect in our applications. The following graph shows values of (5.1) for each month from April 1953 through July 1997.

Figure 1: *Relative Real Returns*



Since we have assumed that uncertainty in future net discount will show a lognormal distortion we can estimate the τ^2 parameter as the variance of the natural logarithms of the amounts in (5.1), plus 1. In this case the result is $\tau^2 = 0.000457$.

5.1.3 Uncertainty in Cost Assumptions

The third area of uncertainty reflected in (3.17) deals with the fact that a_{xt}, the annual payment amounts in current dollars, may themselves be uncertain. In workers compensation claims the indemnity amounts are often specified by statute, so the amounts of those payments for life pension cases may not be subject to change. However, one would probably not expect the same degree of certainty in medical payments either for workers compensation or no-fault benefits.

As noted in Section *3. Introducing Some Uncertainty*, we have assumed that claim annual cost estimates in current dollars have the same uncertainty distribution as reflected by the random variable u. In addition to u, the (present value of) annual payments are also affected by the w_t random variables. From a practical viewpoint, this effectively separates two factors that affect the accuracy of estimates of future costs; unexpected levels of inflation (and/or investment return) and actual costs (or services) differing from what had been expected for reasons other than economic conditions.

This dichotomy suggests a way to estimate the parameter σ^2. We could compare actual annual payments with the forecasts made in previous analyses, after adjustment for trend in the form of some index reflecting underlying cost changes. The following table provides an example of such an approach.

Table 9: Actual vs. Expected Payments

Payment Year	Forecast Year	Annual Payment Actual	Annual Payment Estimated	ln(A/E)
1	0	$ 50,000	$ 45,000	0.1054
2	0	40,000	35,000	0.1335
2	1	40,000	45,000	-0.1178
3	0	30,000	25,000	0.1823
3	1	30,000	35,000	-0.1542
3	2	30,000	30,000	0.0000
			Average	0.0249

In this example, for a single claim, we have actual payments of $50,000, $40,000, and $30,000 in each of the first three years of a claim. In the first analysis (at the beginning of year 1) we estimated payments of $45,000, $35,000, and $25,000 trended to future levels using the selected cost index. The second analysis we adjusted the forecasts for years 2 and 3 to $45,000 and $35,000 respectively, while for the third analysis we estimated $30,000 for the third year.

Since the sample mean of the natural logarithms is the maximum likelihood estimator for the first parameter of a lognormal distribution in our parameterization, and the sample variance is a minimum variance estimator for the second parameter, we could use the sample variance for as an estimator of the σ^2 parameter. We note here that the average does not satisfy the relationship assumed in (3.12). In particular the expected value of the resulting lognormal variable is not unity. Hence our estimates are biased and we should adjust the forecast estimates to remove this indicated bias. Such an adjustment would leave the σ^2 parameter unchanged.

This approach ignores any "aging" considerations. For example, one would expect short-term forecasts to be more accurate than long term ones, all other things being equal. In addition, the longer-term estimates carry less weight in the reserve forecasts due to discounting for mortality if not for investment income.

Also, for medical payments on seriously injured claimants, one would often expect payments in the first years after the accident to be much higher than those in later years after the claimant has medically stabilized. In addition, it could be argued that payments rise during the time just before a claimant's death. The approach we outlined gives equal weight to all forecast errors in estimating the σ^2 parameter. It does, however, have the appeal of a direct comparison of actual versus expected results.

An alternative approach would be to consider the development of claim estimates over time. In such an approach, as in usual incurred loss development, annual cost estimates are gradually replaced by actual payments over the development period. If we take this approach we must keep in mind that we want to separate economic influences from the measurement of movement of claim costs over time.

One such approach would involve recalculating all expected incurred losses each year, replacing expected future payments with actual payments in the annuity calculations and reviewing the development. This would be the most consistent way to handle changes in economic assumptions in the valuations. However, it could be quite time-consuming, especially in situations where there are many claims evaluated over many different development periods, not to mention the need to maintain records of past annual cost estimates for individual claims.

There is an approximation, however, that would allow for the separation of changes in economic assumptions from development in estimates from other causes. At this point we only consider claims having annual cost estimates, since we are trying to quantify the uncertainty in those annual cost estimates. Thus we do not want the development patterns we obtain to be influenced by emergence of new claims, hence aggregation by accident period would not be useful.

This may suggest grouping by report period. However, in that grouping there could be claims reported but which do not yet have individual annual cost estimates attached. The manner in which reserves are set on those claims could influence the review of development on claims having annual cost estimates. Hence report period grouping also seems to be lacking for this purpose.

We thus consider a third alternative, akin to report period. For this we group claims by the period in which they are first case reserved, calling this a reserve period grouping. In the case that there are no "formula" reserves for known claims, this alternative would be equivalent to a report period grouping.

Once claims are grouped in this fashion, we can consider the development of expected incurred losses (calculated using the annuity approach of (2.3)) on fixed groups of claims using a development array format. However, we are faced with several additional difficulties if we wish to focus on the movement and variability in the individual annual cost estimates (the focus of the σ^2 parameter). Those difficulties arise because our reserve estimates may be discounted and because changes in economic or mortality assumptions will cause changes in the expected amounts during the calendar period containing the change and should not be considered when evaluating the variability inherent in the individual annual cost estimates.

Even without changes in underlying assumptions, we are faced with the "unwinding of the discount" phenomenon. By this we mean the fact that incurred losses calculated with discounted reserves will continue to develop upward due to a decreasing effect of discounting, even if all underlying assumptions prove exactly correct. To deal with the unwinding of the discount we discount all amounts to the beginning of the reserve period. This discounting includes the discounting of all payments made to date, as well as discounting of reserves. For convenience we discount to the beginning of the reserve period we are evaluating.

An obvious alternative at this juncture would be to not discount at all. The appeal of discounting at this point, however, is the decreasing influence of remote payments have on the final reserve calculated. As noted above, these remote amounts are probably subject to greater uncertainty. The author recognizes at this point the current discussions regarding the appropriateness of calculating reserves on a discounted basis. None of the methods or results presented here rely on the discount rate being positive. Thus if reserves are carried on a undiscounted basis all the above analysis will apply. However, if the discount rate is negative (implying a significant risk-adjustment due to uncertainty) later payments are given increasing weight in the final expected value calculations.

In any event, however, if we were to discount all amounts to the beginning of the reserve period and if all estimates were exactly correct we would see no development in these amounts over time. In addition, if economic conditions (and assumptions regarding future conditions) remain unchanged all movement in total incurred amounts would reflect changes in future annual cost estimates making up the case reserve estimates. Hence we could quantify variation in those estimates over time, using, for example, techniques developed in Hayne[5], Mack[6] or others.

A practical consideration still remains, however. In reality, assessments of future economic conditions change over time. For example, in the 1980's it may not have been unreasonable to assume that medical cost inflation would remain quite high over a fairly long period of time. However, given the situation in the late 1990's, we may be hard pressed to justify estimates of future inflation at levels experienced in the 1980's. As noted above, such changes would appear as calendar period effects in the development patterns and could mislead estimate of uncertainty in claim cost estimates.

Specific changes such as those in assumed future economic conditions will affect reserve estimates similar to those of currency fluctuations on losses denominated in more than one currency. Borrowing techniques developed to handle such changes, as presented in Duncan and Hayne[7] we can consider a type of two-step development array.

Table 10: Example Two-Stage Development

	Months of Development				
Reserve	12	24		36	
Year	Current	Prior	Current	Prior	Current
1995	$100,000	$110,000	$105,000	$107,100	$109,500
1996	125,000	143,750	137,500		
1997	175,000				

	Development Factors	
	24/12	36/24
1995	1.10	1.02

In this two-stage approach we use "Current" to denote the assumptions inherent in the final selected analysis at the indicated valuation date. For example, $105,000 indicates the total incurred (discounted to the beginning of 1995) using the economic assumptions at the 1996 valuation. Similarly $109,500 represents the discounted incurred (again to the beginning of 1995) using the 1997 economic assumptions.

The "Prior" amounts denote the calculations using the economic assumptions from the prior analysis. For example, the $110,000 represents the forecasts for 1995 claims, using 1996 claim information, but using the economic assumptions inherent in the 1995 (prior) analysis. Thus the difference between $100,000 (1995 at 12 months) and $110,000 is due to the evaluation of the individual claims and not due to different economic assumptions used in calculating the losses. The development factors are then comparisons between the "Prior" at one stage of development with the "Current" at the previous stage. In effect, then, the development isolates changes in economic assumptions from development in underlying cost estimates.

From this point we could use the variation inherent in these development factors to estimate uncertainty in annual cost estimates, and thus the σ^2 parameter.

5.2 Estimating the r and z Parameters

The next portion of total reserves in our consideration is that for known but not-case-reserved claims. If we assume that there is no inherent difference between these claims and those already reported, we could assume their distribution is the same as that for known claims and take $r = 1$ and $z = 0$.

However, there may be other factors considered in setting the formula reserves for these claims. The r and z parameters can then be used to account for these factors and resulting additional uncertainty. For example, assume the formula reserves are set only during the first three years after claim occurrence, using only the most recent three accident years, without any adjustment for trend or differences by report lag. The following then shows one approach to estimating r and z in this case:

Table 11: Estimate of *r* and *z* Parameters

Accident Year	Report Year	Reported Losses	Reported Claims	Loss Average	Loss Standard Deviation
1995	1995	$ 5,000	200	$ 25,000	$ 27,500
1995	1996	5,100	300	17,000	15,300
1995	1997	5,500	250	22,000	23,100
1996	1996	9,800	350	28,000	22,400
1996	1997	4,180	220	19,000	20,900
1997	1997	10,500	350	30,000	31,500
Total		$ 40,080	1,670	$ 24,000	$ 24,635
Expected Without Uncertainty				$ 20,000	$ 18,000
Parameter Estimates:					
r				1.20	
z				0.19	

The estimate for *r* is simply the ratio of the average for the "formula" reserved claims to the expected average (without parameter uncertainty). The estimate of *z* follows from the assumptions regarding the form of uncertainty for these formula reserves. In particular, assuming the random variable *Y* is defined using the notation in Section *4. Additional Areas of Uncertainty* as:

$$(5.2)\qquad Y = \zeta X$$

We then have the following formula for the variance of Y:

$$\begin{aligned}
\mathrm{Var}(Y) &= \mathrm{E}_{\zeta}\left(\mathrm{Var}(X|\zeta)\right) + \mathrm{Var}_{\zeta}\left(\mathrm{E}(X|\zeta)\right) \\
&= \mathrm{E}_{\zeta}\left(\mathrm{Var}(X\zeta)\right) + \mathrm{Var}_{\zeta}\left(\mathrm{E}(X\zeta)\right) \\
&= \mathrm{E}_{\zeta}\left(\zeta^2\,\mathrm{Var}(X)\right) + \mathrm{Var}_{\zeta}\left(\zeta\,\mathrm{E}(X)\right) \\
&= \mathrm{Var}(X)\mathrm{E}\left(\zeta^2\right) + \mathrm{E}(X)^2\,\mathrm{Var}(\zeta) \\
&= \mathrm{Var}(X)\left(\mathrm{Var}(\zeta) + \mathrm{E}(\zeta)^2\right) + \mathrm{E}(X)^2\,\mathrm{Var}(\zeta) \\
&= \mathrm{Var}(X)\left(z + r^2\right) + \mathrm{E}(X)^2 z
\end{aligned}$$

Solving for *z* we obtain:

(5.3) $z = \frac{\text{Var}(Y) - r^2 \text{Var}(X)}{\text{Var}(X) + \text{E}(X)^2}$

5.3 Estimating the c, a, and d Parameters

The final portion of total reserves is for claims that are incurred but not reported. As with known claims with formula reserves, if IBNR reserves are estimated using averages for known claims we could estimate the *a* and *d* parameters similar to the way we estimated the *r* and *z* parameters as described in Section 5.2, *Estimating the r and z Parameters*.

We could estimate the *c* parameter in several ways. One approach starts with the assumption that the number of IBNR claims has a Poisson distribution with a "contagion" parameter similar to that used by Heckman and Meyers.[1] With that assumption we see from Appendix E that with our notation above if N denotes the number of IBNR claims:

(5.4) $\text{E}(N) = \lambda$, and

(5.5) $\text{Var}(N) = \lambda + c\lambda^2$

Solving (5.5) for c we obtain:

(5.6) $c = \frac{\text{Var}(N) - \lambda}{\lambda^2}$

If we estimated the number of IBNR claims using development of reported claims then Hayne[5] provides an approach we could use to estimate total variance in the IBNR estimates, if we are willing to assume independence among the various accident (or exposure) years. Consider the following example:

Table 12: Example Reported Count Development

Accident	Months of Development					
Year	12	24	36	48	60	72
1989	176	363	417	477	500	500
1990	314	384	519	524	550	550
1991	178	294	382	405	425	425
1992	323	472	535	590	620	620
1993	264	492	506	572	600	
1994	253	419	441	495		
1995	137	324	410			
1996	304	415				

Following Hayne, and assuming independence of the age-to-age factors (to keep the calculations simple) we calculate the natural logarithms of the age-to-age factors, their means and standard deviations as parameter estimates for the lognormal distributions of the age-to-age factors. Also, given independence the parameters for the age-to-ultimate factors can then be determined from the parameters of the age-to-age factors by simply summing the means and variances. The following shows these calculations:

Table 13: Logarithms of Claim Age-to-Age Factors

Accident	Months of Development				
Year	24/12	36/24	48/36	60/48	72/60
1989	0.7239	0.1387	0.1344	0.0471	0.0000
1990	0.2012	0.3013	0.0096	0.0484	0.0000
1991	0.5018	0.2618	0.0585	0.0482	0.0000
1992	0.3793	0.1253	0.0979	0.0496	0.0000
1993	0.6225	0.0281	0.1226	0.0478	
1994	0.5045	0.0512	0.1155		
1995	0.8608	0.2354			
1996	0.3113				
Mean	0.5132	0.1631	0.0897	0.0482	0.0000
Std.Dev.	0.2182	0.1056	0.0473	0.0009	0.0000
Cumulative:					
Mean	0.8142	0.3011	0.1380	0.0482	0.0000

Finally, using standard formulae for the lognormal we obtain the following projected number of claims and their corresponding variance:

Table 14: Estimate of *c* Parameter

Accident	Cumulative Parameters		Reported	Forecast	
Year	Mean	Std.Dev.	Claims	Mean	Std.Dev.
1989	0.0000	0.0000	500	500	-
1990	0.0000	0.0000	550	550	-
1991	0.0000	0.0000	425	425	-
1992	0.0000	0.0000	620	620	-
1993	0.0000	0.0000	600	600	-
1994	0.0482	0.0009	495	519	0.5
1995	0.1380	0.0474	410	471	22.3
1996	0.3011	0.1157	415	565	65.5
1997	0.8142	0.2470	282	656	164.6
Total			4,297	4,906	178.6
Indicated IBNR				609	178.6
Indicated *c* Value:					0.084

6. Conclusion

In this paper we have set out one approach that can be used to systematically estimate variation in both total reserve estimates and in payments in individual future years. In explicitly accounting for various components of uncertainty the actuary can adapt these estimates to be used in DFA applications. In such applications economic conditions can form a link between asset and liability models. Explicit recognition of the influence of such factors on loss reserve and payment uncertainty in the liability models will prevent "double counting" of its effect and result in potentially more realistic DFA models.

We have presented this as a first step. There are obviously many simplifying assumptions even in this rather complex presentation. We hope this framework can provide a useful starting point to build and parameterize models of the amount and timing of insured liabilities.

7. *Bibliography*

[1] Heckman, P.E. and Meyers, G.G., "The Calculation of Aggregate Loss Distributions from Claim Severity and Claim Count Distributions," *Proceedings of the Casualty Actuarial Society*, vol. LXX, 1983, pp. 22-61, with additional exhibits presented in vol LXXI, 1984.

[2] Robertson, J.P., "The Computation of Aggregate Loss Distributions," *Proceedings of the Casualty Actuarial Society*, vol. LXXIX, 1992, pp. 57-133.

[3] See the website http://www.neurolaw.com, attributed to the National Spinal Cord Injury Statistical Center, University of Alabama at Birmingham.

[4] Daykin, C.D.; Pentikäinen, T.; and Pesonen, M.; *Practical Risk Theory for Actuaries*, First Edition, 1994, Chapman and Hall, pp. 211-276.

[5] Hayne, R.M.; "An Estimate of Statistical Variation in Development Factor Methods," *Proceedings of the Casualty Actuarial Society*, vol. LXXII, pp. 25-43.

[6] Mack, T.; "Distribution-free Calculation of the Standard Error of Chain Ladder Reserve Estimates," *ASTIN Bulletin*, Volume 23, No. 2, November 1993, pp. 213-226.

[7] Duncan, F. and Hayne, R.; "Some Aspects of Currencies and Exchange Rates in the London Market," *International Topics, Global Insurance Pricing, Reserving and Coverage Issues*, Casualty Actuarial Society, 1991 Discussion Paper Program, pp. 87-144.

APPENDIX A

In this appendix we derive the formula for the variance for an individual life pension claimant, formula (2.4). From our definitions we have:

$$
\begin{aligned}
E(X^2) &= \sum_{t=1}^{\infty} p_{xt} \left(\sum_{s=1}^{t} a_{xs} \right)^2 \\
&= \sum_{t=1}^{\infty} \sum_{s=1}^{t} \sum_{r=1}^{t} p_{xt} a_{xs} a_{xr} \\
&= \sum_{s=1}^{\infty} \sum_{r=1}^{\infty} \sum_{t=\max(s,r)}^{\infty} p_{xt} a_{xs} a_{xr} \\
&= \sum_{s=1}^{\infty} \sum_{r=1}^{\infty} a_{xs} a_{xr} \sum_{t=\max(s,r)}^{\infty} p_{xt} \\
&= \sum_{s=1}^{\infty} \left(\sum_{r=1}^{s} a_{xs} a_{xr} \sum_{t=s}^{\infty} p_{xt} \right) + \left(\sum_{r=s+1}^{\infty} a_{xs} a_{xr} \sum_{t=r}^{\infty} p_{xt} \right)
\end{aligned}
$$

We also have:

$$
\begin{aligned}
E(X)^2 &= \left(\sum_{t=1}^{\infty} p_{xt} \sum_{s=1}^{t} a_{xs} \right)^2 \\
&= \left(\sum_{s=1}^{\infty} \sum_{t=s}^{\infty} p_{xt} a_{xs} \right)^2 \\
&= \sum_{s=1}^{\infty} \sum_{r=1}^{\infty} \left(\sum_{t=s}^{\infty} p_{xt} a_{xs} \right) \left(\sum_{t=r}^{\infty} p_{xt} a_{xr} \right) \\
&= \sum_{s=1}^{\infty} \sum_{r=1}^{\infty} \left(a_{xs} \sum_{t=s}^{\infty} p_{xt} \right) \left(a_{xr} \sum_{t=r}^{\infty} p_{xt} \right) \\
&= \sum_{s=1}^{\infty} \sum_{r=1}^{\infty} a_{xs} a_{xr} \left(\sum_{t=s}^{\infty} p_{xt} \right) \left(\sum_{t=r}^{\infty} p_{xt} \right) \\
&= \sum_{s=1}^{\infty} \left(\sum_{r=1}^{s} a_{xs} a_{xr} \left(\sum_{t=s}^{\infty} p_{xt} \right) \left(\sum_{t=r}^{\infty} p_{xt} \right) \right) + \left(\sum_{r=s+1}^{\infty} a_{xs} a_{xr} \left(\sum_{t=s}^{\infty} p_{xt} \right) \left(\sum_{t=r}^{\infty} p_{xt} \right) \right)
\end{aligned}
$$

Thus we have:

$$
\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}\left(X^2\right) - \mathrm{E}(X)^2 \\
&= \sum_{s=1}^{\infty}\left(\left(\sum_{r=1}^{s} a_{xs}a_{xr}\sum_{t=s}^{\infty} p_{xt}\right) + \left(\sum_{r=s+1}^{\infty} a_{xs}a_{xr}\sum_{t=r}^{\infty} p_{xt}\right)\right. \\
&\quad \left. - \left(\sum_{r=1}^{s} a_{xs}a_{xr}\left(\sum_{t=s}^{\infty} p_{xt}\right)\left(\sum_{t=r}^{\infty} p_{xt}\right)\right) - \left(\sum_{r=s+1}^{\infty} a_{xs}a_{xr}\left(\sum_{t=s}^{\infty} p_{xt}\right)\left(\sum_{t=r}^{\infty} p_{xt}\right)\right)\right) \\
&= \sum_{s=1}^{\infty}\left(\left(\sum_{r=1}^{s} a_{xs}a_{xr}\sum_{t=s}^{\infty} p_{xt}\right) + \left(\sum_{r=s+1}^{\infty} a_{xs}a_{xr}\sum_{t=r}^{\infty} p_{xt}\right) - \left(\sum_{r=1}^{s} a_{xs}a_{xr}\left(\sum_{t=s}^{\infty} p_{xt}\right)\left(\sum_{t=r}^{\infty} p_{xt}\right)\right)\right. \\
&\quad \left. - \left(\sum_{r=s+1}^{\infty} a_{xs}a_{xr}\left(\sum_{t=s}^{\infty} p_{xt}\right)\left(\sum_{t=r}^{\infty} p_{xt}\right)\right)\right) \\
&= \sum_{s=1}^{\infty}\left(\left(\sum_{r=1}^{s} a_{xs}a_{xr}\left(\sum_{t=s}^{\infty} p_{xt} - \left(\sum_{t=s}^{\infty} p_{xt}\right)\left(\sum_{t=r}^{\infty} p_{xt}\right)\right)\right)\right. \\
&\quad \left. + \left(\sum_{r=s+1}^{\infty} a_{xs}a_{xr}\left(\sum_{t=r}^{\infty} p_{xt} - \left(\sum_{t=s}^{\infty} p_{xt}\right)\left(\sum_{t=r}^{\infty} p_{xt}\right)\right)\right)\right) \\
&= \sum_{s=1}^{\infty}\left(\left(\sum_{r=1}^{s} a_{xs}a_{xr}\left(\sum_{t=s}^{\infty} p_{xt}\right)\left(1 - \left(\sum_{t=r}^{\infty} p_{xt}\right)\right)\right) + \left(\sum_{r=s+1}^{\infty} a_{xs}a_{xr}\left(\sum_{t=r}^{\infty} p_{xt}\right)\left(1 - \left(\sum_{t=s}^{\infty} p_{xt}\right)\right)\right)\right) \\
&= \sum_{s=1}^{\infty}\sum_{r=1}^{\infty} a_{xs}a_{xr}\left(\sum_{t=\max(r,s)}^{\infty} p_{xt}\right)\left(1 - \left(\sum_{t=\min(r,s)}^{\infty} p_{xt}\right)\right)
\end{aligned}
$$

APPENDIX B

In this appendix we derive the conditional distribution of θ given z observed open claims from our population of n claims. We also review the asymptotic behavior of this distribution.

1. Conditional Distribution of θ

We first assume that the number of claims remaining open from one year to the next has a binomial distribution with parameter θ. Although we will assume that θ will be uniformly distributed between 0 and 1, the following result holds in the more general case when θ has with a beta distribution with parameters α and β. In this case z, the number of "successes" (or claims remaining open) is given by:

$$f(z) = \binom{n}{z}\theta^z(1-\theta)^{n-z}$$

The *parameter θ then has the distribution:*

$$h(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

The joint distribution for z and θ is then given by:

$$\begin{aligned} k(z,\theta) &= g(z)h(\theta) \\ &= \binom{n}{z}\theta^z(1-\theta)^{n-z}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \\ &= \binom{n}{z}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{z+\alpha-1}(1-\theta)^{n-z+\beta-1} \end{aligned}$$

We will take $y = \theta/E(\theta)$. Now we need to get the distribution of θ given our observed annual closure rate, or conversely, rate of claims that remain open. From Bayes Theorem we obtain:

$$k(\theta|z) = \frac{k(\theta,z)}{\int_0^1 k(\theta,z)d\theta}$$

The integral in the denominator becomes:

$$\int_0^1 \mathrm{k}(\theta,z)d\theta = \int_0^1 \binom{n}{z}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{z+\alpha-1}(1-\theta)^{n-z+\beta-1}d\theta$$

$$= \binom{n}{z}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1 \theta^{z+\alpha-1}(1-\theta)^{n-z+\beta-1}d\theta$$

$$= \binom{n}{z}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(z+\alpha)\Gamma(n-z+\beta)}{\Gamma(\alpha+\beta+n)}$$

This then gives:

$$\mathrm{k}(\theta|z) = \frac{\binom{n}{z}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{z+\alpha-1}(1-\theta)^{n-z+\beta-1}}{\binom{n}{z}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(z+\alpha)\Gamma(n-z+\beta)}{\Gamma(\alpha+\beta+n)}}$$

$$= \frac{\Gamma(\alpha+\beta+n)}{\Gamma(z+\alpha)\Gamma(n-z+\beta)}\theta^{z+\alpha-1}(1-\theta)^{n-z+\beta-1}$$

That is, $\mathrm{k}(\theta|z)$ has a beta distribution with parameters $z+\alpha$ and $n-z+\beta$.

2. Asymptotic Behavior

We first assume that if the portion of claims remaining open tends to a finite limit as the sample size increases then the expected adjustment in (3.16) tends to unity. With this assumption, then, we consider

$$\lim_{n\to\infty}\left(\frac{n+2}{z+1}\right)^s \frac{\Gamma(n+2)\Gamma(z+s+1)}{\Gamma(z+1)\Gamma(s+n+2)}$$

For this evaluation we will use Stirling's approximation for the gamma function for large values of n:

$$\Gamma(n+1) \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

Using this approximation we have:

$$\left(\frac{n+2}{z+1}\right)^s \frac{\Gamma(n+2)\Gamma(z+s+1)}{\Gamma(z+1)\Gamma(s+n+2)} \approx \left(\frac{n+2}{z+1}\right)^s \frac{\sqrt{2\pi(n+1)}\left(\frac{n+1}{e}\right)^{n+1}\sqrt{2\pi(z+s)}\left(\frac{z+s}{e}\right)^{z+s}}{\sqrt{2\pi z}\left(\frac{z}{e}\right)^z\sqrt{2\pi(s+n+1)}\left(\frac{s+n+1}{e}\right)^{s+n+1}}$$

$$= \left(\frac{n+2}{z+1}\right)^s \frac{(n+1)^{n+\frac{3}{2}}(z+s)^{z+s+\frac{1}{2}}e^{-(n+z+s+1)}}{z^{z+\frac{1}{2}}(s+n+1)^{s+n+\frac{3}{2}}e^{-(z+s+n+1)}}$$

$$= \left(\frac{n+2}{z+1}\right)^s \left(\frac{n+1}{s+n+1}\right)^{n+\frac{3}{2}} \left(\frac{1}{s+n+1}\right)^s \left(\frac{z+s}{z}\right)^{z+\frac{1}{2}} (z+s)^s$$

$$= \left(\frac{(n+2)(z+s)}{(z+1)(s+n+1)}\right)^s \left(\frac{n+1}{s+n+1}\right)^{n+\frac{3}{2}} \left(\frac{z+s}{z}\right)^{z+\frac{1}{2}}$$

As n gets large we have:

$$\lim_{n\to\infty}\left(\frac{(n+2)(z+s)}{(z+1)(s+n+1)}\right)^s = \lim_{n\to\infty}\left(\frac{\left(1+\frac{2}{n}\right)\left(\frac{z}{n}+\frac{s}{n}\right)}{\left(\frac{z}{n}+\frac{1}{n}\right)\left(1+\frac{s+1}{n}\right)}\right)^s$$

$$= \left(\frac{\lim_{n\to\infty}\left(1+\frac{2}{n}\right)\lim_{n\to\infty}\left(\frac{z}{n}+\frac{s}{n}\right)}{\lim_{n\to\infty}\left(\frac{z}{n}+\frac{1}{n}\right)\lim_{n\to\infty}\left(1+\frac{s+1}{n}\right)}\right)^s$$

$$= \left(\frac{1\times\theta}{\theta\times 1}\right)^s = 1$$

The limits for the other two terms follow from an alternative definition for the exponential function:

$$e^z = \lim_{n\to\infty}\left(1+\frac{z}{n}\right)^n$$

We thus have:

$$\lim_{n\to\infty}\left(\frac{n+1}{s+n+1}\right)^{n+\frac{3}{2}} = \lim_{n\to\infty}\left(\frac{n+1}{s+n+1}\right)^{\frac{1}{2}-s}\left(1-\frac{s}{s+n+1}\right)^{s+n+1}$$

$$= \left(\lim_{n\to\infty}\left(\frac{1+\frac{1}{n}}{1+\frac{s+1}{n}}\right)^{\frac{1}{2}-s}\right)\lim_{n\to\infty}\left(1-\frac{s}{s+n+1}\right)^{s+n+1} = e^{-s}$$

Similarly we have:

$$\lim_{n\to\infty}\left(\frac{z+s}{z}\right)^{z+\frac{1}{2}} = \lim_{z\to\infty}\sqrt{\frac{z+s}{z}}\left(1+\frac{s}{z}\right)^{z}$$

$$= \sqrt{\lim_{z\to\infty}\left(1+\frac{s}{z}\right)}\lim_{z\to\infty}\left(1+\frac{s}{z}\right)^{z} = e^{s}$$

Thus we obtain:

$$\lim_{n\to\infty}\left(\frac{n+2}{z+1}\right)^{s}\frac{\Gamma(n+2)\Gamma(z+s+1)}{\Gamma(z+1)\Gamma(s+n+2)} = e^{-s}e^{s} = 1$$

APPENDIX C

In this appendix we derive formulae (3.17), using repeated application of the relationships in (3.16). First, we consider (3.16). From the definitions of the conditional distributions it is clear that

$$\begin{aligned} E(Z) &= \int\int z\, f(z|\xi)\, g(\xi) dz d\xi \\ &= E_\xi(E(Z|\xi)) \end{aligned}$$

As for the variance we have

$$\begin{aligned} Var(Z) &= \int\int z^2 f(z|\xi)\, g(\xi) dz d\xi - E(Z)^2 \\ &= E_\xi(E(Z^2|\xi)) - E(Z)^2 \\ &= E_\xi\left(Var(Z|\xi) + E(Z|\xi)^2\right) - E_\xi(E(Z|\xi))^2 \\ &= E_\xi(Var(Z|\xi)) + E_\xi\left(E(Z|\xi)^2 - E_\xi(E(Z|\xi))^2\right) \\ &= E_\xi(Var(Z|\xi)) + Var_\xi(E(Z|\xi)) \end{aligned}$$

From our assumptions we have

$$E(T|\theta,u,w_t) = \sum_x \sum_{s=1}^{\infty} b^*_{xs}\left(\prod_{l=0}^{s-1}(1-q^*_{xl})\right)\left(uy^s \prod_{r=1}^{s} w_r\right)$$

Similarly, we can compute

$$\begin{aligned} Var(T|\theta,u,w_t) &= Var\left(\sum_x X|\theta,u,w_t\right) \\ &= \sum_x Var(X|\theta,u,w_t) \end{aligned}$$

The last sum holds since we assumed the claims are independent for <u>fixed</u> θ, u, and w_t. We thus need only consider the variance for a single claim. We thus have:

$$Var(X|\theta,u,w_t) = E(X^2|\theta,u,w_t) - E(X|\theta,u,w_t)^2$$

From this we have

$$
\begin{aligned}
\mathrm{E}\left(X^2 \middle| \theta, u, w_t\right) &= \sum_{t=1}^{\infty} p_{xt} \left(\sum_{s=1}^{t} b^{*}_{xs} u \prod_{r=1}^{s} w_r \right)^2 \\
&= \sum_{t=1}^{\infty} \sum_{s=1}^{t} \sum_{r=1}^{t} p_{xt} b^{*}_{xs} b^{*}_{xr} u^2 \prod_{q=1}^{s} w_q \prod_{z=1}^{r} w_z \\
&= \sum_{s=1}^{\infty} \sum_{r=1}^{\infty} \sum_{t=\max(s,r)}^{\infty} p_{xt} b^{*}_{xs} b^{*}_{xr} u^2 \prod_{q=1}^{s} w_q \prod_{z=1}^{r} w_z \\
&= \sum_{s=1}^{\infty} \sum_{r=1}^{\infty} b^{*}_{xs} b^{*}_{xr} u^2 \prod_{q=1}^{s} w_q \prod_{z=1}^{r} w_z \sum_{t=\max(s,r)}^{\infty} p_{xt} \\
&= \sum_{s=1}^{\infty} \left(\sum_{r=1}^{s} b^{*}_{xs} b^{*}_{xr} u^2 \prod_{q=1}^{s} w_q \prod_{z=1}^{r} w_z \sum_{t=s}^{\infty} p_{xt} \right) + \left(\sum_{r=s+1}^{\infty} b^{*}_{xs} b^{*}_{xr} u^2 \prod_{q=1}^{s} w_q \prod_{z=1}^{r} w_z \sum_{t=r}^{\infty} p_{xt} \right) \\
&= \sum_{s=1}^{\infty} \left(\sum_{r=1}^{s} b^{*}_{xs} b^{*}_{xr} u^2 \prod_{q=1}^{r} w_q^2 \prod_{z=r+1}^{s} w_z \sum_{t=s}^{\infty} p_{xt} \right) + \left(\sum_{r=s+1}^{\infty} b^{*}_{xs} b^{*}_{xr} u^2 \prod_{q=1}^{s} w_q^2 \prod_{z=s+1}^{r} w_z \sum_{t=r}^{\infty} p_{xt} \right)
\end{aligned}
$$

We also have

$$
\begin{aligned}
\mathrm{E}\left(X \middle| \theta, u, w_t\right)^2 &= \left(\sum_{t=1}^{\infty} p_{xt} \sum_{s=1}^{t} b^{*}_{xs} u \prod_{r=1}^{s} w_r \right)^2 \\
&= \left(\sum_{s=1}^{\infty} u \prod_{r=1}^{s} w_r \sum_{t=s}^{\infty} p_{xt} b^{*}_{xs} \right)^2 \\
&= \sum_{s=1}^{\infty} \sum_{r=1}^{\infty} \left(u \prod_{q=1}^{s} w_q \sum_{t=s}^{\infty} p_{xt} b^{*}_{xs} \right) \left(u \prod_{z=1}^{r} w_z \sum_{t=r}^{\infty} p_{xt} b^{*}_{xr} \right) \\
&= \sum_{s=1}^{\infty} \sum_{r=1}^{\infty} \left(u^2 \prod_{q=1}^{s} w_q \prod_{z=1}^{r} w_z \right) \left(b^{*}_{xs} \sum_{t=s}^{\infty} p_{xt} \right) \left(b^{*}_{xr} \sum_{t=r}^{\infty} p_{xt} \right) \\
&= \sum_{s=1}^{\infty} \left(\sum_{r=1}^{s} b^{*}_{xs} b^{*}_{xr} u^2 \prod_{q=1}^{r} w_q^2 \prod_{z=r+1}^{s} w_z \left(\sum_{t=s}^{\infty} p_{xt} \right) \left(\sum_{t=r}^{\infty} p_{xt} \right) \right) + \left(\sum_{r=s+1}^{\infty} b^{*}_{xs} b^{*}_{xr} u^2 \prod_{q=1}^{s} w_q^2 \prod_{z=s+1}^{r} w_z \left(\sum_{t=s}^{\infty} p_{xt} \right) \left(\sum_{t=r}^{\infty} p_{xt} \right) \right)
\end{aligned}
$$

Thus

$$
\begin{aligned}
\mathrm{E}\left(X^2 \middle| \theta, u, w_t\right) - \mathrm{E}\left(X \middle| \theta, u, w_t\right)^2 &= \sum_{s=1}^{\infty} \left(\sum_{r=1}^{s} b^{*}_{xs} b^{*}_{xr} u^2 \prod_{q=1}^{r} w_q^2 \prod_{z=r+1}^{s} w_z \left(\sum_{t=s}^{\infty} p_{xt} \right) \left(1 - \sum_{t=r}^{\infty} p_{xt} \right) \right) \\
&\quad + \left(\sum_{r=s+1}^{\infty} b^{*}_{xs} b^{*}_{xr} u^2 \prod_{q=1}^{s} w_q^2 \prod_{z=s+1}^{r} w_z \left(\sum_{t=r}^{\infty} p_{xt} \right) \left(1 - \sum_{t=s}^{\infty} p_{xt} \right) \right) \\
&= \sum_{s=1}^{\infty} \sum_{r=1}^{\infty} b^{*}_{xs} b^{*}_{xr} u^2 \prod_{q=1}^{\min(r,s)} w_q^2 \prod_{z=\min(r,s)+1}^{\max(r,s)} w_z \left(\sum_{t=\max(r,s)}^{\infty} p_{xt} \right) \left(1 - \sum_{t=\min(r,s)}^{\infty} p_{xt} \right) \\
&= \sum_{s=1}^{\infty} \sum_{r=1}^{\infty} b^{*}_{xs} b^{*}_{xr} u^2 \prod_{q=1}^{\min(r,s)} w_q^2 \prod_{z=\min(r,s)+1}^{\max(r,s)} w_z \left(y^{\max(r,s)} \prod_{t=0}^{\max(r,s)-1} \left(1 - q^{*}_{xt}\right) \right) \left(1 - \left(y^{\min(r,s)} \prod_{t=0}^{\min(r,s)-1} \left(1 - q^{*}_{xt}\right) \right) \right)
\end{aligned}
$$

Which gives:

$$\mathrm{Var}(T|\theta,u,w_t)=\sum_x\sum_{s=1}^{\infty}\sum_{r=1}^{\infty}b^{*}_{xs}b^{*}_{xr}u^2\prod_{q=1}^{\min(r,s)}w_q^2\prod_{z=\min(r,s)+1}^{\max(r,s)}w_z\left(y^{\max(r,s)}\prod_{l=0}^{\max(r,s)-1}(1-q^{*}_{xl})\right)\left(1-\left(y^{\min(r,s)}\prod_{l=0}^{\min(r,s)-1}(1-q^{*}_{xl})\right)\right)$$

We will apply the Bayesian relationships above for each variable in succession so to as to appropriately track the various dependencies. First we remove the θ dependence:

$$\begin{aligned}
\mathrm{E}(T|u,w_t)&=\mathrm{E}_\theta(\mathrm{E}(T|\theta,u,w_t))\\
&=\mathrm{E}_\theta\left(\sum_x\sum_{s=1}^{\infty}b^{*}_{xs}\left(\prod_{l=0}^{s-1}(1-q^{*}_{xl})\right)\left(uy^s\prod_{r=1}^{s}w_r\right)\right)\\
&=\sum_x\sum_{s=1}^{\infty}b^{*}_{xs}\left(\prod_{l=0}^{s-1}(1-q^{*}_{xl})\right)\left(u\prod_{r=1}^{s}w_r\right)\mathrm{E}(y^s)\\
&=\sum_x\sum_{s=1}^{\infty}b^{*}_{xs}\left(\prod_{l=0}^{s-1}(1-q^{*}_{xl})\right)\left(u\prod_{r=1}^{s}w_r\right)\prod_{i=0}^{s-1}\frac{(n+2)(z+i+1)}{(z+1)(n+i+2)}
\end{aligned}$$

In calculations, the second-to-last representation is probably easier to manage. The variance estimate follows too:

$$\mathrm{Var}(T|u,w_t)=\mathrm{Var}_\theta(\mathrm{E}(T|\theta,u,w_t))+\mathrm{E}_\theta(\mathrm{Var}(T|\theta,u,w_t))$$

From the above relationships.

$$\begin{aligned}
\mathrm{Var}_\theta(\mathrm{E}(T|\theta,u,w_t))&=\mathrm{Var}_\theta\left(\sum_x\sum_{s=1}^{\infty}b^{*}_{xs}\left(\prod_{l=0}^{s-1}(1-q^{*}_{xl})\right)\left(uy^s\prod_{r=1}^{s}w_r\right)\right)\\
&=\mathrm{Var}_\theta\left(\sum_{s=1}^{\infty}\sum_x b^{*}_{xs}\left(\prod_{l=0}^{s-1}(1-q^{*}_{xl})\right)\left(uy^s\prod_{r=1}^{s}w_r\right)\right)\\
&=\mathrm{Var}_\theta\left(\sum_{s=1}^{\infty}y^s\left(u\prod_{r=1}^{s}w_r\right)\sum_x b^{*}_{xs}\left(\prod_{l=0}^{s-1}(1-q^{*}_{xl})\right)\right)\\
&=\sum_{s=1}^{\infty}\mathrm{Var}_\theta\left(y^s\left(u\prod_{r=1}^{s}w_r\right)\sum_x b^{*}_{xs}\left(\prod_{l=0}^{s-1}(1-q^{*}_{xl})\right)\right)\\
&=\sum_{s=1}^{\infty}\left(u\prod_{r=1}^{s}w_r\right)^2\left(\sum_x b^{*}_{xs}\left(\prod_{l=0}^{s-1}(1-q^{*}_{xl})\right)\right)^2\mathrm{Var}(y^s)
\end{aligned}$$

From the definitions of y we have

$$\begin{aligned}\mathrm{Var}\left(y^{s}\right) &= \mathrm{E}\left(y^{2s}\right)-\mathrm{E}\left(y^{s}\right)^{2} \\ &= \prod_{i=0}^{2s-1}\frac{(n+2)(z+i+1)}{(z+1)(n+i+2)}-\left(\prod_{i=0}^{s-1}\frac{(n+2)(z+i+1)}{(z+1)(n+i+2)}\right)^{2} \\ &= \left(\prod_{i=0}^{s-1}\frac{(n+2)(z+i+1)}{(z+1)(n+i+2)}\right)\left(\prod_{i=s}^{2s-1}\frac{(n+2)(z+i+1)}{(z+1)(n+i+2)}-\prod_{i=0}^{s-1}\frac{(n+2)(z+i+1)}{(z+1)(n+i+2)}\right)\end{aligned}$$

Again, the first representation will probably be easier from a coding point of view. This then gives:

$$\mathrm{Var}_{\theta}\left(\mathrm{E}(T|\theta,u,w_{t})\right)=\sum_{s=1}^{\infty}\left(u\prod_{r=1}^{s}w_{r}\right)^{2}\left(\sum_{x}b^{*}_{xs}\left(\prod_{t=0}^{s-1}(1-q^{*}_{xt})\right)\right)^{2}\left(\mathrm{E}\left(y^{2s}\right)-\mathrm{E}\left(y^{s}\right)^{2}\right)$$

As for the other term,

$$\begin{aligned}\mathrm{E}_{\theta}\left(\mathrm{Var}(T|\theta,u,w_{t})\right) &= \mathrm{E}_{\theta}\left(\sum_{x}\sum_{s=1}^{\infty}\sum_{r=1}^{\infty}b^{*}_{xs}b^{*}_{xr}u^{2}\prod_{q=1}^{\min(r,s)}w_{q}^{2}\prod_{z=\min(r,s)+1}^{\max(r,s)}w_{z}\left(y^{\max(r,s)}\prod_{t=0}^{\max(r,s)-1}(1-q^{*}_{xt})\right)\left(1-\left(y^{\min(r,s)}\prod_{t=0}^{\min(r,s)-1}(1-q^{*}_{xt})\right)\right)\right) \\ &= \sum_{x}\sum_{s=1}^{\infty}\sum_{r=1}^{\infty}b^{*}_{xs}b^{*}_{xr}u^{2}\prod_{q=1}^{\min(r,s)}w_{q}^{2}\prod_{z=\min(r,s)+1}^{\max(r,s)}w_{z}\,\mathrm{E}_{\theta}\left(\left(y^{\max(r,s)}\prod_{t=0}^{\max(r,s)-1}(1-q^{*}_{xt})\right)\left(1-\left(y^{\min(r,s)}\prod_{t=0}^{\min(r,s)-1}(1-q^{*}_{xt})\right)\right)\right) \\ &= \sum_{x}\sum_{s=1}^{\infty}\sum_{r=1}^{\infty}b^{*}_{xs}b^{*}_{xr}u^{2}\prod_{q=1}^{\min(r,s)}w_{q}^{2}\prod_{z=\min(r,s)+1}^{\max(r,s)}w_{z}\left(\prod_{t=0}^{\max(r,s)-1}(1-q^{*}_{xt})\right)\times\left(\mathrm{E}\left(y^{\max(r,s)}\right)-\mathrm{E}\left(y^{r+s}\right)\left(\prod_{t=0}^{\min(r,s)-1}(1-q^{*}_{xt})\right)\right)\end{aligned}$$

Combining these two terms we have:

$$\begin{aligned}\mathrm{Var}(T|u,w_{t}) &= \sum_{s=1}^{\infty}\left(u\prod_{r=1}^{s}w_{r}\right)^{2}\left(\sum_{x}b^{*}_{xs}\left(\prod_{t=0}^{s-1}(1-q^{*}_{xt})\right)\right)^{2}\left(\mathrm{E}\left(y^{2s}\right)-\mathrm{E}\left(y^{s}\right)^{2}\right) \\ &+\sum_{x}\sum_{s=1}^{\infty}\sum_{r=1}^{\infty}b^{*}_{xs}b^{*}_{xr}u^{2}\prod_{q=1}^{\min(r,s)}w_{q}^{2}\prod_{z=\min(r,s)+1}^{\max(r,s)}w_{z}\left(\prod_{t=0}^{\max(r,s)-1}(1-q^{*}_{xt})\right)\left(\mathrm{E}_{\theta}\left(y^{\max(r,s)}\right)-\mathrm{E}_{\theta}\left(y^{r+s}\right)\left(\prod_{t=0}^{\min(r,s)-1}(1-q^{*}_{xt})\right)\right)\end{aligned}$$

Now removing w_{t} from the terms:

$$\begin{aligned}
E(T|u) &= E_{w_t}\left(E(T|u,w_t)\right)\\
&= E_{w_t}\left(\sum_x \sum_{s=1}^{\infty} b^{\bullet}_{xs}\left(\prod_{t=0}^{s-1}(1-q^{\bullet}_{xt})\right)\left(u\prod_{r=1}^{s} w_r\right)E(y^s)\right)\\
&= \sum_x \sum_{s=1}^{\infty} b^{\bullet}_{xs}u\left(\prod_{t=0}^{s-1}(1-q^{\bullet}_{xt})\right)E(y^s)E\left(\prod_{r=1}^{s} w_r\right)\\
&= \sum_x \sum_{s=1}^{\infty} b^{\bullet}_{xs}u\left(\prod_{t=0}^{s-1}(1-q^{\bullet}_{xt})\right)E(y^s)\\
&= \sum_x \sum_{s=1}^{\infty} b^{\bullet}_{xs}u\left(\prod_{t=0}^{s-1}(1-q^{\bullet}_{xt})\right)\prod_{i=0}^{s-1}\frac{(n+2)(z+i+1)}{(z+1)(n+i+2)}
\end{aligned}$$

Again the second-to-last term may be easier to work with computationally. This last item follows from the lognormal assumptions regarding w_t. In particular these assumptions imply that:

$$\left(\prod_{r=1}^{s} w_r\right)^p \sim \text{lognormal}\left((-\tfrac{1}{2})(ps\tau^2), p^2 s\tau^2\right)$$

Thus the expectation:

$$E\left(\left(\prod_{r=1}^{s} w_r\right)^p\right) = \exp\left((-\tfrac{1}{2})(sp\tau^2) + \tfrac{1}{2}sp^2\tau^2\right)$$

For p = 1 this gives an expectation of unity, giving the above formula. Similarly for the variance term:

$$\text{Var}(T|u) = \text{Var}_{w_t}\left(E(T|u,w_t)\right) + E_{w_t}\left(\text{Var}(T|u,w_t)\right)$$

Taking the terms one at a time:

$$
\begin{aligned}
\operatorname{Var}_{w_t}\left(\mathrm{E}(T|u,w_t)\right) &= \operatorname{Var}_{w_t}\left(\sum_x \sum_{s=1}^{\infty} b_{xs}^{\bullet}\left(\prod_{t=0}^{s-1}(1-q_{xt}^{\bullet})\right)\left(u\prod_{r=1}^{s} w_r\right)\mathrm{E}(y^s)\right) \\
&= \operatorname{Var}_{w_t}\left(\sum_{s=1}^{\infty}\left(\prod_{r=1}^{s} w_r\right)\sum_x b_{xs}^{\bullet}\left(\prod_{t=0}^{s-1}(1-q_{xt}^{\bullet})\right)u\,\mathrm{E}(y^s)\right) \\
&= \sum_{s=1}^{\infty}\operatorname{Var}_{w_t}\left(\left(\prod_{r=1}^{s} w_r\right)\sum_x b_{xs}^{\bullet}\left(\prod_{t=0}^{s-1}(1-q_{xt}^{\bullet})\right)u\,\mathrm{E}(y^s)\right) \\
&= \sum_{s=1}^{\infty}\left(\sum_x b_{xs}^{\bullet}\left(\prod_{t=0}^{s-1}(1-q_{xt}^{\bullet})\right)u\,\mathrm{E}(y^s)\right)^2 \operatorname{Var}\left(\prod_{r=1}^{s} w_r\right)
\end{aligned}
$$

Again, from the lognormal assumptions:

$$
\begin{aligned}
\operatorname{Var}\left(\prod_{r=1}^{s} w_r\right) &= \mathrm{E}\left(\left(\prod_{r=1}^{s} w_r\right)^2\right) - \mathrm{E}\left(\prod_{r=1}^{s} w_r\right)^2 \\
&= \exp\left(\left(-\tfrac{1}{2}\right)\left(2s\tau^2\right)+\tfrac{1}{2}\left(4s\tau^2\right)\right) - \exp\left(2\left(\left(-\tfrac{1}{2}\right)\left(s\tau^2\right)+\tfrac{1}{2}s\tau^2\right)\right) \\
&= \exp\left(s\tau^2\right)-1
\end{aligned}
$$

We thus obtain:

$$
\operatorname{Var}_{w_t}\left(\mathrm{E}(T|u,w_t)\right) = \sum_{s=1}^{\infty}\left(\sum_x b_{xs}^{\bullet}\left(\prod_{t=0}^{s-1}(1-q_{xt}^{\bullet})\right)u\,\mathrm{E}(y^s)\right)^2\left(\exp\left(s\tau^2\right)-1\right)
$$

As for the second term we have:

$$
\begin{aligned}
\mathrm{E}_{w_t}\left(\operatorname{Var}(T|u,w_t)\right) &= \mathrm{E}_{w_t}\left(\begin{aligned}&\sum_{s=1}^{\infty}\left(u\prod_{r=1}^{s} w_r\right)^2\left(\sum_x b_{xs}^{\bullet}\left(\prod_{t=0}^{s-1}(1-q_{xt}^{\bullet})\right)\right)^2\left(\mathrm{E}(y^{2s})-\mathrm{E}(y^s)^2\right) \\ &+\sum_x\sum_{s=1}^{\infty}\sum_{r=1}^{\infty} b_{xs}^{\bullet}b_{xr}^{\bullet}u^2\prod_{q=1}^{\min(r,s)} w_q^2\prod_{z=\min(r,s)+1}^{\max(r,s)} w_z\left(\prod_{t=0}^{\max(r,s)-1}(1-q_{xt}^{\bullet})\right)\left(\mathrm{E}\left(y^{\max(r,s)}\right)-\mathrm{E}\left(y^{r+s}\right)\left(\prod_{t=0}^{\min(r,s)-1}(1-q_{xt}^{\bullet})\right)\right)\end{aligned}\right) \\
&= \sum_{s=1}^{\infty}\mathrm{E}\left(\left(\prod_{r=1}^{s} w_r\right)^2\right)u^2\left(\sum_x b_{xs}^{\bullet}\left(\prod_{t=0}^{s-1}(1-q_{xt}^{\bullet})\right)\right)^2\left(\mathrm{E}(y^{2s})-\mathrm{E}(y^s)^2\right) \\
&\quad+\sum_x\sum_{s=1}^{\infty}\sum_{r=1}^{\infty} b_{xs}^{\bullet}b_{xr}^{\bullet}u^2\,\mathrm{E}\left(\prod_{q=1}^{\min(r,s)} w_q^2\prod_{z=\min(r,s)+1}^{\max(r,s)} w_z\right)\left(\prod_{t=0}^{\max(r,s)-1}(1-q_{xt}^{\bullet})\right)\left(\mathrm{E}\left(y^{\max(r,s)}\right)-\mathrm{E}\left(y^{r+s}\right)\left(\prod_{t=0}^{\min(r,s)-1}(1-q_{xt}^{\bullet})\right)\right)
\end{aligned}
$$

Now from the lognormality assumptions

$$\prod_{q=1}^{m} w_q^2 \prod_{z=m+1}^{n} w_z \sim \text{lognormal}\left(\left(-\tfrac{1}{2}\right)\left(2m\tau^2 + (n-m)\tau^2\right), 4m\tau^2 + (n-m)\tau^2\right), \text{ thus}$$

$$\prod_{q=1}^{m} w_q^2 \prod_{z=m+1}^{n} w_z \sim \text{lognormal}\left(\left(-\tfrac{1}{2}\right)\left((n+m)\tau^2\right), (n+3m)\tau^2\right)$$

This then gives

$$\begin{aligned} \mathrm{E}\left(\prod_{q=1}^{m} w_q^2 \prod_{z=m+1}^{n} w_z\right) &= \exp\left(\left(-\tfrac{1}{2}\right)\left((n+m)\tau^2\right) + \tfrac{1}{2}\left((n+3m)\tau^2\right)\right) \\ &= \exp\left(m\tau^2\right) \end{aligned}$$

This results in:

$$\begin{aligned} \mathrm{E}_{w_1}\left(\mathrm{Var}(T|u,w_1)\right) = {} & \sum_{s=1}^{\infty} \exp\left(s\tau^2\right) u^2 \left(\sum_x b^*_{xs}\left(\prod_{t=0}^{s-1}\left(1-q^*_{x1}\right)\right)\right)^2 \left(\mathrm{E}\left(y^{2s}\right) - \mathrm{E}\left(y^s\right)^2\right) \\ & + \sum_x \sum_{s=1}^{\infty} \sum_{r=1}^{\infty} b^*_{xs} b^*_{xr} u^2 \exp\left(\min(r,s)\tau^2\right) \left(\prod_{t=0}^{\max(r,s)-1}\left(1-q^*_{xt}\right)\right)\left(\mathrm{E}\left(y^{\max(r,s)}\right) - \mathrm{E}\left(y^{r+s}\right)\left(\prod_{t=0}^{\min(r,s)-1}\left(1-q^*_{xt}\right)\right)\right) \end{aligned}$$

Adding these two terms together we obtain:

$$\begin{aligned} \mathrm{Var}(T|u) = {} & u^2 \sum_{s=1}^{\infty}\left(\sum_x b^*_{xs}\left(\prod_{t=0}^{s-1}\left(1-q^*_{xt}\right)\right)\mathrm{E}\left(y^s\right)\right)^2 \left(\exp\left(s\tau^2\right)-1\right) \\ & + u^2 \sum_{s=1}^{\infty} \exp\left(s\tau^2\right)\left(\sum_x b^*_{xs}\left(\prod_{t=0}^{s-1}\left(1-q^*_{xt}\right)\right)\right)^2 \left(\mathrm{E}\left(y^{2s}\right) - \mathrm{E}\left(y^s\right)^2\right) \\ & + u^2 \sum_x \sum_{s=1}^{\infty} \sum_{r=1}^{\infty} b^*_{xs} b^*_{xr} \exp\left(\min(r,s)\tau^2\right)\left(\prod_{t=0}^{\max(r,s)-1}\left(1-q^*_{xt}\right)\right)\left(\mathrm{E}\left(y^{\max(r,s)}\right) - \mathrm{E}\left(y^{r+s}\right)\left(\prod_{t=0}^{\min(r,s)-1}\left(1-q^*_{xt}\right)\right)\right) \end{aligned}$$

Finally we eliminate u from the formulae. First

$$
\begin{aligned}
\mathrm{E}(T) &= \mathrm{E}_u\left(\mathrm{E}(T|u)\right) \\
&= \mathrm{E}_u\left(\sum_x \sum_{s=1}^{\infty} b^{\bullet}_{xs} u \left(\prod_{t=0}^{s-1}\left(1-q^{\bullet}_{xt}\right)\right)\mathrm{E}\left(y^s\right)\right) \\
&= \sum_x \sum_{s=1}^{\infty} b^{\bullet}_{xs} \left(\prod_{t=0}^{s-1}\left(1-q^{\bullet}_{xt}\right)\right)\mathrm{E}\left(y^s\right)\mathrm{E}(u) \\
&= \sum_x \sum_{s=1}^{\infty} b^{\bullet}_{xs} \left(\prod_{t=0}^{s-1}\left(1-q^{\bullet}_{xt}\right)\right)\mathrm{E}\left(y^s\right) \\
&= \sum_x \sum_{s=1}^{\infty} b^{\bullet}_{xs} \left(\prod_{t=0}^{s-1}\left(1-q^{\bullet}_{xt}\right)\right)\prod_{i=0}^{s-1}\frac{(n+2)(z+i+1)}{(z+1)(n+i+2)}
\end{aligned}
$$

As for the variance formula we have:

$$
\mathrm{Var}(T) = \mathrm{Var}_u\left(\mathrm{E}(T|u)\right) + \mathrm{E}_u\left(\mathrm{Var}(T|u)\right)
$$

Again we consider the two portions separately.

$$
\begin{aligned}
\mathrm{Var}_u\left(\mathrm{E}(T|u)\right) &= \mathrm{Var}_u\left(\sum_x \sum_{s=1}^{\infty} b^{\bullet}_{xs} u \left(\prod_{t=0}^{s-1}\left(1-q^{\bullet}_{xt}\right)\right)\mathrm{E}\left(y^s\right)\right) \\
&= \mathrm{Var}_u\left(u\sum_x \sum_{s=1}^{\infty} b^{\bullet}_{xs} \left(\prod_{t=0}^{s-1}\left(1-q^{\bullet}_{xt}\right)\right)\mathrm{E}\left(y^s\right)\right) \\
&= \left(\sum_x \sum_{s=1}^{\infty} b^{\bullet}_{xs} \left(\prod_{t=0}^{s-1}\left(1-q^{\bullet}_{xt}\right)\right)\mathrm{E}\left(y^s\right)\right)^2 \mathrm{Var}(u)
\end{aligned}
$$

Since u is lognormal with mean 1 we have

$$
\mathrm{Var}(u) = \exp\left(\sigma^2\right) - 1
$$

We thus obtain:

$$
\mathrm{Var}_u\left(\mathrm{E}(T|u)\right) = \left(\sum_x \sum_{s=1}^{\infty} b^{\bullet}_{xs} \left(\prod_{t=0}^{s-1}\left(1-q^{\bullet}_{xt}\right)\right)\mathrm{E}\left(y^s\right)\right)^2 \left(\exp\left(\sigma^2\right) - 1\right)
$$

As for the second term we have:

$$\mathrm{E}_u\left(\mathrm{Var}(T|u)\right) = \mathrm{E}_u\left(\begin{array}{l} u^2\sum_{s=1}^{\infty}\left(\sum_x b_{xs}^*\left(\prod_{t=0}^{s-1}(1-q_{xt}^*)\right)\mathrm{E}(y^s)\right)^2\left(\exp(s\tau^2)-1\right) \\ +u^2\sum_{s=1}^{\infty}\exp(s\tau^2)\left(\sum_x b_{xs}^*\left(\prod_{t=0}^{s-1}(1-q_{xt}^*)\right)\right)^2\left(\mathrm{E}(y^{2s})-\mathrm{E}(y^s)^2\right) \\ +u^2\sum_x\sum_{s=1}^{\infty}\sum_{r=1}^{\infty}b_{xs}^*b_{xr}^*\exp(\min(r,s)\tau^2)\left(\prod_{t=0}^{\max(r,s)-1}(1-q_{xt}^*)\right)\left(\mathrm{E}(y^{\max(r,s)})-\mathrm{E}(y^{r+s})\left(\prod_{t=0}^{\min(r,s)-1}(1-q_{xt}^*)\right)\right)\end{array}\right)$$

$$= \mathrm{E}(u^2)\sum_{s=1}^{\infty}\left(\sum_x b_{xs}^*\left(\prod_{t=0}^{s-1}(1-q_{xt}^*)\right)\mathrm{E}(y^s)\right)^2\left(\exp(s\tau^2)-1\right)$$

$$+\mathrm{E}(u^2)\sum_{s=1}^{\infty}\exp(s\tau^2)\left(\sum_x b_{xs}^*\left(\prod_{t=0}^{s-1}(1-q_{xt}^*)\right)\right)^2\left(\mathrm{E}(y^{2s})-\mathrm{E}(y^s)^2\right)$$

$$+\mathrm{E}(u^2)\sum_x\sum_{s=1}^{\infty}\sum_{r=1}^{\infty}b_{xs}^*b_{xr}^*\exp(\min(r,s)\tau^2)\left(\prod_{t=0}^{\max(r,s)-1}(1-q_{xt}^*)\right)\left(\mathrm{E}(y^{\max(r,s)})-\mathrm{E}(y^{r+s})\left(\prod_{t=0}^{\min(r,s)-1}(1-q_{xt}^*)\right)\right)$$

Since u is lognormal, i.e.

$$u \sim \mathrm{lognormal}\left(-\tfrac{1}{2}\sigma^2,\sigma^2\right)$$

then u^2 is also lognormal and

$$u^2 \sim \mathrm{lognormal}\left(-\sigma^2,4\sigma^2\right)$$

Thus we have

$$\mathrm{E}(u^2)=\exp\left((-\sigma^2)+\tfrac{1}{2}(4\sigma^2)\right)=\exp(\sigma^2)$$

This then gives:

$$
\begin{aligned}
E_u\left(\mathrm{Var}(T|u)\right) = {} & \exp\left(\sigma^2\right)\sum_{s=1}^{\infty}\left(\sum_x b^*_{xs}\left(\prod_{t=0}^{s-1}\left(1-q^*_{xt}\right)\right)E\left(y^s\right)\right)^2\left(\exp\left(s\tau^2\right)-1\right) \\
& +\exp\left(\sigma^2\right)\sum_{s=1}^{\infty}\exp\left(s\tau^2\right)\left(\sum_x b^*_{xs}\left(\prod_{t=0}^{s-1}\left(1-q^*_{xt}\right)\right)\right)^2\left(E\left(y^{2s}\right)-E\left(y^s\right)^2\right) \\
& +\exp\left(\sigma^2\right)\sum_x\sum_{s=1}^{\infty}\sum_{r=1}^{\infty} b^*_{xs}b^*_{xr}\exp\left(\min(r,s)\tau^2\right)\left(\prod_{t=0}^{\max(r,s)-1}\left(1-q^*_{xt}\right)\right)\left(E\left(y^{\max(r,s)}\right)-E\left(y^{r+s}\right)\left(\prod_{t=0}^{\min(r,s)-1}\left(1-q^*_{xt}\right)\right)\right) \\
= {} & \exp\left(\sigma^2\right)\sum_{s=1}^{\infty}\exp\left(s\tau^2\right)\left(\sum_x b^*_{xs}\left(\prod_{t=0}^{s-1}\left(1-q^*_{xt}\right)\right)\right)^2 E\left(y^s\right)^2-\exp\left(\sigma^2\right)\sum_{s=1}^{\infty}\left(\sum_x b^*_{xs}\left(\prod_{t=0}^{s-1}\left(1-q^*_{xt}\right)\right)\right)^2 E\left(y^s\right)^2 \\
& +\exp\left(\sigma^2\right)\sum_{s=1}^{\infty}\exp\left(s\tau^2\right)\left(\sum_x b^*_{xs}\left(\prod_{t=0}^{s-1}\left(1-q^*_{xt}\right)\right)\right)^2 E\left(y^{2s}\right)-\exp\left(\sigma^2\right)\sum_{s=1}^{\infty}\exp\left(s\tau^2\right)\left(\sum_x b^*_{xs}\left(\prod_{t=0}^{s-1}\left(1-q^*_{xt}\right)\right)\right)^2 E\left(y^s\right)^2 \\
& +\exp\left(\sigma^2\right)\sum_x\sum_{s=1}^{\infty}\sum_{r=1}^{\infty} b^*_{xs}b^*_{xr}\exp\left(\min(r,s)\tau^2\right)\left(\prod_{t=0}^{\max(r,s)-1}\left(1-q^*_{xt}\right)\right)\left(E\left(y^{\max(r,s)}\right)-E\left(y^{r+s}\right)\left(\prod_{t=0}^{\min(r,s)-1}\left(1-q^*_{xt}\right)\right)\right) \\
= {} & \exp\left(\sigma^2\right)\sum_{s=1}^{\infty}\left(\sum_x b^*_{xs}\left(\prod_{t=0}^{s-1}\left(1-q^*_{xt}\right)\right)\right)^2\left(\exp\left(s\tau^2\right)E\left(y^{2s}\right)-E\left(y^s\right)^2\right) \\
& +\exp\left(\sigma^2\right)\sum_x\sum_{s=1}^{\infty}\sum_{r=1}^{\infty} b^*_{xs}b^*_{xr}\exp\left(\min(r,s)\tau^2\right)\left(\prod_{t=0}^{\max(r,s)-1}\left(1-q^*_{xt}\right)\right)\left(E\left(y^{\max(r,s)}\right)-E\left(y^{r+s}\right)\left(\prod_{t=0}^{\min(r,s)-1}\left(1-q^*_{xt}\right)\right)\right)
\end{aligned}
$$

Finally putting the two terms together we obtain:

$$
\begin{aligned}
\mathrm{Var}(T) = {} & \left(\sum_x\sum_{s=1}^{\infty} b^*_{xs}\left(\prod_{t=0}^{s-1}\left(1-q^*_{xt}\right)\right)E\left(y^s\right)\right)^2\left(\exp\left(\sigma^2\right)-1\right) \\
& +\exp\left(\sigma^2\right)\sum_{s=1}^{\infty}\left(\sum_x b^*_{xs}\left(\prod_{t=0}^{s-1}\left(1-q^*_{xt}\right)\right)\right)^2\left(E\left(y^{2s}\right)\exp\left(s\tau^2\right)-E\left(y^s\right)^2\right) \\
& +\exp\left(\sigma^2\right)\sum_x\sum_{s=1}^{\infty}\sum_{r=1}^{\infty} b^*_{xs}b^*_{xr}\exp\left(\min(r,s)\tau^2\right)\left(\prod_{t=0}^{\max(r,s)-1}\left(1-q^*_{xt}\right)\right)\left(E\left(y^{\max(r,s)}\right)-E\left(y^{r+s}\right)\left(\prod_{t=0}^{\min(r,s)-1}\left(1-q^*_{xt}\right)\right)\right)
\end{aligned}
$$

APPENDIX D

In this appendix we show the derivation of formulae (3.19) for payments in a particular year. As with the total mean and variance we begin with the mean and variance for fixed parameter values and then, step, by step, remove dependency on the various uncertainty parameters. Without any uncertainty and dropping the explicit i subscript, (2.7) and (2.8) give:

$$E(T_s|\text{Certainty}) = \sum_x a_{xs} \sum_{t=s}^{\infty} p_{xt}$$

$$\text{Var}(T_s|\text{Certainty}) = \sum_x a_{xs}^2 \left(\sum_{t=s}^{\infty} p_{xt} \right) \left(1 - \sum_{t=s}^{\infty} p_{xt} \right)$$

Thus, incorporating cost escalation, discounting, and our uncertainty variables, we have:

$$E(T_s|\theta, u, w_s) = \sum_x b^{*}_{xs} \left(\prod_{t=0}^{s-1} (1 - q^{*}_{xt}) \right) \left(u y^s \prod_{t=1}^{s} w_t \right)$$

$$= \left(u y^s \prod_{t=1}^{s} w_t \right) \sum_x b^{*}_{xs} \left(\prod_{t=0}^{s-1} (1 - q^{*}_{xt}) \right)$$

As with the aggregate,

$$\text{Var}(T_s|\theta, u, w_t) = \text{Var}\left(\sum_x X_s|\theta, u, w_t \right)$$

$$= \sum_x \text{Var}(X_s|\theta, u, w_t)$$

The last sum holds since we assumed the claims are independent for <u>fixed</u> θ, u, and w_t. We thus need only consider the variance for a single claim. We thus have:

$$\text{Var}(X_s|\theta, u, w_t) = E(X_s^2|\theta, u, w_t) - E(X_s|\theta, u, w_t)^2$$

Breaking this into parts then we have:

$$E(X_s^2|\theta, u, w_t) = \left(b^{*}_{xs} u \prod_{t=1}^{s} w_s \right)^2 \left(y^s \prod_{t=0}^{s-1} (1 - q^{*}_{xt}) \right)$$

and

$$E(X_s|\theta,u,w_t)^2 = \left(b^{\bullet}_{xs}u\prod_{r=1}^{s}w_s\right)^2\left(y^s\prod_{t=0}^{s-1}(1-q^{\bullet}_{xt})\right)^2$$

Hence we have

$$\begin{aligned}\mathrm{Var}(X_s|\theta,u,w_t) &= \left(b^{\bullet}_{xs}u\prod_{r=1}^{s}w_s\right)^2\left(y^s\prod_{t=0}^{s-1}(1-q^{\bullet}_{xt})\right)-\left(b^{\bullet}_{xs}u\prod_{r=1}^{s}w_s\right)^2\left(y^s\prod_{t=0}^{s-1}(1-q^{\bullet}_{xt})\right)^2\\ &= \left(b^{\bullet}_{xs}u\prod_{r=1}^{s}w_s\right)^2\left(y^s\prod_{t=0}^{s-1}(1-q^{\bullet}_{xt})\right)\left(1-y^s\prod_{t=0}^{s-1}(1-q^{\bullet}_{xt})\right)\end{aligned}$$

This then gives:

$$\mathrm{Var}(T_s|\theta,u,w_t) = \sum_x\left(b^{\bullet}_{xs}u\prod_{r=1}^{s}w_s\right)^2\left(y^s\prod_{t=0}^{s-1}(1-q^{\bullet}_{xt})\right)\left(1-y^s\prod_{t=0}^{s-1}(1-q^{\bullet}_{xt})\right)$$

We now use the Bayesian relationships to work down the conditional variables. First we remove the θ dependence.

$$\begin{aligned}E(T_s|u,w_s) &= E_\theta\left(\left(uy^s\prod_{r=1}^{s}w_r\right)\sum_x b^{\bullet}_{xs}\left(\prod_{t=0}^{s-1}(1-q^{\bullet}_{xt})\right)\right)\\ &= \left(u\prod_{r=1}^{s}w_r\right)\sum_x b^{\bullet}_{xs}\left(\prod_{t=0}^{s-1}(1-q^{\bullet}_{xt})\right)E(y^s)\\ &= \sum_x b^{\bullet}_{xs}\left(\prod_{t=0}^{s-1}(1-q^{\bullet}_{xt})\right)\left(u\prod_{r=1}^{s}w_r\right)\prod_{i=0}^{s-1}\frac{(n+2)(z+i+1)}{(z+1)(n+i+2)}\end{aligned}$$

In calculations, the second-to-last representation is probably easier to manage. The variance estimate follows too:

$$\mathrm{Var}(T_s|u,w_t) = \mathrm{Var}_\theta(E(T_s|\theta,u,w_t)) + E_\theta(\mathrm{Var}(T_s|\theta,u,w_t))$$

From the above relationships:

$$\mathrm{Var}_\theta\left(\mathrm{E}(T_s|\theta,u,w_t)\right) = \mathrm{Var}_\theta\left(\sum_x b^*_{xs}\left(\prod_{l=0}^{s-1}(1-q^*_{xl})\right)\left(uy^s\prod_{r=1}^{s}w_r\right)\right)$$

$$= \mathrm{Var}_\theta\left(y^s\left(u\prod_{r=1}^{s}w_r\right)\sum_x b^*_{xs}\left(\prod_{l=0}^{s-1}(1-q^*_{xl})\right)\right)$$

$$= \left(u\prod_{r=1}^{s}w_r\right)^2\left(\sum_x b^*_{xs}\left(\prod_{l=0}^{s-1}(1-q^*_{xl})\right)\right)^2 \mathrm{Var}(y^s)$$

From Appendix C we have:

$$\mathrm{Var}(y^s) = \left(\prod_{i=0}^{s-1}\frac{(n+2)(z+i+1)}{(z+1)(n+i+2)}\right)\left(\prod_{i=s}^{2s-1}\frac{(n+2)(z+i+1)}{(z+1)(n+i+2)} - \prod_{i=0}^{s-1}\frac{(n+2)(z+i+1)}{(z+1)(n+i+2)}\right)$$

Again, the first representation will probably be easier from a coding point of view. This then gives:

$$\mathrm{Var}_\theta\left(\mathrm{E}(T_s|\theta,u,w_t)\right) = \left(u\prod_{r=1}^{s}w_r\right)^2\left(\sum_x b^*_{xs}\left(\prod_{l=0}^{s-1}(1-q^*_{xl})\right)\right)^2\left(\mathrm{E}(y^{2s}) - \mathrm{E}(y^s)^2\right)$$

As for the other term,

$$\mathrm{E}_\theta\left(\mathrm{Var}(T_s|\theta,u,w_t)\right) = \mathrm{E}_\theta\left(\sum_x\left(b^*_{xs}u\prod_{r=1}^{s}w_s\right)^2\left(y^s\prod_{l=0}^{s-1}(1-q^*_{xl})\right)\left(1-y^s\prod_{l=0}^{s-1}(1-q^*_{xl})\right)\right)$$

$$= \sum_x\left(b^*_{xs}u\prod_{r=1}^{s}w_s\right)^2 \mathrm{E}_\theta\left(\left(y^s\prod_{l=0}^{s-1}(1-q^*_{xl})\right) - \left(y^{2s}\prod_{l=0}^{s-1}(1-q^*_{xl})^2\right)\right)$$

$$= \sum_x\left(b^*_{xs}u\prod_{r=1}^{s}w_s\right)^2\left(\prod_{l=0}^{s-1}(1-q^*_{xl})\right)\left(\mathrm{E}(y^s) - \mathrm{E}(y^{2s})\left(\prod_{l=0}^{s-1}(1-q^*_{xl})^2\right)\right)$$

Combining these two terms we have:

$$\mathrm{Var}(T_s|u,w_t) = \left(u\prod_{r=1}^{s}w_r\right)^2\left(\sum_x b^*_{xs}\left(\prod_{l=0}^{s-1}(1-q^*_{xl})\right)\right)^2\left(\mathrm{E}(y^{2s}) - \mathrm{E}(y^s)^2\right)$$

$$+\sum_x\left(b^*_{xs}u\prod_{r=1}^{s}w_s\right)^2\left(\prod_{l=0}^{s-1}(1-q^*_{xl})\right)\left(\mathrm{E}(y^s) - \mathrm{E}(y^{2s})\left(\prod_{l=0}^{s-1}(1-q^*_{xl})^2\right)\right)$$

Now removing w_t from the terms:

$$
\begin{aligned}
\mathrm{E}(T_s|u) &= \mathrm{E}_{w_t}\left(\mathrm{E}(T_s|u,w_t)\right)\\
&= \mathrm{E}_{w_t}\left(\left(u\prod_{r=1}^{s} w_r\right)\sum_x b_{xs}^{\bullet}\left(\prod_{t=0}^{s-1}(1-q_{xt}^{\bullet})\right)\mathrm{E}_{\theta}(y^s)\right)\\
&= \sum_x b_{xs}^{\bullet}u\left(\prod_{t=0}^{s-1}(1-q_{xt}^{\bullet})\right)\mathrm{E}(y^s)\mathrm{E}\left(\prod_{r=1}^{s} w_r\right)\\
&= \sum_x b_{xs}^{\bullet}u\left(\prod_{t=0}^{s-1}(1-q_{xt}^{\bullet})\right)\mathrm{E}(y^s)\\
&= \sum_x b_{xs}^{\bullet}u\left(\prod_{t=0}^{s-1}(1-q_{xt}^{\bullet})\right)\prod_{i=0}^{s-1}\frac{(n+2)(z+i+1)}{(z+1)(n+i+2)}
\end{aligned}
$$

Again the second-to-last term may be easier to work with computationally. This last item follows from the lognormal assumptions regarding w_r as in Appendix C.

Similarly for the variance term:

$$
\mathrm{Var}(T_s|u) = \mathrm{Var}_{w_t}\left(\mathrm{E}(T_s|u,w_t)\right) + \mathrm{E}_{w_t}\left(\mathrm{Var}(T_s|u,w_t)\right)
$$

Taking the terms one at a time, using the lognormal relationships in Appendix C:

$$
\begin{aligned}
\mathrm{Var}_{w_t}\left(\mathrm{E}(T_s|u,w_t)\right) &= \mathrm{Var}_{w_t}\left(\left(u\prod_{r=1}^{s} w_r\right)\sum_x b_{xs}^{\bullet}\left(\prod_{t=0}^{s-1}(1-q_{xt}^{\bullet})\right)\mathrm{E}(y^s)\right)\\
&= \mathrm{Var}_{w_t}\left(\left(\prod_{r=1}^{s} w_r\right)\sum_x b_{xs}^{\bullet}\left(\prod_{t=0}^{s-1}(1-q_{xt}^{\bullet})\right)u\mathrm{E}(y^s)\right)\\
&= \left(\sum_x b_{xs}^{\bullet}\left(\prod_{t=0}^{s-1}(1-q_{xt}^{\bullet})\right)u\mathrm{E}(y^s)\right)^2 \mathrm{Var}\left(\prod_{r=1}^{s} w_r\right)\\
&= \left(\sum_x b_{xs}^{\bullet}\left(\prod_{t=0}^{s-1}(1-q_{xt}^{\bullet})\right)u\mathrm{E}(y^s)\right)^2 \left(\exp(s\tau^2)-1\right)
\end{aligned}
$$

As for the second term we have:

$$\mathrm{E}_{w_t}\left(\mathrm{Var}(T_s|u,w_t)\right)=\mathrm{E}_{w_t}\left(\begin{array}{l}\left(u\prod_{t=1}^{s}w_t\right)^2\left(\sum_x b_{xs}^{*}\left(\prod_{t=0}^{s-1}(1-q_{xt}^{*})\right)\right)^2\left(\mathrm{E}(y^{2s})-\mathrm{E}(y^s)^2\right)\\+\sum_x\left(b_{xs}^{*}u\prod_{t=1}^{s}w_s\right)^2\left(\prod_{t=0}^{s-1}(1-q_{xt}^{*})\right)\left(\mathrm{E}(y^s)-\mathrm{E}(y^{2s})\left(\prod_{t=0}^{s-1}(1-q_{xt}^{*})^2\right)\right)\end{array}\right)$$

$$=\mathrm{E}\left(\left(\prod_{t=1}^{s}w_t\right)^2\right)u^2\left(\sum_x b_{xs}^{*}\left(\prod_{t=0}^{s-1}(1-q_{xt}^{*})\right)\right)^2\left(\mathrm{E}(y^{2s})-\mathrm{E}(y^s)^2\right)$$

$$+\sum_x b_{xs}^{*2}u^2\,\mathrm{E}\left(\left(\prod_{t=1}^{s}w_t\right)^2\right)\left(\prod_{t=0}^{s-1}(1-q_{xt}^{*})\right)\left(\mathrm{E}(y^s)-\mathrm{E}(y^{2s})\left(\prod_{t=0}^{s-1}(1-q_{xt}^{*})\right)\right)$$

$$=\exp(s\tau^2)u^2\left(\sum_x b_{xs}^{*}\left(\prod_{t=0}^{s-1}(1-q_{xt}^{*})\right)\right)^2\left(\mathrm{E}(y^{2s})-\mathrm{E}(y^s)^2\right)$$

$$+\sum_x b_{xs}^{*2}u^2\exp(s\tau^2)\left(\prod_{t=0}^{s-1}(1-q_{xt}^{*})\right)\left(\mathrm{E}(y^s)-\mathrm{E}(y^{2s})\left(\prod_{t=0}^{s-1}(1-q_{xt}^{*})\right)\right)$$

Adding these two terms together we obtain:

$$\mathrm{Var}(T_s|u)=\left(\sum_x b_{xs}^{*}\left(\prod_{t=0}^{s-1}(1-q_{xt}^{*})\right)u\,\mathrm{E}(y^s)\right)^2\left(\exp(s\tau^2)-1\right)$$

$$+\exp(s\tau^2)u^2\left(\sum_x b_{xs}^{*}\left(\prod_{t=0}^{s-1}(1-q_{xt}^{*})\right)\right)^2\left(\mathrm{E}(y^{2s})-\mathrm{E}(y^s)^2\right)$$

$$+\sum_x b_{xs}^{*2}u^2\exp(s\tau^2)\left(\prod_{t=0}^{s-1}(1-q_{xt}^{*})\right)\left(\mathrm{E}(y^s)-\mathrm{E}(y^{2s})\left(\prod_{t=0}^{s-1}(1-q_{xt}^{*})\right)\right)$$

Finally we eliminate u from the formulae. First

$$\mathrm{E}(T_s)=\mathrm{E}_u\left(\mathrm{E}(T_s|u)\right)$$

$$=\mathrm{E}_u\left(\sum_x b_{xs}^{*}u\left(\prod_{t=0}^{s-1}(1-q_{xt}^{*})\right)\mathrm{E}(y^s)\right)$$

$$=\sum_x b_{xs}^{*}\left(\prod_{t=0}^{s-1}(1-q_{xt}^{*})\right)\mathrm{E}(y^s)\mathrm{E}(u)$$

$$=\sum_x b_{xs}^{*}\left(\prod_{t=0}^{s-1}(1-q_{xt}^{*})\right)\mathrm{E}(y^s)$$

$$=\sum_x b_{xs}^{*}\left(\prod_{t=0}^{s-1}(1-q_{xt}^{*})\right)\prod_{i=0}^{s-1}\frac{(n+2)(z+i+1)}{(z+1)(n+i+2)}$$

As for the variance formula we have:

$$\text{Var}(T_s) = \text{Var}_u\left(\text{E}(T_s|u)\right) + \text{E}_u\left(\text{Var}(T_s|u)\right)$$

Again we consider the two portions separately.

$$\begin{aligned}
\text{Var}_u\left(\text{E}(T_s|u)\right) &= \text{Var}_u\left(\sum_x b^{\bullet}_{xs} u \left(\prod_{t=0}^{s-1}(1-q^{\bullet}_{xt})\right)\text{E}(y^s)\right) \\
&= \text{Var}_u\left(u \sum_x b^{\bullet}_{xs} u \left(\prod_{t=0}^{s-1}(1-q^{\bullet}_{xt})\right)\text{E}(y^s)\right) \\
&= \left(\sum_x b^{\bullet}_{xs}\left(\prod_{t=0}^{s-1}(1-q^{\bullet}_{xt})\right)\text{E}(y^s)\right)^2 \text{Var}(u) \\
&= \left(\sum_x b^{\bullet}_{xs}\left(\prod_{t=0}^{s-1}(1-q^{\bullet}_{xt})\right)\text{E}(y^s)\right)^2 \left(\exp(\sigma^2)-1\right)
\end{aligned}$$

As for the second term we have:

$$\text{E}_u\left(\text{Var}(T|u)\right) = \text{E}_u\left(\begin{aligned}
&\left(\sum_x b^{\bullet}_{xs}\left(\prod_{t=0}^{s-1}(1-q^{\bullet}_{xt})\right)u\text{E}(y^s)\right)^2\left(\exp(s\tau^2)-1\right) \\
&+\exp(s\tau^2)u^2\left(\sum_x b^{\bullet}_{xs}\left(\prod_{t=0}^{s-1}(1-q^{\bullet}_{xt})\right)\right)^2\left(\text{E}(y^{2s})-\text{E}(y^s)^2\right) \\
&+\sum_x b^{\bullet 2}_{xs}u^2\exp(s\tau^2)\left(\prod_{t=0}^{s-1}(1-q^{\bullet}_{xt})\right)\left(\text{E}(y^s)-\text{E}(y^{2s})\left(\prod_{t=0}^{s-1}(1-q^{\bullet}_{xt})\right)\right)
\end{aligned}\right)$$

$$\begin{aligned}
&= \text{E}(u^2)\left(\sum_x b^{\bullet}_{xs}\left(\prod_{t=0}^{s-1}(1-q^{\bullet}_{xt})\right)\text{E}(y^s)\right)^2\left(\exp(s\tau^2)-1\right) \\
&+\text{E}(u^2)\exp(s\tau^2)\left(\sum_x b^{\bullet}_{xs}\left(\prod_{t=0}^{s-1}(1-q^{\bullet}_{xt})\right)\right)^2\left(\text{E}(y^{2s})-\text{E}(y^s)^2\right) \\
&+\text{E}(u^2)\sum_x b^{\bullet 2}_{xs}\exp(s\tau^2)\left(\prod_{t=0}^{s-1}(1-q^{\bullet}_{xt})\right)\left(\text{E}(y^s)-\text{E}(y^{2s})\left(\prod_{t=0}^{s-1}(1-q^{\bullet}_{xt})\right)\right)
\end{aligned}$$

$$= \exp(\sigma^2)\left(\sum_x b^*_{xs}\left(\prod_{t=0}^{s-1}(1-q^*_{xt})\right)E(y^s)\right)^2\left(\exp(s\tau^2)-1\right)$$

$$+\exp(\sigma^2)\exp(s\tau^2)\left(\sum_x b^*_{xs}\left(\prod_{t=0}^{s-1}(1-q^*_{xt})\right)\right)^2\left(E(y^{2s})-E(y^s)^2\right)$$

$$+\exp(\sigma^2)\sum_x b^{*2}_{xs}\exp(s\tau^2)\left(\prod_{t=0}^{s-1}(1-q^*_{xt})\right)\left(E(y^s)-E(y^{2s})\left(\prod_{t=0}^{s-1}(1-q^*_{xt})\right)\right)$$

$$= \exp(\sigma^2)\left(\sum_x b^*_{xs}\left(\prod_{t=0}^{s-1}(1-q^*_{xt})\right)\right)^2\left(\exp(s\tau^2)E(y^{2s})-E(y^s)^2\right)$$

$$+\exp(\sigma^2)\sum_x b^{*2}_{xs}\exp(s\tau^2)\left(\prod_{t=0}^{s-1}(1-q^*_{xt})\right)\left(E(y^s)-E(y^{2s})\left(\prod_{t=0}^{s-1}(1-q^*_{xt})\right)\right)$$

Finally putting the two terms together we obtain:

$$\mathrm{Var}(T_s)=\left(\sum_x b^*_{xs}\left(\prod_{t=0}^{s-1}(1-q^*_{xt})\right)E(y^s)\right)^2\left(\exp(\sigma^2)-1\right)$$

$$+\exp(\sigma^2)\left(\sum_x b^*_{xs}\left(\prod_{t=0}^{s-1}(1-q^*_{xt})\right)\right)^2\left(\exp(s\tau^2)E(y^{2s})-E(y^s)^2\right)$$

$$+\exp(\sigma^2)\sum_x b^{*2}_{xs}\exp(s\tau^2)\left(\prod_{t=0}^{s-1}(1-q^*_{xt})\right)\left(E(y^s)-E(y^{2s})\left(\prod_{t=0}^{s-1}(1-q^*_{xt})\right)\right)$$

$$=\left(\sum_x b^*_{xs}\left(\prod_{t=0}^{s-1}(1-q^*_{xt})\right)\right)^2\left(\exp(\sigma^2)\exp(s\tau^2)E(y^{2s})-E(y^s)^2\right)$$

$$+\exp(\sigma^2)\sum_x b^{*2}_{xs}\exp(s\tau^2)\left(\prod_{t=0}^{s-1}(1-q^*_{xt})\right)\left(E(y^s)-E(y^{2s})\left(\prod_{t=0}^{s-1}(1-q^*_{xt})\right)\right)$$

APPENDIX E

We will calculate the mean and variance of T in stages. We first consider IBNR claims. As with Heckman & Meyers, Algorithm 3.3 gives:

$$E(N) = E_\chi\left(E(N|\chi)\right) = E_\chi(\lambda\chi) = \lambda E_\chi(\chi) = \lambda$$

and we also have:

$$\begin{aligned} \text{Var}(N) &= E_\chi\left(\text{Var}(N|\chi)\right) + \text{Var}_\chi\left(E(N|\chi)\right) \\ &= E_\chi(\lambda\chi) + \text{Var}_\chi(\lambda\chi) \\ &= \lambda E_\chi(\chi) + \lambda^2 \text{Var}_\chi(\chi) \\ &= \lambda + c\lambda^2 \end{aligned}$$

To ease the notation in what follows we will assume that the claims $X_{N_R+1}, X_{N_R+2}, \ldots, X_{N_R+N_B+N}$ are independently selected from a distribution with mean

$$E(X) = E(T_R|\text{Certainty})/N_R$$

and variance

$$\text{Var}(X) = \text{Var}(T_R|\text{Certainty})/N_R$$

This last relationship follows since

$$\text{Var}(T_R|\text{Certainty}) = \text{Var}\left(\sum_{i=1}^{N_R} X_i|\text{Certainty}\right) = \sum_{i=1}^{N_R} \text{Var}(X_i|\text{Certainty})$$

Now fixing β, ζ, and γ we have:

$$\begin{aligned} E(T|\beta,\gamma,\zeta) &= E_N\left(E\left(\beta\left(X_1 + X_2 + \ldots + X_{N_R} + \zeta\left(X_{N_R+1} + \ldots + X_{N_R+N_B}\right) + \gamma\left(X_{N_R+N_B+1} + \ldots + X_{N_R+N_B+N}\right)\right)\right)\right) \\ &= E_N\left(\beta\left(\sum_{i=1}^{N_R} E(X_i) + \zeta\sum_{i=1}^{N_B} E\left(X_{N_R+i}\right) + \gamma\sum_{i=1}^{N} E\left(X_{N_R+N_B+i}\right)\right)\right) \\ &= E_N\left(\beta\left(N_R E(X) + \zeta N_B E(X) + \gamma N E(X)\right)\right) \\ &= \beta(N_R + \zeta N_B + \gamma\lambda)E(X) \end{aligned}$$

For the variance in this case we have:

$$\begin{aligned}
\mathrm{Var}(T|\beta,\gamma,\zeta) &= \mathrm{E}_N\big(\mathrm{Var}(T|\beta,\gamma,\zeta,N)\big)+\mathrm{Var}_N\big(\mathrm{E}(T|\beta,\gamma,\zeta,N)\big)\\
&= \mathrm{E}_N\left(\mathrm{Var}\left(\beta\left(\sum_{i=1}^{N_R}X_i+\zeta\sum_{i=1}^{N_B}X_{N_R+i}+\gamma\sum_{i=1}^{N}X_{N_R+N_B+i}\right)\right)\right)\\
&\quad+\mathrm{Var}_N\left(\mathrm{E}\left(\beta\left(\sum_{i=1}^{N_R}X_i+\zeta\sum_{i=1}^{N_B}X_{N_R+i}+\gamma\sum_{i=1}^{N}X_{N_R+N_B+i}\right)\right)\right)\\
&= \mathrm{E}_N\left(\beta^2\left(\mathrm{Var}\left(\sum_{i=1}^{N_R}X_i\right)+\mathrm{Var}\left(\zeta\sum_{i=1}^{N_B}X_{N_R+i}\right)+\mathrm{Var}\left(\gamma\sum_{i=1}^{N}X_{N_R+N_B+i}\right)\right)\right)\\
&\quad+\mathrm{Var}_N\left(\beta\left(\mathrm{E}\left(\sum_{i=1}^{N_R}X_i\right)+\mathrm{E}\left(\zeta\sum_{i=1}^{N_B}X_{N_R+i}\right)+\mathrm{E}\left(\gamma\sum_{i=1}^{N}X_{N_R+N_B+i}\right)\right)\right)\\
&= \mathrm{E}_N\big(\beta^2\big(\mathrm{Var}(T_R|\mathrm{Certainty})+N_B\zeta^2\,\mathrm{Var}(X)+N\gamma^2\,\mathrm{Var}(X)\big)\big)\\
&\quad+\mathrm{Var}_N\big(\beta\big(N_R\,\mathrm{E}(X)+N_B\zeta\,\mathrm{E}(X)+N\gamma\,\mathrm{E}(X)\big)\big)\\
&= \beta^2\big(N_R\,\mathrm{Var}(X)+N_B\zeta^2\,\mathrm{Var}(X)+\mathrm{E}_N(N)\gamma^2\,\mathrm{Var}(X)\big)+\beta^2\gamma^2\,\mathrm{E}(X)^2\,\mathrm{Var}_N(N)\\
&= \beta^2\big(N_R+N_B\zeta^2+\lambda\gamma^2\big)\mathrm{Var}(X)+\beta^2\gamma^2\,\mathrm{E}(X)^2\big(\lambda+c\lambda^2\big)
\end{aligned}$$

Similarly we have, for a fixed values of β and ζ we have:

$$\begin{aligned}
\mathrm{E}(T|\beta,\zeta) &= \mathrm{E}_\gamma\big(\mathrm{E}(T|\beta,\gamma,\zeta)\big)\\
&= \mathrm{E}_\gamma\big(\beta(N_R+\zeta N_B+\gamma\lambda)\mathrm{E}(X)\big)\\
&= \beta(N_R+\zeta N_B+a\lambda)\mathrm{E}(X)
\end{aligned}$$

For the variance in this case we have:

$$\begin{aligned}
\mathrm{Var}(T|\beta,\zeta) &= \mathrm{E}_\gamma\big(\mathrm{Var}(T|\beta,\gamma,\zeta)\big)+\mathrm{Var}_\gamma\big(\mathrm{E}(T|\beta,\gamma,\zeta)\big)\\
&= \mathrm{E}_\gamma\big(\beta^2\big(N_R+N_B\zeta^2+\lambda\gamma^2\big)\mathrm{Var}(X)+\beta^2\gamma^2\,\mathrm{E}(X)^2\big(\lambda+c\lambda^2\big)\big)+\mathrm{Var}_\gamma\big(\beta(N_R+\zeta N_B+\gamma\lambda)\mathrm{E}(X)\big)\\
&= \mathrm{E}_\gamma\big(\beta^2\big(N_R+N_B\zeta^2\big)\mathrm{Var}(X)+\gamma^2\beta^2\big(\mathrm{E}(X)^2\big(\lambda+c\lambda^2\big)+\lambda\,\mathrm{Var}(X)\big)\big)+\mathrm{Var}_\gamma\big(\beta(N_R+\zeta N_B+\gamma\lambda)\mathrm{E}(X)\big)\\
&= \beta^2\big(N_R+N_B\zeta^2\big)\mathrm{Var}(X)+\mathrm{E}_\gamma\big(\gamma^2\big)\beta^2\big(\mathrm{E}(X)^2\big(\lambda+c\lambda^2\big)+\lambda\,\mathrm{Var}(X)\big)+\beta^2\lambda^2\,\mathrm{E}(X)^2\,\mathrm{Var}_\gamma(\gamma)\\
&= \beta^2\big(N_R+N_B\zeta^2\big)\mathrm{Var}(X)+\big(d+a^2\big)\beta^2\big(\mathrm{E}(X)^2\big(\lambda+c\lambda^2\big)+\lambda\,\mathrm{Var}(X)\big)+\beta^2\lambda^2\,\mathrm{E}(X)^2d\\
&= \beta^2\Big(\big(N_R+N_B\zeta^2+\big(d+a^2\big)\lambda\big)\mathrm{Var}(X)+\big(\big(d+a^2\big)\big(\lambda+c\lambda^2\big)+\lambda^2d\big)\mathrm{E}(X)^2\Big)
\end{aligned}$$

Now for a fixed β we have:

$$
\begin{aligned}
\mathrm{E}(T|\beta) &= \mathrm{E}_{\zeta}\left(\mathrm{E}(T|\beta,\zeta)\right) \\
&= \mathrm{E}_{\zeta}\left(\beta(N_R + \zeta N_B + \gamma\lambda)\mathrm{E}(X)\right) \\
&= \beta(N_R + rN_B + a\lambda)\mathrm{E}(X)
\end{aligned}
$$

The variance calculation also follows:

$$
\begin{aligned}
\mathrm{Var}(T|\beta) &= \mathrm{E}_{\zeta}\left(\mathrm{Var}(T|\beta,\zeta)\right) + \mathrm{Var}_{\zeta}\left(\mathrm{E}(T|\beta,\zeta)\right) \\
&= \mathrm{E}_{\zeta}\left(\beta^2\left(\left(N_R + N_B\zeta^2 + (d+a^2)\lambda\right)\mathrm{Var}(X) + \left((d+a^2)(\lambda + c\lambda^2) + \lambda^2 d\right)\mathrm{E}(X)^2\right)\right) + \mathrm{Var}_{\zeta}\left(\beta(N_R + \zeta N_B + a\lambda)\mathrm{E}(X\right. \\
&= \beta^2\left(\left(N_R + N_B\mathrm{E}_{\zeta}(\zeta^2) + (d+a^2)\lambda\right)\mathrm{Var}(X) + \left((d+a^2)(\lambda + c\lambda^2) + \lambda^2 d\right)\mathrm{E}(X)^2\right) + \mathrm{E}(X)^2 N_B^2\beta^2\,\mathrm{Var}_{\zeta}(\zeta) \\
&= \beta^2\left(\left(N_R + N_B(z + r^2) + (d+a^2)\lambda\right)\mathrm{Var}(X) + \left((d+a^2)(\lambda + c\lambda^2) + \lambda^2 d\right)\mathrm{E}(X)^2\right) + z\beta^2 N_B^2\,\mathrm{E}(X)^2 \\
&= \beta^2\left(\left(N_R + N_B(z + r^2) + (d+a^2)\lambda\right)\mathrm{Var}(X) + \left((d+a^2)(\lambda + c\lambda^2) + \lambda^2 d + zN_B^2\right)\mathrm{E}(X)^2\right)
\end{aligned}
$$

Thus, combining these results, we have:

$$
\begin{aligned}
\mathrm{E}(T) &= \mathrm{E}_{\beta}\left(\mathrm{E}(T|\beta)\right) \\
&= \mathrm{E}_{\beta}\left(\beta(N_R + rN_B + a\lambda)\mathrm{E}(X)\right) \\
&= \mathrm{E}_{\beta}(\beta)(N_R + rN_B + a\lambda)\mathrm{E}(X) \\
&= (N_R + rN_B + a\lambda)\mathrm{E}(X)
\end{aligned}
$$

Finally we have:

$$
\begin{aligned}
\mathrm{Var}(T) &= \mathrm{E}_\beta\left(\mathrm{Var}(T|\beta)\right) + \mathrm{Var}_\beta\left(\mathrm{E}(T|\beta)\right) \\
&= \mathrm{E}_\beta\left(\beta^2\left(\left(N_R + N_B\left(z + r^2\right) + \left(d + a^2\right)\lambda\right)\mathrm{Var}(X) + \left(\left(d + a^2\right)\left(\lambda + c\lambda^2\right) + \lambda^2 d + zN_B^2\right)\mathrm{E}(X)^2\right)\right) \\
&\quad + \mathrm{Var}_\beta\left(\beta(N_R + rN_B + a\lambda)\mathrm{E}(X)\right) \\
&= \mathrm{E}_\beta\left(\beta^2\right)\left(\left(N_R + N_B\left(z + r^2\right) + \left(d + a^2\right)\lambda\right)\mathrm{Var}(X) + \left(\left(d + a^2\right)\left(\lambda + c\lambda^2\right) + \lambda^2 d + zN_B^2\right)\mathrm{E}(X)^2\right) \\
&\quad + \mathrm{Var}_\beta(\beta)(N_R + rN_B + a\lambda)^2\,\mathrm{E}(X)^2 \\
&= (b+1)\left(\left(N_R + N_B\left(z + r^2\right) + \left(d + a^2\right)\lambda\right)\mathrm{Var}(X) + \left(\left(d + a^2\right)\left(\lambda + c\lambda^2\right) + \lambda^2 d + zN_B^2\right)\mathrm{E}(X)^2\right) \\
&\quad + b(N_R + rN_B + a\lambda)^2\,\mathrm{E}(X)^2 \\
&= (b+1)\left(N_R + N_B\left(z + r^2\right) + \left(d + a^2\right)\lambda\right)\mathrm{Var}(X) \\
&\quad + \left((b+1)\left(\left(d + a^2\right)\left(\lambda + c\lambda^2\right) + \lambda^2 d + zN_B^2\right) + b\left(N_R + rN_B + a\lambda\right)^2\right)\mathrm{E}(X)^2
\end{aligned}
$$

Thus, in terms of estimates for case reserved claims without parameter uncertainty:

$$
\begin{aligned}
\mathrm{Var}(T) &= (b+1)\left(N_R + N_B\left(z + r^2\right) + \left(d + a^2\right)\lambda\right)\frac{\mathrm{Var}\left(T_R|\text{No uncertainty}\right)}{N_R} \\
&\quad + \left((b+1)\left(\left(d + a^2\right)\left(\lambda + c\lambda^2\right) + \lambda^2 d + zN_B^2\right) + b\left(N_R + rN_B + a\lambda\right)^2\right)\left(\frac{\mathrm{E}\left(T_R|\text{No uncertainty}\right)}{N_R}\right)^2 \\
&= (b+1)\left(1 + \frac{N_B\left(z + r^2\right) + \left(d + a^2\right)\lambda}{N_R}\right)\mathrm{Var}\left(T_R|\text{No uncertainty}\right) \\
&\quad + \left(\frac{(b+1)\left(\left(d + a^2\right)\left(\lambda + c\lambda^2\right) + \lambda^2 d + zN_B^2\right)}{N_R{}^2} + b\left(1 + \frac{rN_B + a\lambda}{N_R}\right)^2\right)\mathrm{E}\left(T_R|\text{No uncertainty}\right)^2
\end{aligned}
$$

Modeling Parameter Uncertainty in Cash Flow Projections

Roger M. Hayne, FCAS, MAAA

MODELING PARAMETER UNCERTAINTY IN CASH FLOW PROJECTIONS

by

Roger M. Hayne

Abstract

In order to be complete dynamic financial analysis (DFA) models should deal with both the amount and timing of future loss and loss adjustment expense payments. Even more than asset cash flows, these future payments are very uncertain. However, even with this uncertainty, one would expect to see payments that are somewhat stable from year to year.

This paper presents an approach that can deal with this seeming contradiction. By separating total uncertainty in future cash flows into its parameter and process components we present a method to model future liability cash flows that maintains the desired total uncertainty characteristics. However, it will also result in specific payment flow "paths" having less variation from year to year than would a completely random sample from the expected total payout would indicate.

There is also a companion of this paper, titled "Estimating Uncertainty in Cash Flow Projections" that considers the problem of estimating the distributions, including separate consideration of process and parameter uncertainty.

Biography

Roger is a Fellow of the Casualty Actuarial Society, a Member of the American Academy of Actuaries, and Consulting Actuary in the Pasadena, California office of Milliman & Robertson, Inc. with over twenty-one years of casualty actuarial consulting experience. Roger is a frequent speaker on reserve and DFA related topics and has authored several papers dealing with considerations and estimates of uncertainty in reserve projections. Roger is currently the chair of the CAS Research Policy and Management Committee and has served as chair of both the CAS Committee on Theory of Risk and the CAS/AAA Joint Committee on the Casualty Loss Reserve Seminar.

MODELING PARAMETER UNCERTAINTY IN CASH FLOW PROJECTIONS

Introduction

With the increased focus on dynamic financial analysis (DFA) as a tool to assist in quantifying the financial strength of insurers and other risk bearing entities, comes increased demands on tools for use in those models. As with reserves, insurer cash outflows representing those liabilities are subject to considerable uncertainty. Capturing and appropriately modeling this uncertainty will greatly enhance the accuracy and reliability of DFA models.

The purpose of this paper is to outline a simple approach that can be used to capture various sources of uncertainty and incorporate them into stochastic cash flow models. A simple example should help illustrate this point.

Consider two insurers, both with expected reserves of $90 million, assets of $110 million, ignoring interest, and experiencing the following future payment possibilities:

Table 1: Distribution for Stable Insurer, Inc.

	Year		
Probability	1	2	Total
50.0%	$80	$40	$120
50.0%	40	20	60
Expected	$60	$30	$90

Table 2: Distribution for Random Insurer, Inc.

	Year		
Probability	1	2	Total
25.0%	$80	$40	$120
25.0%	80	20	100
25.0%	40	40	80
25.0%	40	20	60
Expected	$60	$30	$90

Each insurer experiences the same distribution of possible payments in each year. However, the first insurer has a 50% chance of becoming insolvent at the end of two years while the second has only a 25% chance.

The primary difference is that Random Insurer is allowed to experience all possible "futures" with either $80 or $40 paid in the first year and either $40 or $20 paid in the second. Stable Insurer is only allowed two possible "futures," the best and the worst. As we will see, these are simple examples of two approaches to modeling liability cash flows.

If historically the second year's payments were always half of those in the first year, then it could be argued that Stable Insurer's pattern is closer to "reality" than that of Random Insurer. The challenge, then, is to develop methods of modeling liability cash flows that capture the full variation that can be expected in future payments, without "unrealistic" swings in payments from year to year. That is the purpose of this paper.

Types of Uncertainty

There are many ways to categorize uncertainty. Here we will divide uncertainty faced by actuaries into three categories:

1. *Process* – uncertainty present simply from the random nature of a particular process, even if the process itself is known with certainty,
2. *Parameter* – uncertainty that parameters selected for a particular model accurately reflect the reality to be modeled, and
3. *Specification and/or Model* – uncertainty that the models selected themselves accurately reflect the reality to be modeled.

Sometimes the third category is divided into two parts, model and specification where specification refers to the selection of distributions and model refers to the selection of the underlying model itself.

For example, if we throw a fair die, even though we know the underlying physical model with (relative) certainty, there is still an equal chance of each of the six sides showing up. This is an example of process uncertainty.

If, however, the die may be "loaded," but that we know we are observing the throw of a die, we have added parameter uncertainty to the situation. Here we know we will observe throws from one through six, but with one result potentially having higher probability than the others do.

Finally, we could be observing a series of digits from 1 through 6 without knowing the underlying process generating the series. We can still use a loaded die model. However, there is the possibility that some other process is generating the digits that cannot be modeled using a loaded die. For example, the digits could be the last digit from a Geiger counter reading with 1 substituted for 7 and 8 and 6 substituted for 9 and 0. Here we have specification or model uncertainty.

Modeling Process Uncertainty

These categories of uncertainty are increasingly difficult to estimate. Reserves for insurers, or other risk bearing entities, are often set using non-statistical actuarial forecasting methods, including broad application of "actuarial judgment."

Even when statistical methods are used, the information regarding the resulting uncertainty is usually limited to conclusions within the framework of the model. For example, two different statistical models may result in two different probability ranges about their estimates with possibly little or no overlap in the ranges.[1] The same statistical model applied to two different sets of data, paid and incurred losses for example, could even give widely different results and ranges.

Statistical projection methods also tend to concentrate on "squaring the triangle" for a single set of data, usually paid losses. As Berquist and Sherman[2] and many other papers dealing with reserve estimation indicate, there is valuable information in many different insurer statistics. Claim count statistics are extremely valuable in a reserve analysis. Frequency and severity methods are often less volatile than development factor (or link ratio) methods for less mature exposure periods. In addition, claim counts, in conjunction with other insurer data, can help identify changes that could affect one or another projection method. For example, changes in average case reserves per open claim could signal a change in relative reserve adequacy thus affecting projections

---

[1] See, for example, *Transcripts of the 1992 Casualty Loss Reserve Seminar*, pp. 1123-1150. This Advanced Case Study presented two actuaries with the same set of data and asked them to develop reserve and variability estimates. One estimated reserves to be $239 million with a $12.7 million standard error. The other estimated reserves to be $178 million with a standard deviation of $10.7 million.

[2] Berquist, J.R. and Sherman, R.E., "Loss Reserve Adequacy Testing: A Comprehensive, Systematic Approach," *Proceedings of the Casualty Actuarial Society*, LXIV, 1977, pp. 123-184.

based on incurred loss development. Similarly, changes in the rate at which claims are closed will affect methods based on paid losses. The author is unaware of any statistical method that incorporates all these items of information in estimating ultimate losses.

The collective risk model offers a rather easily understood framework to model insured uncertainty. Briefly the collective risk model is based on the following algorithm:

Algorithm 1 – Collective Risk:

1. Randomly select N, the number of claims that will occur.
2. Randomly select N independent claims, $X_1, X_2, \ldots, X_N$ from the selected claim size distribution.
3. Total the amounts $T = \sum_{i=1}^{N} X_i$.
4. Repeat steps 1 through 3 "many" times.

With a minimum of additional assumptions we can derive some very useful relationships between the distributions of the number (N) and size (X) of individual claims and that of the total. In particular, if sufficient moments exist for the various distributions and if all random variables are independent then we have:

(1)
$$\begin{aligned} \mathrm{E}(T) &= \mathrm{E}(N)\mathrm{E}(X) \\ \mathrm{Var}(T) &= \mathrm{E}(N)\mathrm{Var}(X) + \mathrm{Var}(N)\mathrm{E}^2(X) \end{aligned}$$

Similar formulae also hold for higher moments.[3]

The collective risk model also seems to be a logical choice to model process uncertainty in the distribution of insured losses. There has been considerable attention paid to this basic model in the literature and several algorithms have been developed to calculate the distribution of T given distributions of N and X. Probably of greatest interest to

---

[3] See, for example, Mayerson, A.L., Jones, D.A., Bowers, N.L. (Jr.), "The Credibility of the Pure Premium." *Proceedings of the Casualty Actuarial Society*, LV, 1968, p. 179 for these and formulae for third and fourth moments.

practicing casualty actuaries are references by Heckman and Meyers,[4] Panjer and Willmot,[5] Robertson,[6] and the text about to appear by Klugman, Panjer, and Willmot.[7]

The attractiveness of the collective risk model, aside from its description of the insurance process is that it breaks the problem of estimating process variation into more manageable parts, i.e., to estimating the distribution of claim counts and the distribution of the size of claims. As with any model, the collective risk model is an approximation of reality. Many actuaries are concerned with some of its inherent assumptions, not the least of which is the assumption of independence among claims and between the claim size and the claim count distributions. Recent work by Wang[8], sponsored by the Casualty Actuarial Society, addresses this issue. Although derived independently, the methods here follow closely with those presented by Wang.

Some Approaches to Parameter Uncertainty

Probably the most intuitive approach to modeling parameter uncertainty would be Bayesian. Generally one would assume the distribution we wished to model, that of aggregate losses, had a particular distribution with one or more of its parameters being uncertain, itself having a separate distribution. There are many distribution pairs of conditional and prior distributions that mix to closed form mixed distributions. In the appendix to his chapter in *Foundations of Casualty Actuarial Science*, Venter[9] for example has assembled of useful distribution pairs.

---

[4] Heckman, P.E., Meyers, G.G., "The Calculation of Aggregate Loss Distributions From Claim Severity and Claim Count Distributions," *Proceedings of the Casualty Actuarial Society*, LXX, 1983, pp. 22-61

[5] Panjer, G., Willmot, G. *Insurance Risk Models*, Society of Actuaries, Chicago, 1992

[6] Robertson, J.P., "The Computation of Aggregate Loss Distributions," *Proceedings of the Casualty Actuarial Society*, LXXIX, 1992, pp. 57-133

[7] Klugman, S.A., Panjer, H.H., Willmot, G.E., *Loss Models: From Data to Decisions*, John Wiley & Sons, New York, 1998.

[8] Wang, S.S., "Aggregation of Correlated Risk Portfolios: Models & Algorithms," Casualty Actuarial Society at www.casact.org (part of the Committee on Theory of Risk page of the Research portion of the web site).

[9] Venter, G.G., "Credibility," *Foundations of Casualty Actuarial Science*, Casualty Actuarial Society, 1992, Chapter 7, pp. 375-483.

One example here may be helpful. Suppose X has a lognormal distribution with parameters μ and σ^2. By this we mean that X has the probability density function:

(2)
$$f(x) = \frac{\exp\left(-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right)}{x\sigma\sqrt{2\pi}}$$

It is well known that the random variable X is lognormal if and only if the random variable $\ln X$ is normal. In this parameterization the variable $\ln X$ has a normal distribution with mean μ and variance σ^2. If, now, we assume μ is uncertain but has a normal distribution with mean m and variance τ^2, then the random variable X is still lognormal with parameters m and $\sigma^2+\tau^2$. We note that the inclusion of parameter uncertainty in this way has the effect of increasing both the mean and variance of the distribution. This follows from the following results for a lognormal distribution with parameters μ and σ^2.

(3)
$$\begin{aligned} E(X) &= \exp\left(\mu + \tfrac{1}{2}\sigma^2\right) \\ Var(X) &= \exp\left(2\mu + \sigma^2\right)\left(\exp\left(\sigma^2\right) - 1\right) = E^2(X)\left(\exp\left(\sigma^2\right) - 1\right) \end{aligned}$$

As an aside, the reader should note that Venter's parameterization of the lognormal distribution differs from what we use here. The first parameter in our parameterization is the mean of the normal distribution of $\ln X$ whereas Venter's parameter is the exponential of this amount. Thus in the appendix Venter assumes the prior distribution of the parameter is lognormal to conclude the mixed distribution is lognormal. Because of the log transformation between the two paramterizations, and the fact that a variable X is lognormal if and only if the variable $\ln X$ is normal, the two results are actually identical.

Thus one intuitive way to model parameter uncertainty would be to select a pair of distributions (lognormal and normal in this example), use the lognormal to model process uncertainty (as an approximation to the results of a collective risk model). Parameter uncertainty could then be built in by allowing the μ parameter to have a distribution of its own. In this paper we will label method this the Bayesian approach.

Another approach to modeling parameter uncertainty is discussed in Heckman and Meyers.[10] In their approach they separate parameter and process uncertainty by use of additional random variables. The following is a slight modification of the algorithm they present:

Algorithm 2 – Refined Collective Risk:

1. Randomly select N, the number of claims that will occur from a distribution with mean λ and variance $\lambda+c\lambda^2$.

2. Randomly select N independent claims, $X_1, X_2, \ldots, X_N$ from the selected claim size distribution.

3. Randomly select a mixing parameter β from a distribution with mean 1 and variance b.

4. Total the amounts and divide by β, $T = \left(\sum_{i=1}^{N} X_i\right)\Big/\beta$.

5. Repeat steps 1 through 4 "many" times.

Actually, in Heckman and Meyers the authors assume the claim count distribution is a mix of a Poisson prior distribution with a gamma uncertainty distribution for a negative binomial posterior distribution. Their results, however, generalize to situations where the parameter c is negative, which does not make sense in terms of mixed distributions. The algorithm they present for calculating the aggregate distribution does require either a Poisson, binomial, or negative binomial claim count distribution, but the results we use here do not need that assumption.

The primary result we will use, however, is that given *Algorithm 2*, and assuming all the distributions are independent from each other, then we have the following relationships:

$$\text{(4)} \qquad \begin{aligned} \mathrm{E}(T) &= \lambda \mathrm{E}(X) \\ \mathrm{Var}(T) &= \lambda(1+b)\mathrm{E}\left(X^2\right) + \lambda^2(b+c+bc)\mathrm{E}^2(X) \end{aligned}$$

---

[10] Heckman, P.E., Meyers, G.G., ibid

We note that these formulae reduce to formulae (1) in the case that b=0. Rearranging terms in the variance formula we obtain:

(5)

$$
\begin{aligned}
\mathrm{Var}(T) &= \lambda \mathrm{E}(X^2) + b\lambda \mathrm{E}(X^2) + \lambda^2 c \mathrm{E}^2(X) + \lambda^2 b \mathrm{E}^2(X) + \lambda^2 bc \mathrm{E}^2(X) \\
&= \lambda \mathrm{E}(X^2) + \lambda^2 c \mathrm{E}^2(X) + b\left(\lambda \mathrm{E}(X^2) + \lambda^2 c \mathrm{E}^2(X) + \lambda^2 \mathrm{E}^2(X)\right) \\
&= \mathrm{Var}(T|b=0) + b\left(\mathrm{Var}(T|b=0) + \lambda^2 \mathrm{E}^2(X)\right)
\end{aligned}
$$

Which can be used to obtain the following useful relationship for the coefficient of variation (ratio of the standard deviation to the mean) of the respective distributions:

(6)

$$
\begin{aligned}
\mathrm{cv}^2(T) &= \frac{\mathrm{Var}(T)}{\mathrm{E}^2(T)} \\
&= \frac{\mathrm{Var}(T|b=0) + b\left(\mathrm{Var}(T|b=0) + \lambda^2 \mathrm{E}^2(X)\right)}{\mathrm{E}^2(T|b=0)} \\
&= \frac{\mathrm{Var}(T|b=0)}{\mathrm{E}^2(T|b=0)} + b\left(\frac{\mathrm{Var}(T|b=0)}{\mathrm{E}^2(T|b=0)} + \frac{\lambda^2 \mathrm{E}^2(X)}{(\lambda \mathrm{E}(X))^2}\right) \\
&= \mathrm{cv}^2(T|b=0) + b\left(\mathrm{cv}^2(T|b=0) + 1\right)
\end{aligned}
$$

Solving for b we obtain:

(7)

$$
b = \frac{\mathrm{cv}^2(T) - \mathrm{cv}^2(T|b=0)}{\mathrm{cv}^2(T|b=0) + 1}
$$

Recalling that b=0 refers to the situation with only process variation, this formula provides a way to model parameter uncertainty given knowledge of the coefficient of variation for the final distribution and that for the distribution with only process uncertainty.

From this point on we will assume that we know the various means and variances of the distributions with and without parameter uncertainty and concentrate on practical considerations in modeling these sources of uncertainty.

Moving to the example with a lognormal prior distribution mixed with a normal distribution let us consider two different ways of modeling the amounts. We will identify two methods to generate random loss amounts.

Intuitive Method:

1. Randomly pick ν from a normal distribution with mean m and variance τ^2.

2. Randomly pick X from a lognormal distribution with mean ν and variance σ^2.

"Smarter" Method:

1. Randomly pick X from a lognormal distribution with mean m and variance $\sigma^2+\tau^2$.

As we saw above, both methods give exactly the same result. The *Intuitive Method* is simply the Bayesian statement of the problem and the *Smarter Method* is the posterior distribution.

A Dilemma?

Consider a very simple extension of our Bayesian type of algorithm with a lognormal mixed with a normal but for multiple years.

Algorithm 3. Multiple Year Bayesian

1. Assume X_i has a lognormal distribution with parameters μ_i and σ_i^2, with σ_i^2 known but

2. $\mu_i = m_i\beta$ where β has a normal distribution with mean b and variance τ^2, with both b and τ^2 known.

Here the parameter β provides "global" parameter uncertainty. The above discussion leads us to conclude that each X_i has a lognormal distribution with parameters bm_i and $\sigma_i^2 + m_i^2\tau_i^2$. Thus we are tempted to use either the *Intuitive Method* or the *Smarter Method* in modeling. In this case we would have the methods described as:

Intuitive Method:

1. Randomly pick β from a normal distribution with mean b and variance τ^2.

2. Randomly pick X_i from a lognormal distribution with parameters $\mu_i = m_i\beta$ and σ_i^2.

"Smarter" Method:

1. Randomly pick X from a lognormal distribution with parameters bm_i and $\sigma_i^2 + m_i^2\tau_i^2$.

Our reasoning above could lead to the conclusion that the two methods give the same answer. In fact the distributions for each year are identical. However, consider the example where m_i=0.25i, all the σ_i^2=0, and $b=\tau^2$=1. The following graphs make it clear that, at least in this case, the two methods give considerably different answers:

Figure 1: Intuitive Method, First Example

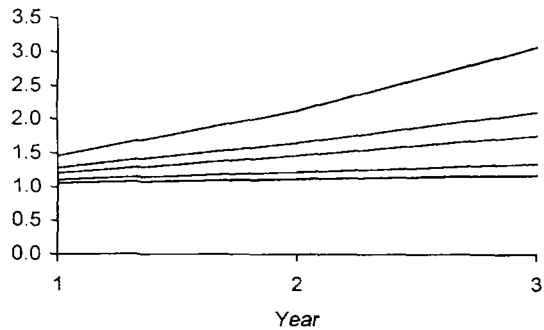
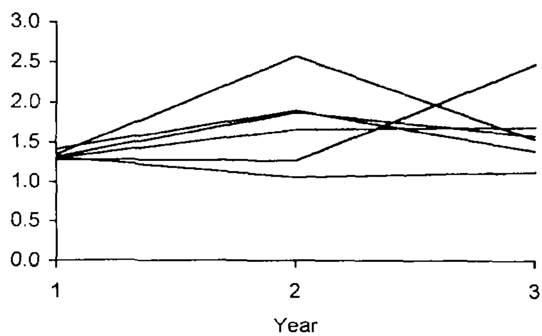


Figure 2: "Smarter" Method, First Example



Even though each year has a lognormal distribution by itself, the structure does not imply that each year is independent of the others. That is the major difference between the *Intuitive* and "*Smarter*" methods. It is also the difference between Stable Insurer and Random Insurer in the *Introduction*.

The above statement of the multiple year algorithm may lead to some ambiguity regarding the role of the uncertainty parameter. The following restatement may help clarify the ambiguity and provide us with a more explicit means to move *Algorithm 2* to a multiple year setting.

Algorithm 4, Refined Multiple Year Bayesian Algorithm

1. Select p with $0 < p < 1$.

2. Set $\mu_i = m_i + \tau_i \Phi^{-1}(p)$, where $\Phi^{-1}(p)$ represents the inverse normal distribution, that is the value such that $\mathrm{P}\left(Z < \Phi^{-1}(p) \middle| Z \sim \mathrm{N}(0,1)\right) = p$.

3. Randomly select X_i from a lognormal distribution with parameters μ_i and σ_i^2, σ_i^2 are known.

4. Repeat steps 2 and 3 for each year to be modeled.

5. Repeat steps 1 through 4 "many" times

We recognize a slight inconsistency in the parametizations of these two versions. Strictly speaking we should have $\mu_i = m_i\left(b + \tau_i \Phi^{-1}(p)\right)$ to be consistent with the first, but this parameterization leads directly to the conclusions for each year individually exactly parallel to those of the single year case.

Implications in Modeling Liabilities

Liabilities for most lines of insurance are characterized by fairly (a *very* relative term) stable payments from year to year. Obvious exceptions are lines subject to catastrophe losses and small liability books with large loss exposure. Even large claims may have extended settlement provisions, affecting the timing and variation of future payments.

If we consider only process variation we see that the law of large numbers soon comes into play. From (1) in the case of the collective risk model we have:

(8)

$$\begin{aligned} cv^2(T) &= \frac{Var(T)}{E^2(T)} \\ &= \frac{E(N)Var(X) + Var(N)E^2(X)}{E^2(X)E^2(N)} \\ &= \frac{Var(X)}{E^2(X)E(N)} + \frac{Var(N)}{E^2(N)} \\ &= \frac{cv^2(X)}{E(N)} + cv^2(N) \end{aligned}$$

If we make the usual assumption now that N has a Poisson distribution with variance equal to the mean then this becomes:

(9)

$$cv^2(T) = \frac{cv^2(X) + 1}{E(N)}$$

Thus, no matter how volatile the claim size distribution is, the total amounts paid could have arbitrarily small relative variation simply by having $E(N)$ sufficiently large. We note the law of large numbers is a special case here where the variance of the number of claims is zero. The same result will follow for any claim count distribution whose standard deviation grows more slowly than the mean, more precisely, whenever

(10)

$$Var(N) = o(E(N)) \text{ as } E(N) \to \infty$$

The power of the law of large numbers should not be underestimated. Even if the claim count distribution were fairly "noisy" with a standard deviation of 5 times the mean, it would only require a Poisson distribution with 100 claims to result in the standard deviation of the total to 51% of the total. With 5,000 claims, not unusual for a fairly large insurer, the standard deviation reduces to 7% of the total. If one would use a rule of thumb that results beyond two standard deviations "rare" in this case it would be rare for actual payments to deviate by more than 15% of the mean.

We recognize that "fairly noisy" is a soft term. Many would argue, and quite persuasively, distributions that are interesting to actuaries may not have finite standard deviations, or maybe not even have finite means. However, with policy limits usually in effect, distributions losses faced by insurers usually have finite means and variances.

The conclusion we reach is the same reached by Meyers and Schenker.[11] For insurers, and larger self-insured entities, the law of large numbers gives process variation much less influence on the overall variation of results than other sources of uncertainty. Thus parameter uncertainty and model or specification uncertainty are more significant issues to insurers than simple process uncertainty.

Realistic modeling of liabilities in a dynamic financial model then must balance two realities. First payments for an insurer are often fairly consistent from year to year. Second the liabilities for insurers or self-insureds often have a high degree of uncertainty, often well beyond that which can be attributed to process variation alone.

One way to look at the problem is to consider payments as falling along various future "paths" with relatively little variation in payments from year-to-year on any given path but with potentially widely varying paths or futures. If this is actually the case, modeling future cash payments should be relatively straightforward. We could assume that variation in payments from year to year would be caused by process variation whereas other sources of uncertainty reflect various possible future paths.

Consider, for example, *Algorithm 4* with a multiple year runoff of reserves, as given by the following table, assuming no parameter uncertainty:

Table 3: Refined Example Data

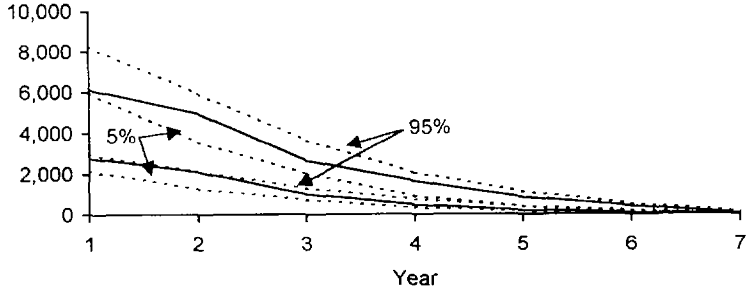
Year	E(X)	E(N)	E(T)	cv(T)
1	5,000	1,000	5,000,000	0.100
2	11,000	300	3,300,000	0.155
3	13,000	150	1,950,000	0.183
4	20,000	50	1,000,000	0.255
5	25,000	20	500,000	0.316
6	30,000	7	210,000	0.423
7	40,000	1	40,000	1.031

If, now for simplicity, we assume that the payments in each year have lognormal distributions, but with "global" parameter uncertainty as described in *Algorithm 4* with $\tau_i = 0.5$ we can then view alternative future reserve runoffs in the following chart:

---

[11] Meyers, G.G., Schenker, N., "Parameter Uncertainty in the Collective Risk Model," *Proceedings of the Casualty Actuarial Society*, LXX, 1983, pp.111-143.

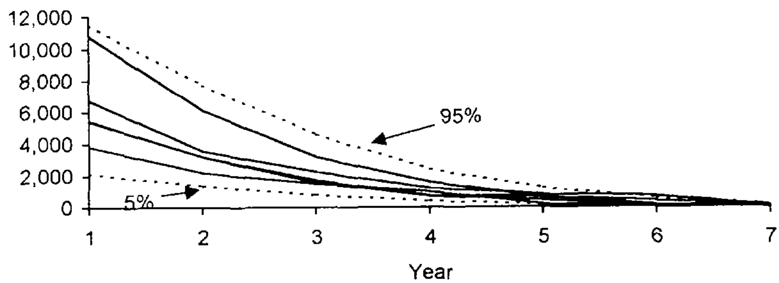
Figure 3: Two "Paths" with Probability Levels



Here the two sets of lines present two of the many possible "futures," corresponding to two different probability levels for the parameter uncertainty. The solid lines indicate the simulated reserve runoff, while the dotted lines represent the 5% and 95% probability bounds accounting only for process uncertainty as defined in the above table. Thus, for these two selected parameter uncertainty levels, we would expect 90% of the possible futures to lie between the dotted lines.

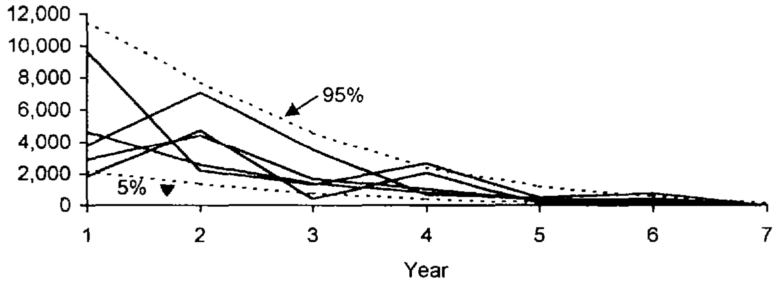
The following graph shows the global 90% range with several simulated runoffs (using our "Intuitive" approach).

Figure 4: Refined Example, Intuitive Method



To show the difference with the "*Smarter*" method the following is a graph showing the fully random lognormal approach:

Figure 5: Refined Example, "Smarter" Method



Again, the *Intuitive* approach gives smoother paths, yet still does provide the total uncertainty expected.

We can also generalize *Algorithm 2* to model multiple year uncertainty.

Algorithm 5 – Multiple Year Refined Collective Risk

1. Assume that payment amount process uncertainty can be modeled by known distributions in each year.

2. Assume that other sources of uncertainty in each year can be reflected by dividing by a "distortion" variable β_i, having mean 1 and known variance b_i.

3. Randomly select $0 < p < 1$

4. Select each β_i from the distortion distributions at probability level p.

5. Randomly select payments in year i, X_i from the assumed distributions.

6. Model amounts by the ratio of X_i and the selected β_i.

We note for each i this model is similar to *Algorithm 2*. The principal difference is the "linkage" between years provided by selecting the distortion variable at the same probability level for each year.

For each year, then, if we can estimate total variation, the variances required in the second step can actually be easily determined using formula (7) above. Of course, estimating total variation is not a trivial matter. There currently may be no agreed-upon method to derive such estimates, however this continues to be an active area of actuarial research.

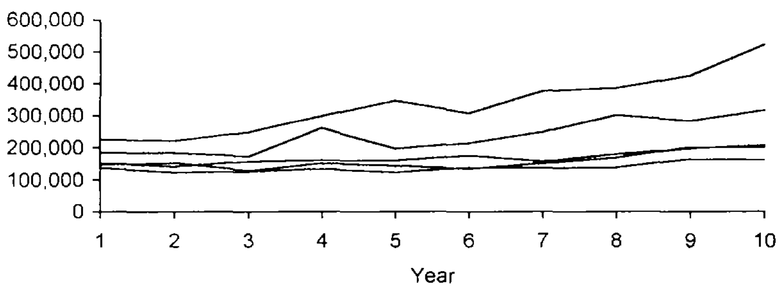
Assuming that we can get the total variance estimates, the following is an example of estimating the b_i values and the resulting graphs. These estimates are based on a fairly comprehensive attempt to estimate process uncertainty as well as other sources of uncertainty in the estimates. All estimates are in current dollars (with the effect of inflation removed) and are for total forecast payments in future years, including those arising from future exposures.

Table 4: Comprehensive Example

	Expected	Standard Deviation		Implied
Year	Paid	Process	Total	b Value
1	$213,000	$5,900	$60,700	0.0804
2	218,000	14,200	96,900	0.1925
3	237,000	22,800	125,000	0.2665
4	255,000	30,700	144,700	0.3031
5	274,000	36,100	167,800	0.3516
6	294,000	38,200	189,300	0.3911
7	316,000	42,900	209,100	0.4118
8	337,000	29,500	228,700	0.4494

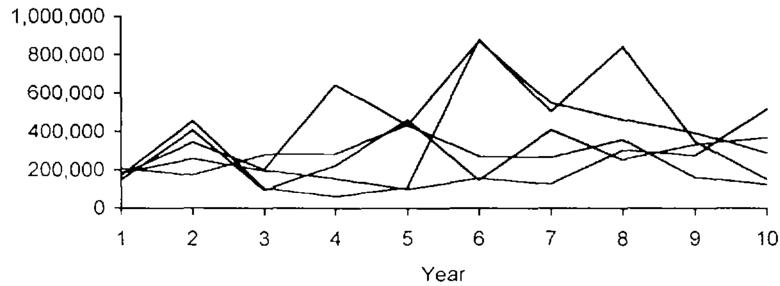
The following graph shows simulations based on *Algorithm 5* using the simplifying assumptions that the uncertainty parameters all have gamma distributions and that process uncertainty can be adequately modeled by a lognormal distribution.

Figure 6: Comprehensive Example, Intuitive Method



This shows relatively moderate variation from year to year but a fairly wide spread of possible outcomes. Both would be expected given the standard deviations shown above. As we compared in other situations, the following graph follows the "*Smarter*" method and results in substantially more variation from year to year than *Algorithm 5*.

Figure 7: Comprehensive Example, "Smarter" Method



As in prior examples of the "*Smarter*" method, there are substantial swings in payments from year to year. If we would expect some predictability of payments then using these simulations in a dynamic financial analysis model may be misleading. In short, the "*Smarter*" model is not really so smart in these situations.

Conclusion

Simply knowing the total distribution of payments in any particular future year does not necessarily give the actuary sufficient information to accurately and adequately model future payments, whether the application be in a full dynamic financial analysis model or in other applications where modeling of reserve payout is important. This paper presents one of many possible alternatives that can be used to separate process variations that will happen even if all information about the model is completely known, from other, potentially more global, influences. Still remaining, however, is significant research into the proper models to be used and in estimating the parameters of those models.

Taking Uncertainty Into Account: Bias Issues Arising from Parameter Uncertainty in Risk Models

John A. Major, ASA, MAAA

Taking Uncertainty Into Account: Bias Issues Arising from Parameter Uncertainty in Risk Models

by John A. Major, ASA

Given a random variable of interest, a historical sample of its realized values, and the desire to model its possible future values, actuarial training provides many methods for selecting a family of probability models (distributions) and determining specific parameter values that best represent it. But how should one take parameter uncertainty (parameter risk) into account? In particular, uncertainty can lead to bias in estimators commonly used by actuaries. This paper examines the problem of adjusting estimated distributions (risk curves) to remove the undesirable bias effects of parameter risk, and shows several solutions. It goes on, however, to critique the very notion of uncertainty-adjusted risk curves, emphasizing that this is an ambiguous concept. The form of the adjustment depends crucially on details of the specific question being addressed, so much so that an estimator can seem to be simultaneously overestimating and underestimating risk. Parameter uncertainty therefore cannot be "taken into account" in an unequivocal manner. It is recommended that parameter risk be held apart from process risk and presented in terms of confidence intervals; only with that as background – and with great care – should bias corrections be attempted.

0. INTRODUCTION

0.1 Parameter Estimation for DFA

For DFA in particular, the problem of parameter estimation occurs in the process of determining the appropriate method for generating random variables in the simulation of a financial security system. For example, if it is desired to investigate the relative efficacy of various reinsurance alternatives, a simulation can be created that tests the alternatives in a series of hypothetical "future histories" of loss experience. To simulate many realizations of possible future losses – many more than have been observed in the past – it is necessary to first create a model of the probability distribution of losses. Such a model would be based, at the very least, on the loss experience observed in the past.

If one can determine an appropriate cumulative probability distribution function (risk curve) $F_X(X,\theta)$ to associate with the random variable of interest X, then random instances of X can be created by the inverse lookup method: $X = F_X^{-1}(U,\theta)$ where U is a uniformly distributed random variable between 0 and 1. For specific distributions, more efficient techniques are available, but inverse lookup will always work when F can be inverted, either analytically or numerically.

Generated variables X, Y, Z, etc., can be combined in pro forma financial statements or other actuarial calculations to simulate financial results $R=\varphi(X,Y,Z,\ldots)$. After a sufficient number of simulation cycles, the empirical distribution of R values can be used to assess the risk to the financial system, answering questions such as "What value of R is not likely to be exceeded with probability q?" and "What is the probability that R will be greater than (a fixed value) L?"

0.2 Randomness and Uncertainty

"The uncertainty associated with a stochastic model has two distinct sources: the inherent variability of the phenomenon, [and] incomplete knowledge... of the probabilities.... Sometimes these sources of uncertainty are referred to as 'process risk' and 'parameter risk,' respectively. The terms 'risk' and 'uncertainty,' respectively, have also been used...." [Committee on Principles, 1997] In this paper, the terms "randomness" and "uncertainty" are used.

Standard statistical theory, as taught to actuaries, offers many methods for fitting risk models (distributions) to data. With parametric models, there are a variety of techniques for estimating the parameters and assessing the uncertainty in those estimates. What is relatively lacking, however, is advice on how to incorporate uncertainty information into the risk model itself, or more generally, into the advice being given to the user of the risk model.

The predictive approach to probability modeling is one such method for embedding uncertainty (parameter risk) into the (process risk) model for a random variable. The random variable's assumed family of distributions and its parameters are augmented to include variation in the estimation process itself. A familiar example of this is the construction of a prediction interval for a yet-to-be observed time series or regression value. The formula for the variance of the predicted value includes terms for both the residual error (noise) variance and the variance of the estimator for the mean value. Another way of saying this is that the estimated risk curve for the random variable is modified somehow to account for the phenomenon of parameter uncertainty.

It is the purpose of this paper to critique the predictive approach (or indeed any model-embedded approach) to "taking uncertainty into account" in parameter estimation and risk curve construction. In so doing, it will emphasize that this is not an unambiguous operation. The desired form of the risk curve adjustment depends crucially on subtle details of the specific question being addressed, so much so that a risk curve can seem to be simultaneously overestimating and underestimating risk.

0.3 Contents

This paper consists of six parts. The remainder of the introduction discusses previous literature in this area. In particular, a seminal work by Kreps [1997] is summarized. Part 1 discusses estimation and bias in the context of probability distribution parameters and percentiles. To help clarify theory, an exponential example and a lognormal example are worked out in some detail. The lognormal example is the same one used by Kreps [1997]. Part 2 presents some motivation for "adjusting for uncertainty." Taking a Frequentist approach, it casts the issue in terms of a particular type of bias and works out the necessary – predictive – adjustment for the two examples. While Frequentist, it draws strong parallels to the Bayesian approach in Kreps [1997]. Part 3 extends the bias concerns of part 2 in other directions and reveals the existence of an apparent paradox, making the case against adjustment. Part 4 discusses confidence intervals as an alternative to "adjusting for uncertainty." Confidence intervals for parameters, percentiles, and exceedance probabilities are given for the two examples. Part 5 concludes with advice to the DFA practitioner.

0.4 Previous Research

Previous actuarial literature has addressed "parameter uncertainty," but it is sometimes not clear what the term is intended to encompass.

0.4.1 The View from PCAS 1983

Venter [1983] refers to the possibility of modeling "parameter risk" in the context of transformed gamma and beta models for losses where "because of uncertain trend (or other factors) there is substantial uncertainty about the scale parameter λ...." He goes on to suggest putting a gamma distribution on λ^{α} and mixing the loss distribution over λ, as a "practical technique for quantifying this uncertainty." The parameters for the distribution of λ itself can be estimated through percentile matching or, alternatively, an examination of industry or sub-sector loss ratios.

Meyers & Schenker [1983] and Heckman & Meyers [1983] discuss parameter uncertainty in the collective risk model. "Parameter uncertainty can arise from sampling variability and changes... over time.... [or] when some members of the group have different [expectations]." Their model uses a "contagion parameter" c in the claim count distribution and a "mixing parameter" b in the claim severity distribution. Specifically, λ, the expected number of claims (say, from a Poisson distribution), is multiplied by χ, a gamma-distributed random variable with mean 1 and variance c. Z, the claim amount, is divided by β, a gamma-distributed random variable whose inverse has mean 1 and variance b.

Meyers & Schenker [1983] provide three examples of fitting the parameters b and c to empirical data. In the most general form, their model treats r_i years of experience of insureds $i = 1,\ldots,T$ as manifesting T independent draws of the χ and β random variables. Their equations then estimate b and c through variance components (random effects ANOVA).

Thus, we seem to have three sources of parameter uncertainty which perhaps should be carefully distinguished: sampling error, nonstationarity, and heterogeneity. The recommended mathematical treatment is to interpret uncertainty as a hierarchical random effect. While this method admirably represents nonstationarity and heterogeneity, it does not appear to address sampling error. Sampling error is distinct from heterogeneity; it determines the accuracy with which b, c, λ, etc., can be estimated. The standard errors of the estimates will diminish with increasing numbers of insureds T. The values of b and c themselves, however, will not converge to zero with increasing T.

0.4.2 Kreps 1997

Kreps [1997] discusses parameter uncertainty in normal and lognormal distributions. In his introduction, he states "One of the most ubiquitous sources of parameter uncertainty is the fact that samples in real life are never infinite." Here, he is explicitly addressing sampling error, and develops a theory of predictive distributions "with" parameter uncertainty.[1] He concludes that "the effect of parameter uncertainty is to push probability away from the mean out into the tail." As will be seen below, the

---

[1] Mathematically, his technique is again to treat uncertainty as a hierarchical random effect, however, with the imprimatur of explicitly Bayesian justifications.

predictive approach can be interpreted as creating percentile estimators that are unbiased in a probabilistic sense.

For a case study, he analyzes Best's reserving data. IBNR is assumed to be distributed lognormally. Based on n=5 years, the maximum likelihood estimates of the mean and standard deviation of X = ln(IBNR) are 23.01923 and 0.06653, respectively. This "point estimate" implies a probability of IBNR exceeding $11.5 billion equal to 1.39%. For Kreps, taking parameter uncertainty into account, "the exact result... is 12.78%. To get to the true 1.39% level, it is necessary to reserve $14.1 billion!"

Subsequent sections will follow through on this example and parallels to Kreps's work will be sketched in more detail.

1. ESTIMATION

This section discusses the estimation of parameters and percentage points. While the estimation of parameters is the usual goal, the theory of point estimation applies equally well to the estimation of functions of the parameters. Because of the typical DFA interest in tail behavior of variables, the estimation of percentiles (specific points on the risk curve) is arguably more important than the estimation of parameters per se. At the very least, the choice of parameter estimation technique should be informed by the effect it has on percentile estimates. *Bias* is defined and illustrated in both parameter and percentile contexts. The concept of a *risk curve* is formally defined and examples are presented. The specific notions of X-unbiased risk curves and estimation techniques are defined and illustrated.

1.1 Estimation of Parameters

While various techniques are available for estimating parameters, we focus here on Maximum Likelihood due to its general applicability and widespread use. Consider a family of probability density functions $g(x;\theta)$ where x is a real variable and θ is a (possibly vector) parameter. Given a sample $\{x_1, x_2, \ldots, x_n\}$, the Maximum Likelihood Estimate (MLE) of the parameter θ is the value $\hat{\theta}$ that maximizes the joint likelihood

$$L(\theta) = \prod_{i=1}^{n} g(x_i;\theta). \tag{1.1}$$

The sampling distribution of $\hat{\theta}$ has (asymptotically, i.e. with large samples) a dispersion matrix equal to the inverse of the matrix of second derivatives (with respect to θ) of the natural log of the likelihood. Thus, standard errors of the MLE may be computed nearly as easily as the estimator itself. In many commonly-used families of distributions, the MLEs are the obvious moment estimators.

For the typical distributions in use by actuaries, MLEs are *asymptotically efficient*. This means that for large samples, they uniformly provide the most accuracy, regardless of the true parameter value. However, they tend not to have strong small-sample justifications [Lehmann, 1983].

1.1.1 The Exponential Case

Consider a random variable X distributed as exponential with scale parameter λ:

$$\Pr\{X \le x\} = F_X(x;\lambda) = 1 - \exp(-x/\lambda). \tag{1.2}$$

Given a sample $\{x_1, x_2, \ldots, x_n\}$, the likelihood function is given by

$$L(\lambda) = \prod_{i=1}^{n} \frac{1}{\lambda} \cdot \exp(-x_i/\lambda) = \left(\frac{1}{\lambda}\right)^n \cdot \exp\left(-\frac{1}{\lambda} \cdot \sum_{i=1}^{n} x_i\right). \tag{1.3}$$

Differentiating by λ and setting to zero, we can see that the value of λ that maximizes the likelihood is given by

$$T = \hat{\lambda} = \frac{1}{n} \sum_i x_i . \tag{1.4}$$

This is also the same estimator obtained by equating first moments of the theoretical distribution and the sample.

1.1.2 The Normal Case

Consider a random variable X distributed according to the normal cumulative distribution function:

$$\Pr\{X \le x\} = F\left(x, \begin{bmatrix} \mu \\ \sigma^2 \end{bmatrix}\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2 \cdot \pi}} \cdot \exp\left\{-\frac{1}{2} \cdot \frac{(z-\mu)^2}{\sigma^2}\right\} dz \tag{1.5}$$

The likelihood can be written

$$L\left(\begin{bmatrix} \mu \\ \sigma^2 \end{bmatrix}\right) = \exp\left\{-\frac{n}{2} \cdot \ln(2 \cdot \pi) - n \cdot \ln(\sigma) - \frac{1}{2} \cdot \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{\sigma^2}\right\} \tag{1.6}$$

Differentiating the expression inside the exponential and setting to zero, we get the so-called *likelihood equations*:

$$\frac{\partial \ln(L)}{\partial \mu} = \frac{1}{\sigma^2} \cdot \left(\sum_{i=1}^{n} x_i - n \cdot \mu \right) = 0 \tag{1.7}$$

$$\frac{\partial \ln(L)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \cdot \sum_{i=1}^{n} (x_i - \mu)^2 = 0$$

The solutions, the maximum likelihood estimators, consist of the sample mean and variance, respectively:

$$\hat{\mu} = \frac{1}{n} \sum_i x_i \tag{1.8}$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_i (x_i - \hat{\mu})^2$$

Again, this gets the same result as moment matching. For a lognormal variable Y = exp(X), the sample mean and variance of ln(Y) make up the MLE. This follows from an invariance property of MLEs.

In the example set out in Kreps [1997], we have the log of IBNR modeled as a normal distribution with $\hat{\mu}$ = 23.01923 and $\hat{\sigma}$ = 0.06653 based on n = 5 sample points.

1.2 Estimating a Percentile

Typically, actuarial risk calculations concern themselves with one tail of a distribution. In DFA, the "interesting" or "risky" behavior of the system will often be driven by the upper or lower extreme values of one or more key variables. For example, in the context of reserving, it is common to ask, what level of the loss variable will only be exceeded with specified low probability? This sort of quantity is also known in other financial disciplines as the *value at risk*.

The $100(1\text{-}q)^{th}$ percentile X_q of a distribution is given by solving $1\text{-}q = F_X(X_q,\theta)$. However, this requires knowing the true value of θ. In practice, we only have some estimator $\hat{\theta}$ of θ, therefore we are left with the problem of constructing estimators of X_q.

1.2.1 The Exponential Case

Given the parameter λ, it is readily determined that $X_q = -\lambda \ln(q)$. This suggests an obvious estimator:

$$\hat{X}_q = -T \cdot \ln(q)\,. \tag{1.9}$$

1.2.2 The Normal Case

For normal variables, $X_q = \mu + z_q\sigma$ where z_q is the $100(1\text{-}q)^{th}$ percentage point of the standard normal distribution, e.g., $z_{0.05}$=1.645. Again, this suggests an obvious estimator:

$$\hat{X}_q = \hat{\mu} + z_q \cdot \hat{\sigma} \tag{1.10}$$

For the lognormal, we simply transform by $\hat{Y}_q = \exp(\hat{X}_q)$. Kreps's example notes that the probability of exceeding Y = \$11.5 billion is 1.39% (if the estimated parameters are exactly correct). Equivalently, $\hat{X}_{0.0139} = 23.166$ or $\hat{Y}_{0.0139} = 11.5 \cdot 10^9$.

1.3 Bias in Parameter and Percentile Estimators

Since estimators are themselves random variables, it is meaningful to inquire into their sampling behavior (distributional properties). Imagine there are modelers, m = 1,...,M, each drawing an independent sample $\{x_{1m},\ldots,x_{Nm}\}$ from some fixed distribution.

Each modeler assumes (correctly) the form $F(x;\theta)$ of the distribution, but must estimate the parameter θ based solely on his or her own sample. Each modeler will then, presumably, have a different estimate for θ and some will get closer to the actual value of θ than others.

An estimator S for a quantity $f(\theta)$ is said to be unbiased if

$$E_{\theta}[S - f(\theta)] = 0 \tag{1.11}$$

where the notation $E_{\theta}[\,]$ denotes mathematical expectation with respect to the distribution characterized by θ. Note that θ, hence $f(\theta)$, is a fixed number and S is a random variable. In the example of the M modelers, unbiasedness means that the average estimate obtained among modelers, as M gets arbitrarily large, will converge to the true value of the parameter. Unbiasedness is only one property that an estimator may possess, and not having it does not necessarily make an estimator inferior to ones that do.[2]

Note that the definition of unbiasedness applies to estimators of any quantity associated with a distribution, parameters as well as percentiles, exceedance probabilities, etc.

---

[2] "Bias" is such a loaded word that statisticians would have been better off with a more technical term like "expectation neutrality." Alas, we are stuck with the baggage of historical usage.

1.3.1 The Exponential Case

The distribution of T, the MLE for the exponential scale parameter λ, can be shown to be a gamma with scale parameter λ/n and shape parameter n,

$$\Pr\{T \le t\} = \int_0^t \frac{\left(\frac{z}{\lambda/n}\right)^{N-1} \exp\left(-\frac{z}{\lambda/n}\right)}{(n-1)!\,\lambda/n} dz . \quad (1.12)$$

The mean of T is therefore λ, and the variance is λ^2/n. T is therefore an unbiased estimator for λ. Because T is unbiased for λ, $\hat{X}_q$ is also unbiased for X_q.

1.3.2 The Normal Case

The sample mean of a normal distribution is distributed as a normal with mean μ and variance σ^2/n, therefore it is unbiased for μ. The sample variance is distributed as σ^2/n times a χ^2(n-1) variable; the MLE for σ is therefore biased. We can distinguish several alternatives. If an unbiased estimate of the *variance* (σ^2) is desired, then we want the familiar

$$\hat{\sigma}_s = \sqrt{\frac{n}{n-1}} \cdot \hat{\sigma} \quad (1.13)$$

This gives us a value of 0.07439 in the Kreps example.

Unbiased estimation of the *standard deviation* (σ) is much less familiar to beginning students of statistics. Lehmann [1983] gives a general form for unbiased σ^k estimation[3] which specializes for k = 1 to:

$$\hat{\sigma}_s = \sqrt{\frac{n}{2}} \cdot \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \cdot \hat{\sigma} \tag{1.14}$$

This gives us a value of 0.07911 in the Kreps example.

We may generalize our percentile estimator by considering

$$\hat{X}_q = \hat{\mu} + z_q \cdot \hat{\sigma}_{()} \tag{1.15}$$

where we have a choice of estimators $\hat{\sigma}_{()}$ for σ. Recall that the ML estimator of the 1.39% exceedance point ($z_{0.0139}$ = 2.2) is X = 23.166 translating to an IBNR of Y = exp(23.166) = 11.5 billion.

An unbiased estimator for X_q uses $\hat{\sigma}_{()} = \hat{\sigma}_s$ which yields 23.193, translating to an IBNR of 11.82 billion. This is not unbiased for IBNR, however, because an unbiased X does not imply an unbiased *exp(X)*. This author is not aware of an unbiased estimator for Y_q. We can estimate the magnitude of the bias, however, by noting that if the normal parameters were indeed equal to their ML estimators, then, approximately,

---

[3] Johnson, Kotz, and Balakrishnan [1994] discuss the special case of k=1 and present a simpler approximation.

$$\frac{E\left[\exp(\hat{X}_{q,s})\right]}{Y_q} = \exp\left(\frac{\sigma^2}{2}\cdot\left[\frac{1}{n}+z_q{}^2\cdot\left\{\frac{n-1}{2}\cdot\left(\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\right)-1\right\}\right]\right) \tag{1.16}$$

where the subscript s indicates we are using the unbiased estimator for σ. This is only an approximation because it assumes that $\hat{\sigma}_s$ is distributed as a normal variable; for $\sigma<1$, however, it is accurate to within 5%. In our example, for values of σ in the neighborhood[4] of the ML value, the ratio of equation 1.16 is within 1.3% of unity, indicating little bias. However, for larger values of σ, the bias can be substantial.

1.4 The Risk Curve and X-Unbiasedness

We can present the results of many percentage point estimators in graphical form. The locus of points $\{<X_q,q>\}$ is known as the risk curve or exceedance probability (EP) curve. We place the exceedance probability q on the vertical axis and the percentile estimate $\hat{X}_q$ on the horizontal axis. Depending on the range of interest, we may want to plot one or both axes logarithmically. An alternative for the vertical axis is to plot the return period, 1/q, in units of time, e.g., years if the variable represents an annually measured quantity.

[4] Specifically, for values of the parameter within a two-tailed 90% confidence interval, as defined in section 4.1.

If, for every q, the percentile estimator $\hat{X}_q$ is unbiased, we say that the risk curve is X-unbiased, or unbiased in the X domain. If a parameter estimation technique leads to an X-unbiased risk curve, we will call it an X-unbiased technique.

1.4.1 The Exponential Case

Having developed T, the locus of points $\langle -T \cdot \ln(q), q \rangle$ is the ML risk curve. This risk curve is unbiased in the X domain. The MLE technique for exponentially-distributed data is thus X-unbiased.

1.4.2 The Normal Case

Depending on which $\hat{\sigma}_{()}$ is used, there are corresponding alternatives for the risk curve. Figure 1 shows the MLE-based curve as a thin solid line and the X-unbiased (approximately Y-unbiased) curve as a thick dotted line. For reference, the target <\$1.152 billion, 1.39%> probability point is marked with a box. Note that the two versions of the curve differ markedly. Around the reference point, the difference amounts to \$300mm on the dollar axis or 1.7% on the probability axis. The MLE technique for normally distributed data is therefore not X-unbiased, but an X-unbiased alternative, based on equation 1.14, is available.

2. THE CASE FOR ADJUSTMENT

Unbiased estimation in the X (log) or Y (dollar) domain may or may not be appropriate for the decisions to be made in a real application of the theory. For example, while the American Academy of Actuaries [1993] says, "Consideration must also be given to any [statistical] bias in the reserves or premiums," it doesn't specify in what manner this consideration should be given. This section considers a different sort of bias, leading to the notion of P-unbiasedness, and how that can be achieved through the predictive distribution approach.

2.1 Probabilistic Bias and Predictive Bounds

We can ask a slightly different question about estimators for X_q (equivalently, Y_q), based on the property they purport to represent, namely, an exceedance probability of q: What is the expected value of this probability? In particular, we might like estimators that are "probabilistically unbiased" (P-unbiased) in the sense that

$$E\left[1 - F(\hat{X}_q, \theta)\right] = q . \qquad (2.1)$$

Such probabilistically unbiased estimators do exist. They are known as *prediction bounds*, because

$$E\left[F(\hat{X}_q, \theta)\right] = \Pr\left\{X \le \hat{X}_q\right\} \qquad (2.2)$$

where X is another draw from the population, independent of the sample upon which the estimator is based. Since X_q is the point satisfying $1\text{-}q = F_X(X_q,\theta)$, if $\hat{X}_q$ is unbiased for X_q, it is natural to assume that the probability of $X > \hat{X}_q$ is also equal to q. This is not generally the case: X-unbiasedness does not imply P-unbiasedness. By establishing the true "predictive probability" of an estimator $\hat{X}_q$

$$\wp(q,\theta) = \Pr\left\{X > \hat{X}_q\right\} = E_\theta\left[1 - F_X(\hat{X}_q;\theta)\right] \qquad (2.3)$$

we might be able to solve for an adjusted q* satisfying $\wp(q^*,\theta) = q$. Then, $\hat{X}_{q^*}$ may serve as a P-unbiased estimator for X_q. Other routes are available, also. If, for every q, the percentile estimator $\hat{X}_{q^*}$ is P-unbiased, we say that the risk curve is P-unbiased. If a parameter estimation technique leads to an P-unbiased risk curve, we will call it an P-unbiased technique.

2.1.1 The Exponential Case

The predictive probability for an exponential percentile MLE is independent of the parameter:

$$\Pr\left\{X > \hat{X}_q\right\} = \left(\frac{n}{n - \ln(q)}\right)^n. \tag{2.4}$$

For example, with n = 20 and nominal q = 0.01, the true predictive probability is 0.016.

Inverting the relationship, we get the adjusted q* for a "probabilistically unbiased" $\hat{X}_{q^*}$:

$$q^* = \exp\left(n \cdot \left(1 - q^{-1/n}\right)\right). \tag{2.5}$$

For example, with n = 20 and q = 0.01, the computed q* = 0.006. The adjusted (P-unbiased) risk curve is then the plot of $\left\langle -T \cdot n \cdot \left(1 - q^{-1/n}\right), q\right\rangle$.

Can we find a P-unbiased estimator for the exponential parameter? In other words, can we compute T in such a way that the straightforward $\hat{X}_q$ from equation 1.9 is

P-unbiased? Not in general; there is no solution S to the equation $-S \cdot \ln(q) = -T \cdot n \cdot \left(1 - q^{-1/n}\right)$ that holds for all 0<q<1 simultaneously (although as n increases without bound, $S=T$ is an asymptotic solution). This means that there is no parameter estimation technique *within the exponential distribution* that yields a P-unbiased risk curve.

That is because the predictive distribution for an exponential variable is not an exponential distribution, it is a Pareto! This can be seen by solving $X = -T \cdot n \cdot \left(1 - q^{-1/n}\right)$ for q in terms of X:

$$q = \left(1 + \frac{X}{T \cdot n}\right)^{-n}. \tag{2.6}$$

In summary: to create an X-unbiased risk curve from presumed exponential data, first determine the MLE T of the exponential parameter as in equation 1.4. Then substitute T for λ in equation 1.2. This is not P-unbiased, however, because the true exceedance probability at an estimated percentile is affected by parameter estimation uncertainty. For a P-unbiased risk curve, construct the Pareto distribution corresponding to equation 2.6. Drawing simulated values X from the Pareto instead of the exponential will "take uncertainty into account" in the sense that the true exceedance probabilities of the simulated percentage points will be accurate in expectation.[5]

---

[5] Here, "in expectation" means "averaged over all random samples of data from the same exponential population."

2.1.2 The Normal Case

A prediction bound which a single future, independently selected normal variable will not exceed with probability q is given by:

$$\tilde{X}_q = \hat{\mu} + t_{q,n-1} \cdot \hat{\sigma}_s \cdot \sqrt{1 + \frac{1}{n}} \tag{2.7}$$

where t is the 100(1-q)th percentile of a Student t distribution with n-1 degrees of freedom. For our example of n = 5 and q = 1.39%, we get t = 3.379 and the prediction bound is $\tilde{X}$ = 23.295, corresponding to $13.08 billion.

This must mean the estimators in section 1 are probabilistically biased. Indeed, by setting $\tilde{X}_{q^*} = \hat{X}_q$ (equations 2.7 and 1.15, respectively) we may compute the predictive probability q^* corresponding to the nominal q probability for the estimator $\hat{X}_q$. The ML estimator for the q = 1.39% exceedance point, shown previously to be biased in the log domain, has an expected actual exceedance of q^* = 7.34%. Thus, in probabilistic terms, it is drastically biased – downward – underestimating the tail risk.

What about the (log) unbiased estimator based on $\hat{\sigma}_s$? This is a little better, with expected actual exceedance of q^* = 4.98%, but it is still far from being unbiased in the probabilistic sense.

Figure 2 adds $\langle \tilde{X}_q, q \rangle$ to the plot of risk curves as a dash-dot line. This represents a dramatic increase in estimated risk. Values of IBNR exceeding $1.35 billion, essentially inconceivable according to the MLE and dollar-unbiased curves (20,000-year return period or higher), are now seen as a distinct possibility (100-year return period).

Is there a P-unbiased estimator for σ? As with the exponential case, no. The predictive distribution is from the Student t family, not the normal family (although, again, in the limit as n increases without bound, there is convergence). To create an X-unbiased (or log-unbiased) risk curve from presumed normal data, the methods of section 1.3.2 suffice. For a P-unbiased risk curve, however, one must construct the Student t distribution corresponding to equation 2.7.

2.2 Discussion

A specific family of distributions will lead to a specific form for the predictive distribution. However, there is an approximation method which can bypass the analysis. By sampling the parameters (according to an estimate of their distribution) as well as the object random variable (according to the particular parameter values selected in their most recent draw), one can create a random variable drawn from a mixture.[6] This mixture represents a predictive distribution insofar as it incorporates variability in the random variable (process risk) as well as uncertainty in the parameters (parameter risk).

Making this sort of adjustment – analytically or numerically – is often what is meant by "taking uncertainty into account." Notice correcting this new sort of bias is a matter of increasing an understated (on average) risk. For typical actuarial distributions with decreasing density in the upper tail and small enough q, on average, the true exceedance probability $1 - F_X(\hat{X}_q, \theta)$ for the quoted value of an unbiased estimator $\hat{X}_q$

[6] cf. Venter's recommendation discussed in section 0.4.1.

will be higher than the nominal probability q from which the estimate is developed. The adjusted value $\hat{X}_{q^*}$ will therefore be higher (farther up in the tail); this is why it is often claimed that "uncertainty fattens the tails."[7]

Why does this happen? The function F_x is nonlinear in its X argument. Values of an X-unbiased $\hat{X}_q$ deviate from the true value in a balanced fashion between high and low; the average is the true value X_q. However, a deviation on the high side contributes less to the expectation of $F_X(\hat{X}_q,\theta)$ than an equally large deviation on the low side diminishes it, due to the curvature of F_x. Therefore the expectation is not the same as the function evaluated at the true value X_q.

As mentioned in section 0.4.2, Kreps [1997] addresses this issue from a Bayesian perspective. His result for the "true" 1.39% exceedance point is $14.1 billion, about a billion higher than was calculated in section 2.1.2. It is interesting to note that Kreps [1997] summarizes his computations of percentage points with analogous expressions $\hat{\mu} + z \cdot \hat{\sigma}$ involving the MLEs of the parameters. For the MLE of the percentage point, z is the corresponding percentile of a unit normal. For the predictive distribution, Kreps's z is z_{eff}, the percentile from a normal with variance $(n+1)/(n+\tau-4)$, where τ is a parameter defining the "uninformative" Bayesian prior distribution on σ, typically 0 or 1 (he used zero). Since a t distribution with ν degrees of freedom has variance ν/(ν-2), the

[7] cf. Kreps's comment, discussed in section 0.4.2.

equivalent Frequentist coefficient $t_{q,n-1} \cdot \sqrt{\frac{n+1}{n-1}}$ (derived from equation 2.7) can be considered analogous to the Bayesian z_{eff} with $\tau = 1$.

Bayesians feel free to treat uncertainty in the parameters on an equal footing with the stochastic behavior of the random variable. Above, we saw how Frequentist mathematics can, in effect, yield the same results. If probabilistically unbiased estimation (or simulation) is the goal, it is appropriate to utilize the predictive distribution, rather than the ML-estimated distribution, to look up percentiles (or generate random variables). This is the Frequentist rationale for "adjusting the risk curve for uncertainty."

3. THE CASE AGAINST ADJUSTMENT

In this section, the search for hidden forms of bias continues. The concept of Q-unbiasedness will be defined. It will be seen that the adjustments of section 2 can lead to worsening of estimator behavior with respect to Q-unbiasedness. Moreover, it will be seen how it is typically impossible to make an adjustment which simultaneously improves the two competing measures of bias.

3.1 Estimating Exceedance Probabilities and Q-Unbiasedness

Rather than divulge a dollar limit X_q corresponding to a given exceedance probability q, we may view a risk curve as telling us a probability Q_L of exceeding a specific threshold L. This might be the perspective, say, in a ruin-theoretic analysis. The decisionmaker could have in mind that \$1.152 billion is the most that could be lost without dire consequences, and might request an estimate of the probability of suffering them. As far as the geometry of the risk curve is concerned, this new situation is simply a matter of entering the graph from a different axis, treating the locus of points as $<L,Q_L>$ rather than $<X_q,q>$.

If an estimator $\hat{Q}_l$ is unbiased, we will say that a risk curve constructed from such estimators is Q-unbiased. If a DFA model aims at constructing risk curves for both X_q and Q_L lookups, then Q-unbiasedness and P-unbiasedness are arguably equally desirable.

A natural point estimator is $\hat{Q}_L = 1 - F_X(L,\hat{\theta})$. Indeed, if $\hat{\theta}$ is the MLE of θ, then $\hat{Q}_l$ is the MLE of Q_L. It should come as no surprise that $E_\theta\left[\hat{Q}_L\right]$ does not in general equal Q_L, again, due to nonlinearity of F_X – this time in its θ argument.

In the two examples it will be seen that, on average, the estimated exceedance probability $\hat{Q}_L$ for the specified loss threshold L will be *higher* than the true probability Q_L. To correct for this bias, an adjusted probability estimate $\hat{Q}_L^*$ will have to be *lower* than the estimate $\hat{Q}_L$ computed from maximum likelihood. Thus, this variety of bias is in the direction of *overstating* the risk, in marked contrast with the case of the previous section, which understated the risk. Q-unbiasedness is not the same as P-unbiasedness. In a sense, they are duals, if not opposites, of each other.

3.1.1 The Exponential Case

The point estimator $\hat{Q}_L$ is obtained from equation 1.2 as exp(-L/T). This is biased, and Johnson, Kotz, and Balakrishnan [1994] give the minimum variance unbiased (MVU) estimator as:

$$\hat{Q}_{LUB} = \left(1 - \frac{L}{n \cdot T}\right)^{n-1}. \tag{3.1}$$

This represents the risk curve as a form of beta distribution. As with P-unbiasedness, there is no estimator of the exponential parameter to make a Q-unbiased exponential risk curve. However (and again, similarly), in the limit as n increases without bound, equation 3.1 approaches an exponential. With λ = 1, n = 20, and L = 4.605, the true value of Q_L is 1%; were T to equal λ, this estimator would produce the value 0.69%.

This estimator has the unfortunate property that if L is greater than nT then the estimated exceedance probability is zero, making very-high-tail estimates impractical. By taking a Taylor expansion, we may approximate

$$E_\lambda\left[\hat{Q}_L\right] = E_\lambda\left[1 - F_X(L;T)\right] \approx \exp\left(-\frac{L}{\lambda}\right)\left(1 + \frac{L(L - 2 \cdot \lambda)}{2 \cdot n \cdot \lambda^2}\right) \tag{3.2}$$

Unfortunately, the "bias correction" term in this approximation is dependent on the true value of λ, which is unknown. By substituting T for λ, we may compute an approximately unbiased estimate as:

$$\hat{Q}_L{}^* = \exp\left(-\frac{L}{T}\right) \bigg/ \left(1 + \frac{L(L-2\cdot T)}{2\cdot n\cdot T^2}\right). \tag{3.3}$$

For tail thresholds L greater than twice the estimated mean T, the denominator is greater than one and the estimated probability is therefore less than the MLE. In this numerical example (n = 20, L = 4.605, and λ = 1), simulation shows this estimator to average 1.1% versus the true 1%. For n = 20, L = 4.605, and T = 1, this adjusted estimator produces 0.77%, about three-fourths of the ML-estimated probability, and 11% higher than the MVU estimator.

3.1.2 The Normal Case

Again we have a variety of estimates

$$\hat{Q}_{()}(x) = 1 - \Phi\left(\frac{x-\hat{\mu}}{\hat{\sigma}_{()}}\right) \tag{3.4}$$

available, depending on the estimator used for σ. Here, Φ is the cumulative (standard) normal probability function corresponding to the integral in equation 1.5. The ML version of this estimator gives us an exceedance probability estimate at L = $11.5 billion ($x$ = ln(L)) of 1.39%.

At this point, readers should not be surprised to learn that the MLE is biased. Lehmann [1983] and Johnson, Kotz, and Balakrishnan [1994] provide the minimum variance unbiased estimator for the exceedance probability of a normal distribution:

$$\hat{Q}_{UB}(x) = 1 - \sqrt{\frac{n}{n-1}} \cdot \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\cdot\Gamma\left(\frac{n-2}{2}\right)} \cdot \int_{-S}^{U}\left(1 - \frac{n}{n-1}\cdot z^2\right)^{n/2-2} dz \qquad (3.5)$$

$$\textit{where} \quad S = \sqrt{\frac{n-1}{n}} \quad \textit{and} \quad U = \min\left(S, \frac{x-\hat{\mu}}{\sqrt{n}\cdot\hat{\sigma}}\right)$$

Unfortunately, for values of x such that U = S, (which includes our numerical example) this estimator takes on the value of zero. Again, this is likely to be unacceptable in the typical actuarial application.

Alternatively, we can, by numerical integration or simulation, estimate the bias of the MLE (assuming various parameter values). For parameters taking on their MLE values in our example, the expected value for the ML exceedance probability estimator is approximately 1.83%, versus the hypothesized 1.39% – a ratio of 1.3. For other parameter values in the neighborhood[8] of the ML values, this ratio is at least 0.9, usually greater than one, often greater than two, and sometimes greater than 10. This means we should suspect the MLE of being biased high in the situation representing our data, that is, *over*estimating the tail risk. This is in contrast to the MLE percentile estimator, which was biased low, *under*estimating the tail risk.

What about the alternative estimators? Using an unbiased estimator for σ, we get an expected exceedance estimate (again, assuming parameters at the ML values) of 3.17%, high by a factor of 2.3, substantially worse. This is because the unbiased estimate

---

[8] See footnote 4.

of σ is greater than the ML estimator, decreasing the Z-score, hence the cumulative probability, and hence increasing the exceedance probability.

What about inverting the prediction bound equation? This is the equivalent of "looking up" exceedance probabilities from the predictive distribution. This is worse still, with an expected exceedance estimate of 6.13%, high by a factor of 4.41.

Applying the same strategy as with the exponential distribution, we can take a 2nd order Taylor series approximation to the exceedance probability and express the relative bias as

$$\frac{E_{\mu,\sigma}\left[\hat{Q}_L\right]}{Q_L} = 1 + \frac{1}{Q_L} \cdot \frac{\exp\left(-\frac{1}{2} \cdot z^2\right)}{\sqrt{2 \cdot \pi \cdot n}} \cdot W \tag{3.6}$$

$$where \quad W = \left(5 \cdot z - 2 \cdot z^3\right) \cdot \frac{\sqrt{2 \cdot n} \cdot \Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} + (2 \cdot n - 1) \cdot z^3 - 5 \cdot z \cdot n + 3 \cdot z .$$

$$and \quad z = \frac{L - \mu}{\sigma}.$$

As in the case with the exponential, we may substitute the ML estimators for μ and σ, obtaining

$$\hat{Q}_{ML} = 1 - \Phi\left(\frac{x - \hat{\mu}}{\hat{\sigma}}\right) \tag{3.7}$$

$$\hat{Q}^* = \frac{\hat{Q}_{ML}}{1 + \frac{1}{\hat{Q}_{ML}} \cdot \frac{\exp\left(-\frac{1}{2} \cdot \hat{z}^2\right)}{\sqrt{2 \cdot \pi \cdot n}} \cdot \hat{W}}$$

This estimator, while not exactly unbiased, does manage to shrink the bias of the ML estimator, typically by 60-90% in the neighborhood of the ML values.

Figure 3 extends our family of candidate risk curves to include Q* and Q_{UB}.

3.2 The Paradox

We have seen that to fit a model to data from an assumed distribution, the ML estimators of parameters led us to straightforward construction of risk curves. However, "taking uncertainty into account" in the parameter estimates led us to a profusion of sometimes opposing adjustments.

The ML estimators *under*estimate tail risk in one or two ways. First, the MLE of a normal 100(1-q)th percentile (for small enough q) is, on average, too low. Second, even the unbiased version (or the naturally unbiased estimator in the case of an exponential distribution) still provides "too low" of an estimate because the true exceedance probability of this estimator (the predictive probability) is, on average, greater than the specified amount q.

On the other hand, an MLE of exceedance probability at a (high enough) prespecified threshold is, on average, too high, thereby *over*estimating the tail risk. The substitution of a predictive distribution, corresponding to the probabilistic bias correction for estimating percentiles, is even *more* biased than the MLE.

The search to achieve simultaneous X-, P-, and Q-unbiasedness, even approximately, leads us in conflicting directions.

Consider the implications of this in practice. An actuary has performed a Dynamic Financial Analysis of a client's balance sheet. Numerous sources of random variation in liability and asset values were modeled, each of them having been fit to historical data. After explaining the methodology and walking through various charts and tables, the actuary summarizes: "There appears to be a 1% chance that your surplus will experience a drop exceeding $1 billion."

In an atypical response, the client might remind the actuary that there is sampling error in the various historical estimates and that actual probability distributions may well be different from the point estimates used in the model. Is this not another source of risk? Should the analysis not be adjusted to "take uncertainty into account?"

They meet a few days later, after the actuary has had a chance to enhance the old "Certainty Model" to include uncertainty adjustments. The following dialog between the client (C) and the actuary (A) ensues.

> *C: OK, now that you've taken uncertainty into account, what is an unbiased estimate of my 1% exceedance point?*
>
> *A: It's still $1 billion. That's an unbiased estimate.*
>
> *C: But isn't it true that exceedance points including sampling error should be higher than exceedance points without?*
>
> *A: Yes, that makes sense. See, the probability of your experiencing a loss greater than the point the Certainty Model picks out as the 1% point, that is to say, the predictive probability, is actually greater than 1%, so the $1 billion figure is probably too low. A better answer is more like $1.2 billion.*
>
> *C: So an unbiased estimate of the probability of exceeding $1 billion is actually greater than 1%?*
>
> *A: No, exceedance estimation in the Certainty Model is biased upwards. An unbiased estimate of the probability of exceeding $1 billion is more like 0.8%.*

C: First you tell me $1 billion is an unbiased estimate for the 1% point. Then you tell me the risk is worse, that the probability is actually greater than 1%. Then you tell me the risk is better, that an unbiased estimate is less than 1%. Now tell me why I shouldn't report you to the Actuarial Board for Counseling and Discipline!

What, then is the correct response? How is uncertainty to be taken into account?

4. CONFIDENCE INTERVALS

The classical approach to expressing parameter uncertainty is through summaries of the estimator distributions, either moments or selected percentage points. The latter become *confidence intervals* when couched in terms of the probability that the quoted percentage points bracket the true quantity.

Following Hahn & Meeker [1991], we may define a confidence interval as an interval bracketed by two estimators (functions of the sample data) intended to contain an unknown characteristic of the sampled population. Such a characteristic could be a parameter of the distribution, e.g., the mean or standard deviation of a normal distribution, or a function of those parameters, e.g., a percentile or an exceedance probability. The interval will contain the true value of the characteristic with a specified "confidence," e.g., 99%. This can be interpreted by Frequentists in terms of sampling, because the interval endpoints are random variables.[9] If independent samples were repeatedly drawn, and the interval computed from the samples, then the interval would contain the true value of the characteristic with the specified frequency, e.g., 99% of the time.

4.1 CI for Parameters

4.1.1 The Exponential Case

A 100(1-α)% confidence interval for the exponential parameter θ is given by

$$\hat{\theta}_{\pm} = \frac{T \cdot n}{\gamma_{\left(\frac{1}{2} \mp \frac{1-\alpha}{2},\, n\right)}} \tag{4.1}$$

where γ is the 100(α/2)th or 100(1-α/2)th percentile from a gamma distribution with shape parameter n. For our numerical example with T=1, the interval is [0.717, 1.509].

4.1.2 The Normal Case

A 100(1-α)% confidence interval for the mean is given by

$$\hat{\mu}_{\pm} = \hat{\mu} \pm t_{1-\frac{\alpha}{2}, n-1} \cdot \frac{\hat{\sigma}}{\sqrt{n-1}} \tag{4.2}$$

---

[9] Interpretation is even easier for Bayesians, because they are free to treat the parameters themselves as random variables.

where t is the 100(1-α/2)th percentile from a Student's t distribution with n-1 degrees of freedom. A 90% confidence interval for the mean in Kreps's reserving example (see previous section) is therefore [22.948, 23.090].

A 100(1-α)% confidence interval for the standard deviation is given by

$$\hat{\sigma}_{\pm} = \hat{\sigma} \cdot \sqrt{\frac{n-1}{\chi^2_{\left(\frac{1}{2} \pm \frac{1-\alpha}{2}\right),(n-1)}}} \tag{4.3}$$

where χ^2 is the 100(α/2)th or 100(1-α/2)th percentile from a chi-square distribution with n-1 degrees of freedom. A 90% confidence interval for σ in our example is [0.0483, 0.1765].

4.2 CI for Percentiles

4.2.1 The Exponential Case

Since the exponential is defined by only one parameter θ, a confidence interval for a q-exceedance percentile can be obtained directly from the confidence interval for the parameter by substituting endpoints:

$$\hat{X}_{q\pm} = \frac{-T \cdot n}{\gamma_{\left(\frac{1}{2} \pm \frac{1-\alpha}{2},n\right)}} \cdot \ln(q) \tag{4.4}$$

4.2.2 The Normal Case

Since the normal is defined by two parameters that must be estimated, the situation is a bit more complex. A 100(1-α)% confidence interval for X_q is given by

$$\hat{X}_{q\pm} = \hat{\mu} + g'_{\frac{1}{2}\pm\frac{1-\alpha}{2},1-q,n} \cdot \hat{\sigma}_v \tag{4.5}$$

where tables of g' are available in Hahn & Meeker [1991]. More complete tables, as well as the underlying theory based on the noncentral t, are available in Odeh & Owen [1980]. Johnson, Kotz, and Balakrishnan [1994] also give the distribution of $\hat{X}_q$ in terms of the noncentral t. A 90% confidence interval for the 1.39% exceedance point in the example is [23.113, 23.43]. This translates to an IBNR interval of [10.91 billion, 14.98 billion].

4.3 CI for Exceedance Probabilities

4.3.1 The Exponential Case

Again, since the exponential is defined by only one parameter θ, a confidence interval for exceedance probabilities can be obtained directly from the confidence interval for the parameter by substituting endpoints:

$$\hat{Q}_{L\pm} = \exp\left(-\frac{L \cdot \gamma_{\left(\frac{1}{2}\pm\frac{1-\alpha}{2},n\right)}}{T \cdot n} \right). \tag{4.6}$$

4.3.2 The Normal Case

A 100(1-α)% confidence interval for Q is given by

$$\left[\hat{Q}_{L-}, \hat{Q}_{L+}\right] = \left[h_{1-\frac{\alpha}{2},\frac{x-\hat{\mu}}{\hat{\sigma}_v},n} \; , \; 1 - h_{1-\frac{\alpha}{2},-\frac{x-\hat{\mu}}{\hat{\sigma}_v},n} \right] \tag{4.7}$$

where values of h are tabulated in Odeh & Owen [1980]. For the reserving example, a 90% confidence interval for exceeding Y = \$11.5 billion is [0.000617, 0.28351]. This is a stupefyingly large confidence interval, encompassing a factor of 459 between the two extremes. Figure 4 adds the upper and lower 90% confidence risk curves to the previous risk curves.

5. CONCLUSION

This paper examined the general problem of estimating parameters of probability distributions and the specific problem of estimating the actuarially interesting percentage points and exceedance probabilities as captured in the notion of a "risk curve." The choice of risk curve translates directly into the generation of random variables in DFA if the inverse lookup method is used, or, indirectly, as it affects the selection of distributional parameters for other methods. In particular, the paper showed how parameter uncertainty (parameter risk), stemming from sampling variability, can induce bias in estimators. It presented three varieties of bias that a risk curve could exhibit, depending on what aspect of the curve is considered relevant. It demonstrated that, at least in the common examples of exponential and normal/lognormal distributions, there is no way to correct these biases, even approximately, in a single "uncertainty-adjusted" risk curve. The conclusion, that a risk curve estimation procedure can seem to be simultaneously overestimating or underestimating risk, appeared as something of a paradox.

The resolution of this paradox is to examine our intuitive expectations about uncertainty-induced bias.[10] It is not the case that a single "uncertainty-adjusted" curve can replace the "point-estimated" curve, yielding better estimators all the way around. Uncertainty (parameter risk), it seems, cannot be put on a par with randomness (process risk). The problem is inherent in the nature of parameter uncertainty; like a carpet too big for a room, attempts to "flatten it out" in one spot will only make it "bulge up" somewhere else.

The solution that would-be DFA model builders should consider is to make explicit the distinction between uncertainty and randomness by placing (uncertainty) confidence intervals around the (randomness) estimates. For directly fitted distributions, confidence intervals can be calculated as was done in section 4. For DFA outputs, the situation is not so straightforward. The model can be "stress tested" by substituting extreme, but not implausible (see section 4.1), values of the parameters (equivalently, versions of the risk curve) and observing how the results change. More thoroughly, multiple runs, with parameters selected randomly from estimates of their distributions (again, refer to section 4.1) and fixed within each run, can provide multiple versions of the results. These multiple results can be summarized in terms of percentiles of their empirical distribution, giving, in effect, confidence intervals on the model outputs.

---

[10] Bayesians would say that the resolution is to not be concerned about bias; that bias as a statistical concept is problematical per se. I suspect few actuaries would feel totally comfortable with this advice.

After showing a client stress test or confidence interval results, bias can be addressed according to the particular goals of the problem. Given that bias is typically small compared to confidence intervals, a proper appreciation of confidence intervals would tend to dampen concern over the minutiae of bias adjustments.

REFERENCES

American Academy of Actuaries [1993]. *Report on Reserve and Underwriting Risk Factors*, Property/Casualty Risk-Based Capital Task Force.

Committee on Principles [1997]. *General Principles of Actuarial Science Discussion Draft*, Casualty Actuarial Society and Society of Actuaries.

Hahn, Gerald J. and William Q. Meeker [1991]. *Statistical Intervals: A Guide for Practitioners*, New York: John Wiley & Sons.

Heckman, Philip and Glenn Meyers [1983]. "The Calculation of Aggregate Loss Distributions from Claim Severity and Claim Count Distributions," Proceedings of the Casualty Actuarial Society LXX.

Johnson, Norman L., Samuel Kotz, and N. Balakrishnan [1994]. *Continuous Univariate Distributions, volume 1*, New York: John Wiley & Sons.

Kreps, Rodney E. [1997]. "Parameter Uncertainty in (Log)normal Distributions," presented at the 1997 Annual Meeting of the Casualty Actuarial Society.

Lehmann, E. L. [1983]. *Theory of Point Estimation*, New York: John Wiley & Sons.

Meyers, Glenn and Nathaniel Schenker [1983] . "Parameter Uncertainty in the Collective Risk Model," Proceedings of the Casualty Actuarial Society LXX.

Odeh, Robert E. and D. B. Owen [1980]. *Tables for Normal Tolerance Limits, Sampling Plans, and Screening*, New York: Marcel Dekker.

Venter, Gary [1983]. "Transformed Beta and Gamma Distributions and Aggregate Losses," Proceedings of the Casualty Actuarial Society LXX.

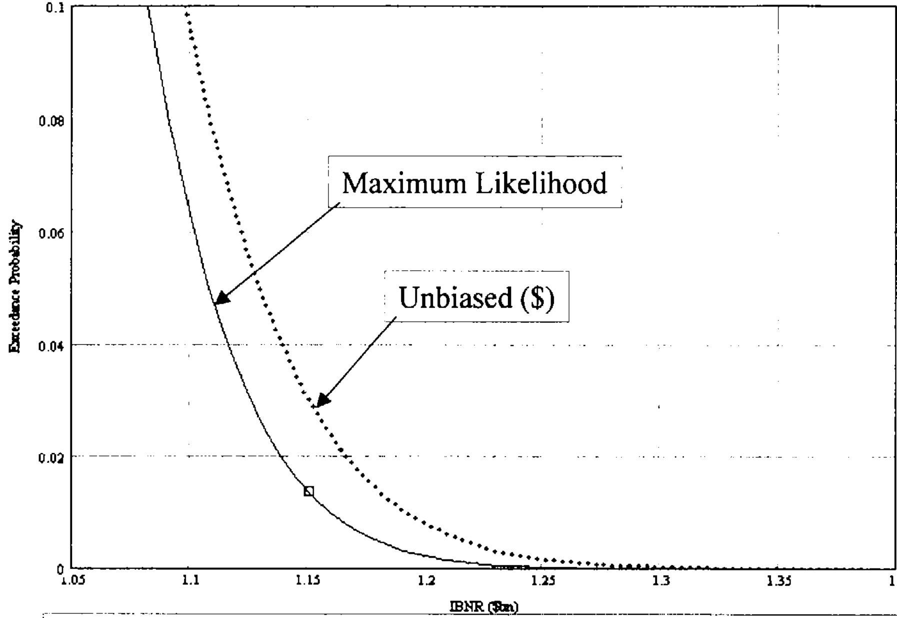


Figure 1: Risk Curves from Lognormal Example (Kreps, 1997)

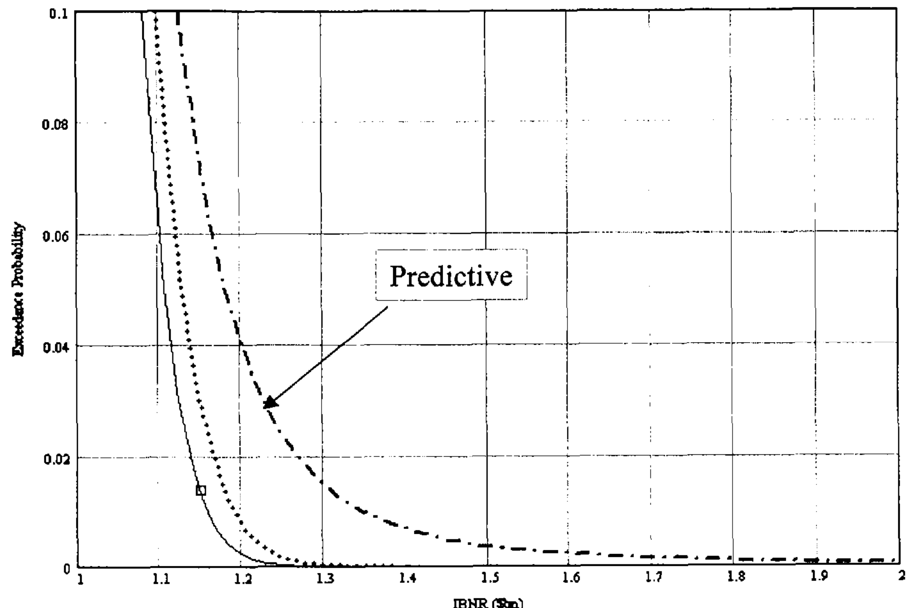


Figure 2: Adding Predictive Distribution Risk Curve

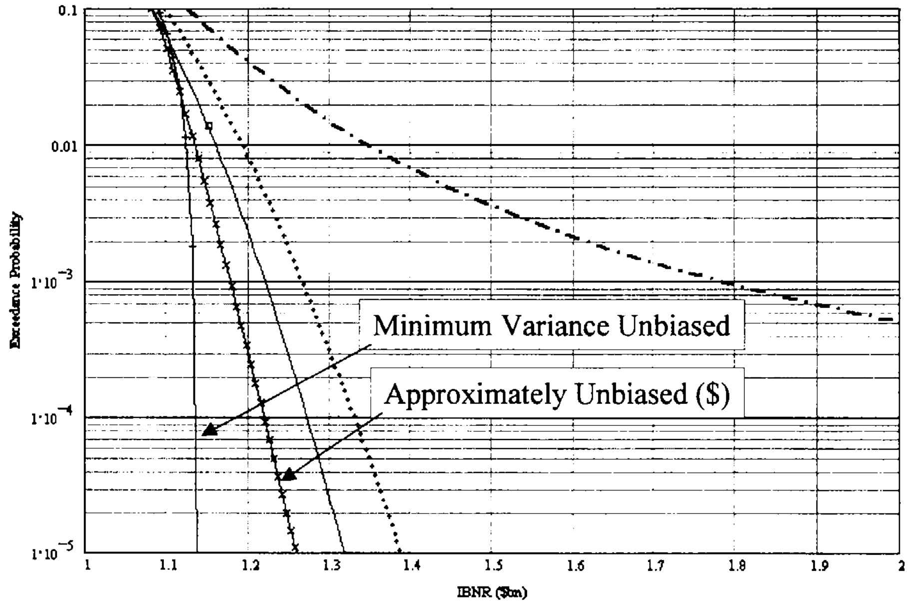


Figure 3: Adding Curves Unbiased for Exceedance

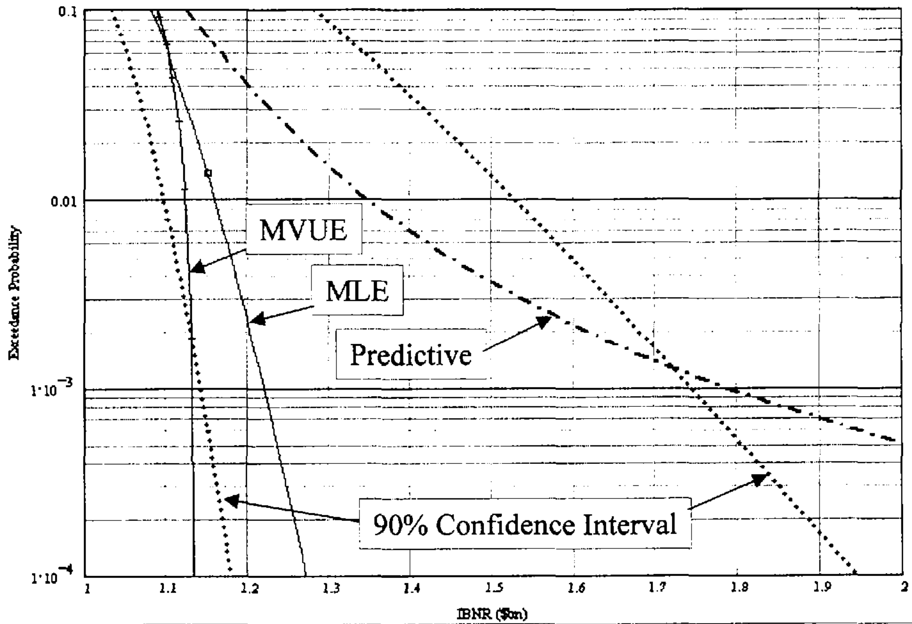


Figure 4: Confidence Intervals Compared to Estimators

Estimating Between Line Correlations Generated by Parameter Uncertainty

Glenn Meyers, FCAS, MAAA

Estimating Between Line Correlations

Generated by Parameter Uncertainty

by

Glenn Meyers
Insurance Services Office, Inc.

Abstract

When applying the collective risk model to an analysis of insurer capital needs, it is crucial to consider the effect of correlation between lines of insurance. Recent work sponsored by the Committee on the Theory of Risk has sparked the development of methods that include correlation in the collective risk model. One of these methods is built around the view that correlation is generated by parameter uncertainty affecting several lines of insurance simultaneously.

This paper uses simulation analyses to explore the properties of both classical and Bayesian methods of quantifying parameter uncertainty. We conclude that in order to get sufficient accuracy to determine the necessary capital, one must use the combined data of several insurers. Using the combined data of several insurers forces us to consider a collective risk model where parameter uncertainty affects several insurers – as well as several lines of insurance – simultaneously.

1. Introduction

The collective risk model has long been one of the primary tools of actuarial science. One can view this model as a computer simulation where one first picks a random number of claims and then sums the random loss amounts for each claim.

The early uses of the collective risk model were mostly theoretical illustrations of the role of insurer surplus and profit margins. Such illustrations are still common today in insurance educational readings such as Bowers, Gerber, Jones, Hickman and Nesbitt [1997, Ch 13].

By the late 1970's, members of the Casualty Actuarial Society were beginning to use the collective risk model as input for real-life insurance decisions. The early applications of the collective risk model included retrospective rating, e.g. Meyers [1980], and aggregate stop loss reinsurance, e.g. John and Patrik [1980] which is also described by Patrik [1996]. Bear and Nemlick [1990] provide further examples of the use of the collective risk model in the pricing of reinsurance contracts. Meyers [1989] begins to apply the collective risk model to an analysis of insurer capital.

This paper is part of a collective effort to extend the use of the collective risk model to Dynamic Financial Analysis (DFA). One goal of DFA is the management of an insurer's capital. An insurer requires sufficient capital so that its chance of insolvency is reasonably remote. An insurer can manage its capital needs by structuring its business so that it has an acceptably remote chance of a large loss. This structuring can include the use of reinsurance.

While the collective risk model arose from theoretical exercises in insurer solvency, it has not been widely used in practice for setting solvency standards. The main reason for this has been that it requires that individual lines of insurance be independent. Almost nobody believes this to be true. And as we shall demonstrate below, assuming independence can lead to a significantly understated solvency standard.

Recognizing this problem, the CAS Committee on the Theory of Risk commissioned Dr. Shaun Wang to develop versions of the collective risk model that do not require one to assume independence between lines of insurance. This work led to a paper titled "Aggregation of Correlated Risk Portfolios: Models & Algorithms" which is to appear in the next volume of the *Proceedings of the Casualty Actuarial Society.*

Inspired by Dr. Wang's work, we followed with a discussion of his paper, Meyers [1999], that focused on a version of the collective risk model where the claim count distribution for each line of insurance was conditionally independent given a parameter α. Treating α as a random variable leads to a particular kind of dependence between lines of insurance.

In this paper we propose a methodology for estimating the variance of α and explore the data requirements necessary to provide reliable estimates of this variance.

2. The Collective Risk Model

For the h^{th} line of insurance let:

μ_h = Expected claim severity;

σ_h^2 = Variance of the claim severity distribution;

λ_h = Expected claim count; and

$\lambda_h + c_h \cdot \lambda_h^2$ = Variance of the claim count distribution.

Following Heckman and Meyers [1983], we call c_h the contagion parameter. If the claim count distribution is:

Poisson, then $c_h = 0$;

negative binomial, then $c_h > 0$; and

binomial with n trials, then $c_h = -1/n$.

A good way to view the collective risk model is by a Monte-Carlo simulation.

Simulation Algorithm #1

The Collective Risk Model Assuming Independence Between Lines of Insurance

1. For lines of insurance 1 to n, select a random number of claims, K_h, for each line of insurance h.
2. For each line of insurance h, select random claim amounts Z_{hk}, for k = 1, ... K_h. Each Z_{hk} has a common distribution $\{Z_h\}$.
3. Set $X_h = \sum_{k=1}^{K_h} Z_{hk}$.
4. Set $X = \sum_{h=1}^{n} X_h$;

The collective risk model describes the distribution of X.

Meyers [1999] shows that if K_h is independent of K_d for $d \neq h$, and Z_h is independent of K_h we have:

$$\text{Var}[X_h] = \lambda_h \cdot \sigma_h^2 + \mu_h^2 \cdot (\lambda_h + c_h \cdot \lambda_h^2); \tag{2.1}$$

and

$$\text{Cov}[X_d, X_h] = 0 \text{ for } d \neq h. \tag{2.2}$$

We now introduce parameter uncertainty that affects the claim count distribution that affects several lines of insurance simultaneously. We partition the lines of insurance into covariance groups $\{G_i\}$. Our next version of the collective risk model is defined as follows.

Simulation Algorithm #2

The Collective Risk Model with Parameter Uncertainty

in the Claim Count Distributions

1. For each covariance group i, select $\alpha_i > 0$ from a distribution with:

$$E[\alpha_i] = 1 \text{ and } Var[\alpha_i] = g_i.$$

   g_i is called the covariance generator for the covariance group i.

2. For line of insurance h in covariance group i, select a random number of claims K_{hi} from a distribution with mean $\alpha_i \cdot \lambda_{hi}$.

3. For each line of insurance h in covariance group i, select random claim amounts Z_{hik} for $k = 1, \ldots K_{hi}$. Each Z_{hik} has a common distribution $\{Z_{hi}\}$.

4. Set $X_{hi} = \sum_{k=1}^{K_{hi}} Z_{hik}$.

5. Set $X_{\bullet i} = \sum_{h \in G_i} X_{hi}$.

6. Set $X = \sum_{i=1}^{n} X_{\bullet i}$.

Meyers [1999] shows that for $d \neq h$:

$$Cov[X_{di}, X_{hi}] = g_i \cdot \lambda_{di} \cdot \mu_{di} \cdot \lambda_{hi} \cdot \mu_{hi}. \tag{2.3}$$

For d = h:

$$Cov[X_{hi}, X_{di}] = Var[X_{hi}] = \lambda_{hi} \cdot \sigma_{hi}^2 + \mu_{hi}^2 \cdot (\lambda_{hi} + (1+g_i) \cdot c_{hi} \cdot \lambda_{hi}^2) + g_i \cdot \lambda_{hi}^2 \cdot \mu_{hi}^2 \tag{2.4}$$

And for $i \neq j$:

$$Cov[X_{di}, X_{hj}] = 0. \tag{2.5}$$

The ultimate purpose of this paper is to discuss the estimation of the g_i's from claim count data, so we remove claim severity from the above equations by setting each $\mu_h = 1$ and $\sigma_h^2 = 0$. This gives us:

$$\mathrm{Cov}\left[K_{di}, K_{hi}\right] = g_i \cdot \lambda_{di} \cdot \lambda_{hi}. \tag{2.6}$$

and for d = h:

$$\mathrm{Cov}\left[K_{hi}, K_{hi}\right] = \mathrm{Var}\left[K_{hi}\right] = \lambda_{hi} + \left(c_{hi} + g_i + c_{hi} \cdot g_i\right) \cdot \lambda_{hi}^2. \tag{2.7}$$

and for $i \neq j$:

$$\mathrm{Cov}\left[K_{di}, K_{hj}\right] = 0. \tag{2.8}$$

3. The Impact of the Covariance Generator on Required Capital

The purpose of this paper is to give some estimators of the covariance generator, g. To this end, we give an example on a hypothetical insurer writing four lines of insurance. The insurer expects 1,000 claims in each line, and the contagion parameter for each line is equal to 0.02. The covariance generator is equal to 0.04. The claim severity distributions are given in Meyers [1999]. Tables 3.1 and 3.2 give various summary statistics of the insurer's aggregate loss distribution

Table 3.1
Aggregate Summary Statistics

Aggregate Mean	101,581,230
Aggregate Std. Dev.	23,270,489

Table 3.2
Claim Severity and Claim Count Statistics

Distribution Name	E[Count]	Std[Count]	E[Severity]	Std[Severity]	E[Total Loss]
GL-$1M	1000	248.60	36,966.16	124,853.59	36,966,160
GL-$5M	1000	248.60	40,348.87	160,218.51	40,348,870
AL-$1M	1000	248.60	11,456.65	76,434.03	11,456,650
AL-$5M	1000	248.60	12,809.55	99,730.27	12,809,550

Table 3.3 and 3.4 give the correlations between each of the lines of insurance for the claim counts, and for the total losses.

Table 3.3
Claim Count Correlation Matrix

	GL-$1M	GL-$5M	AL-$1M	AL-$5M
GL-$1M	1.000	0.647	0.647	0.647
GL-$5M	0.647	1.000	0.647	0.647
AL-$1M	0.647	0.647	1.000	0.647
AL-$5M	0.647	0.647	0.647	1.000

Table 3.4
Total Loss Correlation Matrix

	GL-$1M	GL-$5M	AL-$1M	AL-$5M
GL-$1M	1.000	0.531	0.453	0.423
GL-$5M	0.531	1.000	0.440	0.410
AL-$1M	0.453	0.440	1.000	0.351
AL-$5M	0.423	0.410	0.351	1.000

We now consider some capital requirement formulas. Let X be a random variable representing the insurer's aggregate loss. Let:

$$F(x) = \Pr\{X \le x\}$$
$$f(x) = F'(x)$$
$$\sigma = \text{Standard Deviation of } X$$
$$C = \text{Required Insurer Capital}$$

Then the required capital can be defined by one of the following equations

1. Probability of Ruin Formula: $F(C + E[X]) = 1 - \varepsilon$.

2. Expected Policyholder Deficit Formula: $\dfrac{\int_{C+E[X]}^{\infty} (x - C - E[X]) \cdot f(x)dx}{E[X]} = \eta$.

3. Standard Deviation Formula $C = T \cdot \sigma$.

The probability of ruin is a common textbook capital requirement formula in actuarial mathematics. The standard deviation formula is the probability of ruin formula, when applied to a normal approximation of the insurer's aggregate loss distribution. The expected policyholder deficit formula is more recent, and takes into account the amount of insolvency as well as the probability of insolvency.

We calculated the distribution of X using the Heckman/Meyers algorithm [1983] as modified by Meyers [1999]. We then calculated the capital requirements using the above formulas (with $\varepsilon = 0.01$, $\eta = 0.001$ and $T = 2.32$) for the insurer using various values of g. The results are in Tables 3.5 and 3.6.

Table 3.5
The Effect of g on Capital Requirements

g	Standard Deviation	Probability of Ruin	Expected Policyholder Deficit
0.02	42,388,424	43,179,285	46,210,851
0.03	48,535,720	52,492,867	49,606,674
0.04	53,987,534	57,818,856	55,052,911
0.05	58,937,183	62,516,435	59,858,191
0.06	63,502,198	66,763,256	64,205,165

Table 3.6
The Effect of g on Capital Requirements
% Deviations from the Base g = 0.04

g	Standard Deviation	Probability of Ruin	Expected Policyholder Deficit
0.02	-21.5%	-25.3%	-16.1%
0.03	-10.1%	-9.2%	-9.9%
0.04	0.0%	0.0%	0.0%
0.05	9.2%	8.1%	8.7%
0.06	17.6%	15.5%	16.6%

The above tables show that the value of g can have a significant effect on the required surplus.

4. The Likelihood Function for a Multivariate Claim Count Distribution

From this point forward, we shall assume there is only one covariance group and drop the subscripts i and j in Simulation Algorithm #2.

As we estimate the g parameter across different lines in a covariance group, we will be estimating the parameters, λ_h and c_h, of each claim count distribution simultaneously. In effect, we will be estimating the parameters of a multivariate distribution on the random vector $\vec{\mathbf{K}} = \{K_h\}$

At this point, it is helpful to adopt the vector notation $\vec{c} = \{c_h\}$ and $\vec{\lambda} = \{\lambda_h\}$

The negative binomial claim count distribution, conditional on α, will be obtained from the standard negative binomial distribution by multiplying its mean, λ_h, by α.

Following Meyers [1999], we shall use the following form of the negative binomial distribution for the probability of k_h conditional on α.

$$\Pr\{K_h = k_h | \alpha\} = \frac{\Gamma(1/c_h + k_h)}{\Gamma(1/c_h) \cdot \Gamma(k_h + 1)} \cdot \frac{(c_h \alpha \lambda_h)^{k_h}}{(1 + c_h \alpha \lambda_h)^{1/c_h + k_h}} \tag{4.1}$$

Given $g \geq 0$, define[1]:

$$\alpha_1 = 1 - \sqrt{3g},\ \alpha_2 = 1, \text{ and } \alpha_3 = 1 + \sqrt{3g},$$

and (4.2)

$$\Pr\{\alpha = \alpha_1\} = 1/6,\ \Pr\{\alpha = \alpha_2\} = 2/3, \text{ and } \Pr\{\alpha = \alpha_3\} = 1/6.$$

One can easily verify that $E[\alpha] = 1$ and $Var[\alpha] = g$.

The conditional likelihood of a claim count vector $\vec{\mathbf{k}}|\alpha = \{k_h|\alpha\}$ is given by:

$$\ell(\vec{\mathbf{k}}; \vec{\lambda}, \vec{c} | \alpha) = \prod_h \Pr(K_h = k_h | \alpha). \tag{4.3}$$

---

[1] As pointed out in Meyers [1999], this discrete distribution for α was motivated by the Gauss-Hermite numerical integration formula. One can easily derive similar distributions with more points.

The unconditional likelihood of a claim count vector $\vec{\mathbf{k}} = \{k_h\}$ is given by:

$$f(\vec{\mathbf{k}};\vec{\lambda},\vec{\mathbf{c}},g) = \frac{f(\vec{\mathbf{k}};\vec{\lambda},\vec{\mathbf{c}}|\alpha_1)}{6} + \frac{2 \cdot f(\vec{\mathbf{k}};\vec{\lambda},\vec{\mathbf{c}}|\alpha_2)}{3} + \frac{f(\vec{\mathbf{k}};\vec{\lambda},\vec{\mathbf{c}}|\alpha_3)}{6}. \tag{4.4}$$

As we go about the computational efforts described below, we will work with the log-likelihood functions:

$$L(\vec{\mathbf{k}};\vec{\lambda},\vec{\mathbf{c}}|\alpha) = \ln(f(\vec{\mathbf{k}};\vec{\lambda},\vec{\mathbf{c}}|\alpha)); \text{ and} \tag{4.5}$$

$$L(\vec{\mathbf{k}};\vec{\lambda},\vec{\mathbf{c}},g) = \ln\left(\frac{e^{L(\vec{\mathbf{k}};\vec{\lambda},\vec{\mathbf{c}}|\alpha_1)}}{6} + \frac{2 \cdot e^{L(\vec{\mathbf{k}};\vec{\lambda},\vec{\mathbf{c}}|\alpha_2)}}{3} + \frac{e^{L(\vec{\mathbf{k}};\vec{\lambda},\vec{\mathbf{c}}|\alpha_3)}}{6}\right). \tag{4.6}$$

5. Maximum Likelihood Estimation

Under the assumption that claims are generated by the process described in Simulation Algorithm #2, an insurer wishing to estimate the parameters $\vec{\lambda}$, $\vec{\mathbf{c}}$ and g might gather data like that in the following table from its own claims experience.

Table 5.1
Insurer Data for Estimating c and g

	Exposure by Line and Year			
Year	Line 1	Line 2	Line 3	Line 4
1998	100	80	40	20
1997	100	80	40	20
1996	100	80	40	20
1995	100	80	40	20
1994	100	80	40	20
	Claim Count by Line and Year			
	Line 1	Line 2	Line 3	Line 4
1998	153	131	53	31
1997	96	77	41	20
1996	53	89	45	16
1995	92	72	45	30
1994	92	90	43	16
Estimated Frequency	0.9720	1.1475	1.1350	1.1300

We estimated the insurer's frequency by line of insurance by dividing the total claim count by the total exposure. We then assumed that $c_h \equiv c$ for all h.

Let $\vec{\mathbf{k}}_y$ and $\vec{\lambda}_y$ be respectively, an observed claim count vector and an estimated expected claim count vector for the year y.

In Table 5.1 the observed claim count vector, $\vec{\mathbf{k}}_{1998}$, is equal to $(153, 131, 53, 31)^T$. The expected claim count vector, $\vec{\lambda}_{1998}$, is equal to $(100 \cdot 0.9720, 80 \cdot 1.1475, 40 \cdot 1.1350, 20 \cdot 1.1300)^T$ which is equal to $(97.2, 91.8, 45.4, 22.6)^T$. The parameter vector, $\vec{\mathbf{c}}$, is equal to $(c, c, c, c)^T$. The maximum likelihood estimates $\hat{c}$ and $\hat{g}$ of c and g are the values of c and g that maximizes:

$$\sum_y L\left(\vec{\mathbf{k}}_y \vec{\lambda}_y, \vec{\mathbf{c}}, g\right). \tag{5.1}$$

Using Excel Solver™, we found the maximum likelihood estimate (MLE), $\hat{c}$, of c to be 0.0169 and the maximum likelihood estimate $\hat{g}$ of g to be 0.0245.

We should note that the data in Table 5.1 was not generated from actual insurer data. It was taken from five random drawings from Simulation Algorithm #2 with the "true" frequencies set equal to 1.0000 for each line of insurance, the "true" value of c set equal to 0.0200, and the "true" value of g set equal to 0.0400. We repeated the simulation 100 times with the following results.

Table 5.2
Properties of MLE's for c and g
Derived from 100 Simulations
of a Single Insurer's Data

	c	g
True Value	0.0200	0.0400
Average MLE	0.0134	0.0226
Std. Dev. of the MLE	0.0126	0.0208

One can see from Tables 3.5 and 3.6 that the estimation errors can lead to a significant understating of the required surplus.

Based on this and other similar simulations we conclude that estimating c and g in this manner can lead to biased and highly volatile results.

We now examine some other estimation methods.

The first alternative is to combine the data of several "similar" insurers. Let A be the set of insurers and let $a \in A$. We created 40 nearly identical "copies" of our insurer and simulated the MLE's for c and g. Table 5.3 below shows the exposures and claim counts for the first two insurers in a typical simulation.

When combining the data of several insurers we maximize the log-likelihood expression:

$$\sum_{a,y} L\left(\vec{k}_y^a; \vec{\lambda}_y^a, \vec{c}, g\right). \qquad (5.2)$$

Table 5.3
Multi-Insurer Data for Estimating c and g

Insurer #1

Year	Line 1	Line 2	Line 3	Line 4
	Exposure by Line and Year			
1998	100	80	40	20
1997	100	80	40	20
1996	100	80	40	20
1995	100	80	40	20
1994	100	80	40	20
	Claim Count by Line and Year			
	Line 1	Line 2	Line 3	Line 4
1998	69	69	53	20
1997	99	80	51	17
1996	101	78	68	18
1995	129	94	42	17
1994	82	76	30	15

Insurer #2

Year	Line 1	Line 2	Line 3	Line 4
	Exposure by Line and Year			
1998	20	100	80	40
1997	20	100	80	40
1996	20	100	80	40
1995	20	100	80	40
1994	20	100	80	40
	Claim Count by Line and Year			
	Line 1	Line 2	Line 3	Line 4
1998	25	108	64	45
1997	18	88	75	42
1996	22	87	94	44
1995	22	130	69	47
1994	30	147	111	68

Insurer #3

Year	Line 1	Line 2	Line 3	Line 4
	Exposure by Line and Year[2]			
⇓	⇓	⇓	⇓	⇓
Estimated Frequency	1.0088	1.0077	1.0088	0.9877

We ran 100 simulations of data like that in Table 5.3 and calculated the maximum likelihood estimators for c and g with the following results.

Table 5.4
Properties of MLE's for c and g
Derived from 100 Simulations
of 40 Insurers' Data

	c	g
True Value	0.0200	0.0400
Average MLE	0.0199	0.0399
Std. Dev. of the MLE	0.0022	0.0030

Based on this and other similar simulations we conclude that we can obtain accurate estimates of c and g — if we can get the combined results of several "similar" insurers[3].

The existence (or non-existence) of similar insurers opens up a host of issues. We now explore a few of these issues.

6. Bayesian Estimation

We suspect few insurers would agree that they are sufficiently "similar" to any other group of insurers to fully accept the results of an analysis like that given above. They might accept the results because they have no quantitative alternative, and then judgmentally modify the results. Since we consider it likely that judgment will enter the picture, we consider a Bayesian approach to the problem.

Consider a grid (c_i,g_j) of possible values of c and g. Let $\{\vec{k}_y\}$ be a set of observations needed to calculate the likelihood function for each point (c_i,g_j). Let p_{ij} be the "prior" probability of each point (c_i,g_j).

---

[2] We varied the exposure for the lines in the pattern: 100,80,40,20; 20,100,80,40; 40,20,100,80; and 80,40,20,100.

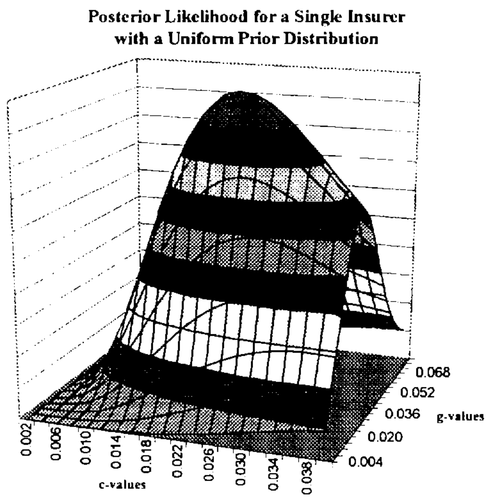
[3] The reader may observe that the expected claim counts for the insurer in this simulated sample were significantly smaller than the insurer discussed in Section 3 above. We also did a simulation where the insurers were 10 times as large. We obtained Std Dev$[\hat{c}]$ = 0.0011 and Std Dev$[\hat{g}]$ = 0.0022.

Then according to Bayes' Theorem, the posterior likelihood of each (c_i,g_j) will be proportional to[4]

$$\prod_{y} \ell\left(\vec{\mathbf{k}}_y;\vec{\lambda}_y,c_i,g_j\right)\cdot p_{ij}. \tag{6.1}$$

As an illustration, suppose that we choose a prior so that the p_{ij}'s are equally likely. For one simulated $\left\{\vec{\mathbf{k}}_y\right\}$ based on a single insurer's exposure we obtained the following posterior distribution of (c_i,g_j), which we show (part of) graphically.

Graph 6.1



[4] For the time being we are assuming that the expected claim count is known. We will address this problem below

As an example, we construct a prior distribution so that

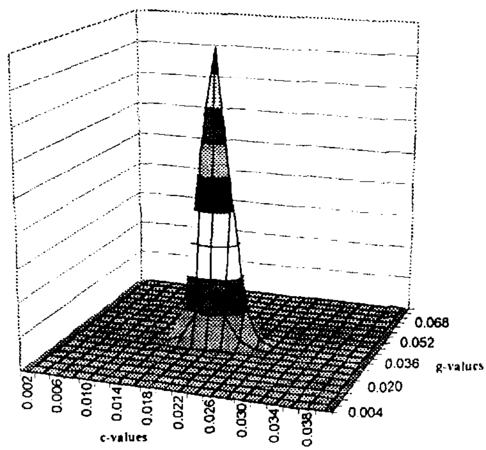
$$p_{ij} \propto \prod_{a,y} \ell\left(\vec{\mathbf{k}}_y^a; \vec{\lambda}, c_i, g_j\right), \tag{6.2}$$

where $\left\{\vec{\mathbf{k}}_y^a\right\}$ comes from the (simulated here, but in practice real) data of the 40 "peer group" insurers given above. We obtained the following posterior distribution for the same insurer that we show graphically.

Below, we will show how to use the posterior distribution as input into the collective risk model, as described in Simulation Algorithm #2.

Graph 6.2

Posterior Likelihood for a Single Insurer with a Prior Distribution Based on Industry Data



7. Industry Drivers of Correlation

The likelihood Equation 3.6 was derived under the assumption that the "driver" of the correlation, i.e. the random variable α, was independent for each individual insurer. This section considers the consequences of the random variable α being common to all insurers. To this end, we replace Steps 1 and 2 of Simulation Algorithm #2 with the more complicated process.

Simulation Algorithm #3
The Collective Risk Model with Parameter Uncertainty
in the Claim Count Distributions
Driven by Industry and Insurer Parameter Uncertainty

1. For each covariance group i, select $\alpha_i^{\wedge}$ and α_i as follows.

    1.1. Select $\alpha_i^{\wedge}$ from a distribution with $E[\alpha_i^{\wedge}] = 1$ and $\mathrm{Var}[\alpha_i^{\wedge}] = g_i^{\wedge}$. $g_i^{\wedge}$ is called the industry covariance generator for covariance group i.

    1.2. Select α_i from a distribution with $E[\alpha_i] = 1$ and $\mathrm{Var}[\alpha_i] = g_i$. g_i is called the insurer covariance generator for covariance group i.

2. For line of insurance h in covariance group i, select a random number of claims K_{hi} from a distribution with mean $\alpha_i^{\wedge} \cdot \alpha_i \cdot K_{hi}$.

3. For each line of insurance h in covariance group i, select random claim amounts Z_{hik} for k = 1, ... K_{hi}. Each Z_{hik} has a common distribution $\{Z_{hi}\}$.

4. Set $X_{hi} = \sum_{k=1}^{K_{hi}} Z_{hik}$.

5. Set $X_{\bullet i} = \sum_{h \in G_i} X_{hi}$.

6. Set $X = \sum_{i=1}^{n} X_{\bullet i}$.

We now calculate the moments of the aggregate loss distribution described by Simulation Algorithm #3.

$$E\left[\alpha_i^A \cdot \alpha_i\right] = E_{\alpha_i^A}\left[\alpha_i^A \cdot \alpha_i \middle| \alpha_i^A\right] = E_{\alpha_i^A}\left[\alpha_i^A\right] = 1 \tag{7.1}$$

$$\begin{aligned} Var\left[\alpha_i^A \cdot \alpha_i\right] &= E_{\alpha_i^A}\left[Var\left[\alpha_i^A \cdot \alpha_i \middle| \alpha_i^A\right]\right] + Var_{\alpha_i^A}\left[E\left[\alpha_i^A \cdot \alpha_i \middle| \alpha_i^A\right]\right] \\ &= E_{\alpha_i^A}\left[\left(\alpha_i^A\right)^2 Var\left[\alpha_i\right]\right] + Var_{\alpha_i^A}\left[\alpha_i^A\right] \\ &= \left(1 + g_i^A\right) \cdot g_i + g_i^A \\ &= g_i + g_i^A + g_i \cdot g_i^A \end{aligned} \tag{7.2}$$

To calculate the variances and covariances analogous to Simulation Algorithm #2, we simply replace the variance g_i in Equations 2.3, 2.4, 2.6 and 2.7 with the expression $g_i + g_i^A + g_i \cdot g_i^A$.

Let $\vec{\mathbf{k}}_y^A$ be a vector of observed claim counts for the "industry" in year y. An example of such a vector based on Table 5.3 is $\vec{\mathbf{k}}_{1998}^A = (69, 69, 53, 20, 25, 108, 64, 45, \ldots)^T$. Similarly let $\vec{\lambda}_y^A$ be a vector of expected claim counts for the "industry" in year y.

The likelihood function of $\vec{\mathbf{k}}_y^A$ conditional on α^A is given by:

$$\ell\left(\vec{\mathbf{k}}_y^A ; \vec{\lambda}_y^A, \vec{\mathbf{c}}, g \middle| \alpha^A\right) = \prod_a \ell\left(\vec{\mathbf{k}}_y^a ; \alpha^A \vec{\lambda}_y^a, \vec{\mathbf{c}}, g\right). \tag{7.3}$$

The associated log-likelihood function is given by:

$$L\left(\vec{\mathbf{k}}_y^A ; \vec{\lambda}_y^A, \vec{\mathbf{c}}, g \middle| \alpha^A\right) = \sum_a L\left(\vec{\mathbf{k}}_y^a ; \alpha^A \vec{\lambda}_y^a, \vec{\mathbf{c}}, g\right) \tag{7.4}$$

Given $g^A \geq 0$ define

$$\alpha_1^A = 1 - \sqrt{3g^A},\ \alpha_2^A = 1, \text{ and } \alpha_3^A = 1 + \sqrt{3g^A},$$

and (7.5)

$$\Pr\{\alpha^A = \alpha_1^A\} = 1/6,\ \Pr\{\alpha^A = \alpha_2^A\} = 2/3, \text{ and } \Pr\{\alpha^A = \alpha_3^A\} = 1/6.$$

The unconditional log-likelihood function is then given by:

$$L\left(\vec{k}_y^A; \vec{\lambda}_y^A, \vec{c}, g, g^A\right) = \ln\left(\frac{e^{L(\vec{k}_y^A, \vec{\lambda}_y^A, \vec{c}, g|\alpha_1^A)}}{6} + \frac{2 \cdot e^{L(\vec{k}_y^A, \vec{\lambda}_y^A, \vec{c}, g|\alpha_2^A)}}{3} + \frac{e^{L(\vec{k}_y^A, \vec{\lambda}_y^A, \vec{c}, g|\alpha_3^A)}}{6}\right) \tag{7.6}$$

8. Maximum Likelihood Estimation Revisited

Consider the following two situations.

1. $g = r > 0$ and $g^A = 0$.
2. $g = 0$ and $g^A = r > 0$.

From the insurer's point of view, the two situations are identical. Its expected claim counts are multiplied by a random number each year.

But from the point of view of one who is trying to estimate the variance of the random multiplier, the situations are different. In the first situation, a new α is picked for each insurer for each year. In the second situation, α^A is picked *once* each year for all insurers. The estimator should use the log-likelihood function in Equation 4.6. In the second situation the estimator should use the log-likelihood function in Equation 7.6.

We did 100 simulations of our 40 insurers where the claim counts are generated by Simulation Algorithm #3, with c = 0.02, g = 0 and $g^A = 0.04$. We then estimated c and "g" using maximum likelihood on Equation 4.6, with the following results.

Table 8.1
Properties of MLE's for c and g
Derived from 100 Simulations of 40 Insurers' Data
with Industrywide Parameter Uncertainty

	c	g	g^A
True Value	0.0200	0.0000	0.0400
Average MLE	0.0218	0.0249	—
Std. Dev. of the MLE	0.0039	0.0158	—

We next did 100 simulations of our 40 insurers where the claim counts are generated by Simulation Algorithm #3, with c = 0.02, g = 0.01 and g^A = 0.03. We then estimated c, g and g^A using maximum likelihood on the "correct" Equation 7.6, with the following results.

Table 8.2
Properties of MLE's for c, g and g^A
Using Estimated Frequencies
Derived from 100 Simulations of 40 Insurers' Data
with Industrywide Parameter Uncertainty

	c	g	g^A
True Value	0.0200	0.0100	0.0300
Average MLE	0.0201	0.0114	0.0213
Std. Dev. of the MLE	0.0023	0.0026	0.0090

If you used the estimated g and g^A in equation 7.2 instead of the true value of g and g^A, you could significantly understate your capital requirements.

It may occur to one that the reason for this downward bias is due to the fact that we use estimated frequencies, rather than true frequencies. To test this we repeated the simulation using the "true" frequency rather than the estimated frequency and obtained the following results.

Table 8.3
Properties of MLE's for c, g and g^A
Using "True" Frequencies
Derived from 100 Simulations of 40 Insurers' Data
with Industrywide Parameter Uncertainty

	c	g	g^A
True Value	0.0200	0.0100	0.0300
Average MLE	0.0200	0.0104	0.0298
Std. Dev. of the MLE	0.0023	0.0029	0.0033

This simulation indicates that the bias is indeed caused by using estimated frequencies in the MLE. However, in practice the "true" mean is not known.

9. Bayesian Estimation Revisited

Consider a grid $\left(\vec{\lambda}_{iy}^{A}, c_i, g_i, g_i^{A}\right)$ of possible values of $\vec{\lambda}^{A}$, c, g and g^{A}. Let $\left\{\vec{k}_y^{A}\right\}$ be a set of observations needed to calculate the likelihood function for each point $\left(\vec{\lambda}_{iy}^{A}, c_i, g_i, g_i^{A}\right)$. Let p_i be the "prior" probability of each point $\left(\vec{\lambda}_{iy}^{A}, c_i, g_i, g_i^{A}\right)$.

Then according to Bayes' Theorem, the posterior likelihood of each $\left(\vec{\lambda}_{iy}^{A}, c_i, g_i, g_i^{A}\right)$ is proportional to:

$$\prod_{y} \ell\left(\vec{k}_y^{A}; \vec{\lambda}_{iy}^{A}, c_i, g_i, g_i^{A}\right) \cdot p_i \qquad (9.1)$$

Let $\vec{e}_y^{A}$ be a vector of exposures for the set of insurers, A, in year y. Let $\vec{f}_i^{A}$ be vector of claim frequencies. Then each coordinate of the expected claim count vector $\vec{\lambda}_{iy}^{A}$ is equal to the product of the corresponding coordinates of $\vec{e}_y^{A}$ and $\vec{f}_i^{A}$. Since the exposures are known and the claim frequencies are unknown, we should put a prior distribution on the grid $\left(\vec{f}_i^{A}, c_i, g_i, g_i^{A}\right)$.

Let $\mathcal{P}_i$ be the posterior probability of each point in the grid $\left(\vec{f}_i^{A}, c_i, g_i, g_i^{A}\right)$. Then one can obtain estimates of $\vec{f}_i^{A}, c_i, g_i$ and g_i^{A} by the following formulas

$$\begin{aligned} \hat{\vec{f}}_i^{A} &= \sum_i \vec{f}_i^{A} \cdot \mathcal{P}_i \\ \hat{c} &= \sum_i c_i \cdot \mathcal{P}_i \\ \hat{g} &= \sum_i g_i \cdot \mathcal{P}_i \\ \hat{g}^{A} &= \sum_i g_i^{A} \cdot \mathcal{P}_i \end{aligned} \qquad (9.2)$$

We then tested the variability of these estimators on our simulated set of 40 insurers. The grid was constructed by varying $\vec{\mathbf{f}}_i^A, c_i, g_i$ and g_i^A in the following manner.

1. Each component of $\vec{\mathbf{f}}_0^A$ was set equal to 0.9875. Each component of $\vec{\mathbf{f}}_4^A$ was set equal to 1.0125. The components for i = 1,2 and 3 were equally spaced in between.

2. c_0 was set equal to 0.0100. c_4 was set equal to 0.0300. The components for i=1, 2 and 3 were equally spaced in between.

3. g_0 was set equal to 0.0020. g_4 was set equal to 0.0180. The components for i=1, 2 and 3 were equally spaced in between.

4. g_0^A was set equal to 0.0200. g_4^A was set equal to 0.0400. The components for i=1, 2 and 3 were equally spaced in between.

In total, the grid had $5^4 = 625$ points. We assumed all points in the grid were equally likely[5].

We made 100 simulated estimates with the following results.

Table 8.4
Properties of Bayesian Estimates for c, g and g^A
Using "True" Frequencies
Derived from 100 Simulations of 40 Insurers' Data
with Industrywide Parameter Uncertainty

	c	g	g^A
True Value	0.0200	0.0100	0.0300
Average Estimate	0.0201	0.0105	0.0303
Std. Dev. of the Estimate	0.0021	0.0020	0.0027

Here we see that one can obtain stable and unbiased (in the classic statistical sense) by an appropriate use of Bayes' Theorem.

---

[5] This "equally likely" is as subjective as any other assumption that one can make. The spacing of the grid is one part of the subjectivity. Another subjective assumption is that the frequencies for the four lines of insurance move together.

9. Using Real Data

This paper has taken a version of the collective risk model, in which the lines of insurance are correlated and explored some methods of estimating parameters of the claim count distributions. The data used in these methods consisted of both exposures and claim counts that span several years.

We explored the use of maximum likelihood on a single insurer's data to estimate the parameters and concluded that the random variation of the estimates were too large to derive a reliable estimate of the insurer's required surplus. One can obtain more stable estimates of the parameters by combining the data of several insurers.

We drew these conclusions from experiments performed on simulated "data."

We now raise some of the issues that one must address when estimating these parameters of the collective risk model with real data from several insurers.

1. Claim Count Development

   When analyzing several years of claim count data, one must take care to distinguish the random variation from the systematic claim count development that occurs because of delays in reporting claims.

2. Insurer Class Differences

   Different insurers can focus on different classes of business. When analyzing the data of several insurers, one must take care to distinguish the random variation from the systematic differences that occur because of the different classes of business that insurers write

3. Insurer Strategy Changes

   When analyzing the data of several insurers, one must take care to note that *planned* changes in insurer strategy that result in changes in claim counts. This can be difficult because insurers usually keep their strategy changes to themselves.

We are in the process of fitting this model to the data of several insurers. We are not yet in a position to say how we are addressing these and other issues. Suffice it to say that we are using our judgment, and we anticipate that the ultimate users of this information will want to impose their own judgment. The Bayesian methodology provides a framework for making these judgments.

In spite of the judgments that one must make, we do feel that parameter estimates using the combined data of several insurers provides a useful starting point for insurers as they go about doing their Dynamic Financial Analysis.

References

1. Bear, Robert A., and Nemlick, Kenneth J., 1990, "Pricing the Impact of Adjustable Features and Loss Sharing Provisions of Reinsurance Treaties" *PCAS LXXVII*.
2. Bowers, N.L.; Gerber, H.U.; Hickman, J.C.; Jones, D.A.; and Nesbitt, C.J., 1997, *Actuarial Mathematics* (Second Edition), Society of Actuaries.
3. Heckman, Philip E., and Meyers, Glenn G., 1983, "The Calculation of Aggregate Loss Distributions from Claim Severity Distributions and Claim Count Distributions" *PCAS LXX*.
4. John, Russell T., and Patrik, Gary S., 1980, "Pricing Excess-of-Loss Casualty Working Cover Reinsurance Treaties" *Casualty Actuarial Society Discussion Paper Program*.
5. Klugman, Stuart A.; Panjer, Harry H. and Willmot, Gordon E., 1998, Loss Models: From Data to Decisions. John Wiley & Sons.
6. Meyers, Glenn G., 1980, "An Analysis of Retrospective Rating" *PCAS LXVII*.
7. Meyers, Glenn G., 1989, "An Analysis of the Capital Structure of an Insurance Company" *PCAS LXXVI*, 1989.
8. Meyers, Glenn G., 1999, "A Discussion of Aggregation of Correlated Risk Portfolios: Models & Algorithms, by Shaun S. Wang," to appear in the *Proceedings of the Casualty Actuarial Society*.
9. Patrik, Gary S., 1996, "Reinsurance", *Foundations of Casualty Actuarial Science, Third Edition*, Ch.6, Casualty Actuarial Society.
10. Wang, Shaun S., 1998, "Aggregation of Correlated Risk Portfolios: Models & Algorithms," to appear in the 1998 *Proceedings of the Casualty Actuarial Society*.

Calibration of Stochastic Scenario Generators for DFA

John M. Mulvey, Ph.D.,
François Morin, FCAS, MAAA, and
Bill Pauling

Calibration of Stochastic Scenario Generators for DFA

John M. Mulvey, Ph.D.
Princeton University
Princeton, NJ 08544
Email: mulvey@princeton.edu

François Morin, FCAS, MAAA
Tillinghast – Towers Perrin
Hartford, CT
Email: Morinf@towers.com

Bill Pauling
Towers Perrin
Philadelphia, PA
Email: paulinb@towers.com

April 1999

Abstract

Leading actuarial companies employ stochastic simulation models to evaluate the viability of pension plans and insurance companies over a set of projected scenarios. A critical element involves generating future scenarios. We show that the problem of calibrating a stochastic scenario system can be posed as a special optimization model, and illustrate the process by means of the Towers Perrin – Tillinghast CAP:Link system. We briefly discuss solution algorithms for the resulting non-convex problem. Areas for future research are indicated.

I. Introduction to Integrated Financial Risk Management

Over the past several years, innovative insurance companies have begun building integrated financial risk management systems. These efforts aim to evaluate the company's activities within a common framework. The goal is to maximize shareholder wealth by focusing on the overall risks and rewards to the organization as measured in several ways. Ideally, the major areas affecting the company's results should be integrated: asset allocation, business management, corporate structure, and re-insurance. Doing so provides the best opportunity to achieve the company's goals over time.

For some insurance businesses, the importance of linking asset and liability risks is well understood. An example is an annuity whose payoff is set at a proportion of the US S&P500 stock return above a set index. It would be foolhardy to invest the assets for this business without understanding the risks of mismatching assets and liabilities. Given recent performance, it is a simple matter to assign assets that exactly match the product's payoff patterns, through options or futures contracts, or funding reserves by dynamically purchasing or selling the stock. Asset and liability management for this type of insurance is a clear and obvious concern. For all insurance companies, there is need to manage the assets and liabilities so that surplus will grow at a rapid pace, as compared with maintaining the surplus at a constant or slowly growing trajectory. In addition, shareholders will seek out insurance companies that grow rapidly and possess diversification benefits.

A dynamic financial analysis (DFA) system consists of three major elements: a stochastic scenario generator, a multi-period simulator, and an optimization module (Figure 1). The first two elements form the corporate simulation system; these are deployed before the optimization module searches for the best compromise decisions given the relevant business, policy, and regulatory constraints. In effect, the optimization runs the simulation by identifying the combination of decisions that best fits the proposed objective function over the multi-period planning horizon.

A critical issue involves constructing the economic scenarios. Each scenario depicts a single coherent path for the primary uncertainties, such as interest rates, inflation, and business activity. Typically, the scenarios are generated by sampling from a system of stochastic differential or difference equations. As a simple example, we could generate short government interest rates by means of a mean reverting equation:

$$dr_t = a(r_\mu - r_t)\,dt + \sigma\, dZ \qquad (1.1)$$

where dZ = Wiener white noise term
r_t = interest rates at time t.

Here, the equation shows that the change in interest rates at time t depend upon three factors – the distance to the target reversion parameter (r_μ), the drift parameter (a), and the instantaneous volatility (σ). Thus, there are three parameters associated with equation (1.1). These parameters dictate the characteristics of the sample paths[1]. The calibration process determines the appropriate values for these parameters. We call the approach – integrated parameter estimation (IPE). The basics are taken up in the next section.

[1] Three sources of errors must be considered in a DFA system: model error, calibration error, and sampling error. We are solely concerned with the second source in this paper. See Mulvey and Madsen (1999) for a further discussion of addressing errors in DFA systems.

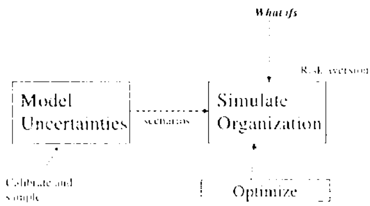


Figure 1
Components of DFA Technology

2. An Optimal Fitting Model for Calibration

This section describes the calibration of the scenario generator as a special optimization model. The primary concept is to match summary statistics and other indicators such as inter-quartile ranges and quantiles as closely as possible, while defining the model parameters as decision variables in the optimization model. The approach directly traces to traditional fitting models, including maximum likelihood, method of moments, and simulated moment estimation. As with these approaches, the model parameters are determined by reference to specialized optimal fitting problems.

Judgement is necessary when determining the parameters of a stochastic model. Fixing parameters is equivalent to setting assumptions. Ideally, we would test the impact of various settings of the parameters on the model's recommendations as shown in Figure 2.

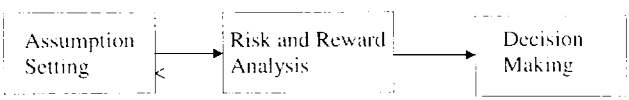


Figure 2: The Three Stages for Conducting a DFA

The process entails combining feedback and revision in order to become comfortable with the recommendations.

2.1 The Estimation Problem

This section reviews the generalized method of moments (GMM) of Hansen (1982) and the simulated moments estimation (SME) of Duffie and Singleton (1993). The notation follows Duffie and Singleton (1993).

Consider a function $H: \Re^N \times \Re^P \times \Theta \rightarrow \Re^N$, with parameter set $\Theta \subset \Re^Q$. Also consider an observation function $f: \Re^{NL} \times \Theta \rightarrow \Re^M$. A scenario process $\{Y_t\}_{t=1}^{\infty}$ is generated by the difference equation

$$Y_{t+1} = H(Y_t, \varepsilon_{t+1}, \beta_0) \tag{2.1}$$

where the parameter vector β_0 is to be estimated, and $\{\varepsilon_t\}$ is an i.i.d. sequence of random variables defined on a given probability space. Let $Z_t = (Y_t, Y_{t-1}, \ldots, Y_{t-L+1})$ defining the state of the process over time. Estimation of β_0 is based on the statistics of the observed process $f_t^* \equiv f(Z_t, \beta)$. For example, we might be interested in the means and standard deviations of the asset returns.

2.1.1 Generalized Method of Moments

Define $\beta \in \Theta$ to be an arbitrary parameter setting. When $f: \beta \mapsto E[f(Z_t, \beta)]$ is analytically known and time independent, the estimation of β_0 can be done with the generalized method of moments (GMM). For these cases, the estimator is:

$$b_t = \arg \min_{\beta \in \Theta} GMM_T(\beta)' W_T(\beta) GMM_T(\beta) \tag{2.2}$$

where $f_t^* \equiv f(Z_t, \beta_0)$, $GMM_T(\beta) = \frac{1}{T}\sum_{t=1}^{T} f_t^* - E[f(Z_t, \beta)]$, $W_T(\beta)$ is a $M \times M$ positive-definite symmetric weighting matrix, and T is the actual number of observations on f_t^*. In words, the vector b_T is the solution[2] to the minimization model defined by the least square GMM function. Hansen (1982) shows that the above minimization produces the estimator with the smallest asymptotic covariance matrix if $W_T(\beta) = \left[E[GMM_T(\beta) GMM_T'(\beta)]\right]^{-1}$.

2.1.2 Simulated Moments Estimator

For wider classes of problems where the GMM assumptions fail, the mapping $f: \beta \mapsto E[f(Z_t, \beta)]$ may be replaced by its simulated version. We assume a sequence $\{\hat{\varepsilon}_t\}$ of

[2] The arg min notation refers to the solution of the posed optimization model.

random variables with a distribution identical to and independent of $\{\varepsilon_t\}$. The simulated state process $\{Y_t^\beta\}$ occurs by choosing a starting point Y_1^β, and letting

$$Y_{t+1}^\beta = H(Y_t^\beta, \hat{\varepsilon}_{t+1}, \beta) \quad t = 1, 2, \ldots \tag{2.3}$$

The simulated observations (summary statistics) are defined as $f_t^\beta = f(Z_t^\beta, \beta)$. If $\{f_t\}$ and $\{f_t^\beta\}$ are governed by the law of large numbers, and identification conditions (Duffie and Singleton (1993)) are met then, $\lim_{T\to\infty} SME_T(\beta) = 0 \Leftrightarrow \beta = \beta_0$.

The SME estimator is,

$$b_T = \arg \min_{\beta\in\Theta} SME_T(\beta)' W_T(\beta) SME_T(\beta) \tag{2.4}$$

where $SME_T(\beta) = \frac{1}{T}\sum_{t=1}^{T} f_t - \frac{1}{\tau(T)}\sum_{t=1}^{\tau(T)} f_t^\beta$, and $\tau(T)$ is the simulation sample size for a given T. The b_T vector is the solution to the equation (2.4). Proofs of consistency (strong and weak) and asymptotic normality are given in Duffie and Singleton (1993). Additional discussions of parameter estimation through simulation can be found in Hansen and Singleton (1982), Pakes and Pollard (1989), McFadden (1989), Gregory and Smith (1990), and Lee and Ingram (1991).

2.2 Integrated Parameter Estimation

The integrated parameter estimation (IPE) approach extends simulated moments estimation in two ways. First, the target vector includes a variety of descriptive statistics besides moments, for example, serial correlation, distribution percentiles, and range estimates. Lee and Ingram (1991) allow for serial correlation in the data set, but they require that the criterion function be continuous in the mean. In contrast IPE does not require a continuous objective function; any general function or even process can be employed. Second, we place bounds on the values of selected parameters. These constraints assist in the search for the best solution.

Given a vector of parameters (β), as decision variables the IPE estimator solves the following optimization model :

$$b_T = \arg \min_{\beta\in\Theta} GSME_T(\beta)' W_T(\beta) GSME_T(\beta) \tag{2.5}$$

$$s.t. \quad L_i \le T_i - S_i \le U_i, \quad i \in \Psi$$

$$\beta_l \le \beta \le \beta_u$$

where $GSME_T(\beta) = \mathbf{m}(T_i - S_i)$, $\mathbf{m}(\cdot)$ is the IPE objective function, S_i denotes the model statistics, T_i denotes the target statistics, β_l and β_u are bounds on the parameter vector β, and Ψ indexes the pertinent statistics. When the objective function equals a distance metric, the set Ψ includes only moments, and the feasible region is unconstrained, IPE is equivalent to SME. If the weighting matrix W_1 is diagonal, then (2.5) reduces to $\arg \min_{\beta\in\Theta} \sum_{i\in\Psi} \hat{W}_i [\mathbf{m}(T_i - S_i)]$, where $\hat{W}_i$ are the diagonal elements.

The IPE approach fits simulated samples from the stochastic model to a given set of descriptive statistics. Each of these descriptive statistics serves as a target, and deviations from the targets are expressed as constraints in (2.5) with tolerances L_i and U_i. The feasible region is determined by a user specified tolerance level, the maximum allowable difference between a given summary statistic and its target. One can also penalize constraints, rather than keeping them explicit. Suitable penalty functions include absolute value and quadratic functions which penalizes underage differently than overage, and asymmetric risk measure. The penalty function relatives must account for differences in scale as well as serving as importance factors. See Mulvey, Rosenbaum, and Shetty (1996) for further discussion, and Berger and Madsen (1999) for a similar approach.

The actual parameter setting process combines the actuary's judgement with the computational ability of the IPE calibration tool. Typically, several iterations are required in order to find the most desired combination of penalties and constraints to meet the target goals (equation 2.5).

3. Dealing with Non-Convexity

The solution of the calibration model is complicated by the presence of non-convexities. At its simplest, non-convexity causes standard hill climbing algorithms to stall at local optimal points. Thereby, software systems such as Microsoft's Excel solver may not find the global optimal solution. Figure 3 shows an example of numerous local optimal solutions.

The search procedure must extend itself in order to cope with non-convexities. To do this, we employ the Tabu search method, one of the most successful methods for overcoming these difficulties. The approach depends upon several memory functions that guide the search and pass through local optimal points as needed. Both long-term and short-term memory are employed.

Tabu search has proven effective for solving combinatorial optimization problems; see Glover 1990 and 1995. The procedure provides for an efficient search of a feasibility region by monitoring key attributes of the points that comprise the search history. Potential search iterates possessing attributes that are undesirable with respect to those already visited become tabu; appropriate penalties discourage the search from visiting them.

Consider a general non-convex optimization problem of the form: minimize$_x$ $f(x)$, $x \in X$. (The function $f(x)$ indicates the responses of a system to a given strategy or decision vector x.) For deterministic problems, there is a single response associated with any x. Our adaptation of tabu search has three basic elements:

1) a function $g(x) = f(x) + d(x) + t(x)$. The function $d(x)$ penalizes x for infeasibility. The function $t(x)$ penalizes x for being labeled tabu.
2) the current iterate x_c, and
3) a neighborhood of the current point N_c. The procedure generates a new iterate x_{new} by selecting the element of N_c for which $g(x)$ is smallest.

The tabu restrictions represented in $t(x)$, can address short-, intermediate-, and long-term components of the search history. Short-term monitoring is designed to prevent the search from returning to recently visited points, allowing the procedure to "climb out of valleys" associated with local minima. Short-term monitoring can also serve as a rudimentary diversification vehicle. Intermediate- and long-term monitoring techniques provide for a much more effective diversification of search over the feasible region. The $t(x)$ function in our version of tabu search relies on exploitation of short-term search history. Details of three processes are required to define our adaptation: formation of the neighborhood of the current point, assignment of tabu penalties, and termination of search procedure. See Glover, Mulvey, and Hoyland (1995).

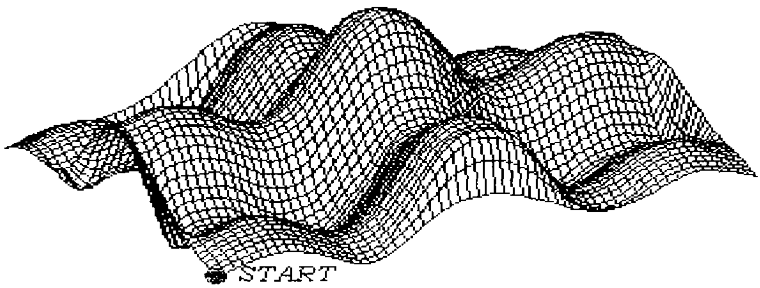


Figure 3
A Non-convex Region with Numerous Local Solutions

4. Calibration Example

The Towers Perrin – Tillinghast company employs the CAP:Link/OPT:Link system for helping pension plan and insurance clients understand the risks and opportunities related to capital market investments. The scenarios generated by CAP:Link contain key economic variables such as price and wage inflation, interest rates for twelve maturities (real and nominal), stock dividend yields and growth rates, and currency exchange rates through each year for a period of up to 20 years. We model returns on asset classes and liability projections consistent with the underlying economic factors, especially interest rates and inflation. The economic variables are simultaneously determined for multiple economies within a common global framework. Long-term asset and liability management is the primary application.

The global CAP:Link system forms a linked network of single country modules. The three major economic powers – the United States, Germany, and Japan -- occupy a central role, with the remaining countries designated as home or other countries. We assume that the other countries are affected by, but do not impact the economies of the three major countries. The basic stochastic differential equations are identical in each country, although the parameters reflect unique characteristics of each particular economy. Additional countries can be readily included in the framework by increasing the number of other countries.

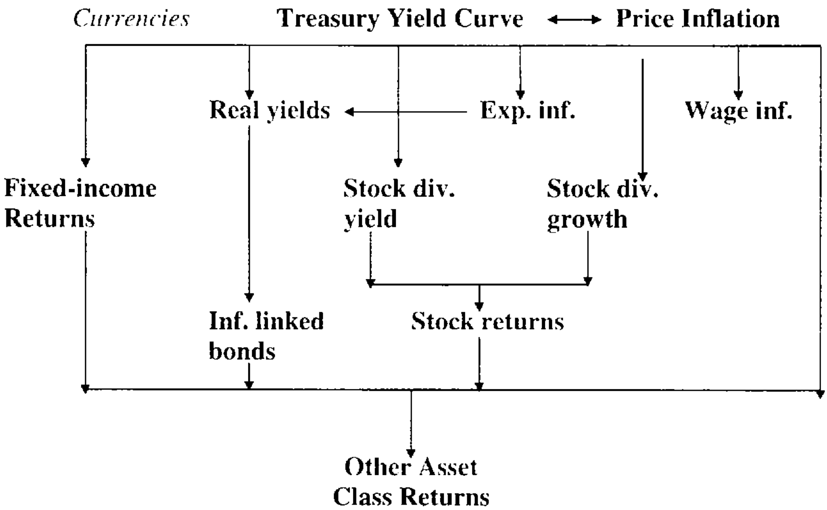


Figure 4: The cascade CAP:Link structure within a single country. Each country in Global CAP:Link depicts a common heritage.

Within each country, the basic economic structure is illustrated in Figure 4. Variables at the top of the structure influence those below, but not vice-versa. This approach eases the task of calibrating the model's parameters. The ordering does not reflect causality between economic variables, but rather captures significant co-movements. Linkages across countries occur at various levels of the model -- for example, interest rates and stock returns. These connections are discussed in Mulvey and Thorlacius (1998). Roughly, the economic conditions in a single country are more or less affected by those of its neighboring countries and by its trading partners. The degree of interaction depends upon the country under review.

The structure is based on a cascade format. Modules above and equal to that module can affect each sub-module within the system. Briefly, the first level consists of short and long interest rates, and price inflation. Interest rates are a key attribute in modeling asset returns and especially in coordinating the linkages between asset returns and liability investments. To calculate a pension plan or an insurance company's surplus, we must be able to discount the projected liability cash flows at a discount rate that is consistent with bond returns, under each scenario. Also, since dynamic relationships are essential in risk analysis, the interest rate model forms a critical element.

The second level entails real yields, currency exchange rates and wage inflation. At the third level, we focus on the components of equity returns: dividend yields and dividend growth.

Returns for the remaining asset classes form the next level, with fixed income assets reflecting the term structure of interest rates and other mechanisms. Each economic variable is projected by means of a stochastic differential equation -- relating the variable through time and with the stochastic elements of the equation and, of course, to other variables and factors at the same or higher levels in the cascade.

A critical feature for a global scenario generator is the currency model. Several complicating issues arise when modeling currency exchange rates. First, currencies must enforce the arbitrage free condition among spot exchange rates and among forward rates with differential interest rates. The second issue involves symmetry and numeraire independence: we must create a structure in which the distribution of currency returns from country A to B has the same distribution as returns from B to A. Both issues limit the form of the currency exchange models, especially when integrating three or more currencies. To avoid these problems, we focus on the strength of each country's currency. Exchange rates follow as the ratio between the strengths of any two countries. The absolute strength of any currency is a notional concept: the relative levels reflect the difference in the exchange rates. See Mulvey and Thorlacius (1998) for further details.

4.1 Example of Calibrating a Scenario Generator with both Assets and Liabilities

We now present an example of calibrating a DFA model that includes both asset and liabilities. In this example, we calibrate the CAP:Link model to produce liability growth, as well as asset returns. We then use the OPT:Link system to find a set of efficient portfolios for a hypothetical property/casualty insurer. These efficient portfolios comprise the asset-liability efficient frontier (ALEF™) for the DFA. The IPE approach forms the basis for the automatic calibration tool.

4.1.1 Form of the Liability Model

For this example, we are interested in modeling a line of insurance that relates to medical and legal inflation. Liability inflation is modeled as a function of its value in the prior period, price inflation, and random volatility. The user inputs consist of an initial rate of inflation, and an assumption of future inflation. The model has two additional calibration parameters: a parameter that determines the sensitivity to modeled price inflation, and a parameter that determines the amount of random volatility. These two parameters will be calibrated in conjunction with the standard CAP:Link parameters. The Lattice optimization solver carries out the non-convex search.

4.1.2 The Calibration Process

We propose four steps for conducting a calibration exercise as shown below. It is advisable to get actuaries and users involved in the process at an early stage so that everyone understands the issues and is comfortable with the resulting model parameters.

Step 1: Analyze Historical Data

The first step to any calibration process should start with historical data. We analyzed historical data to determine the characteristics of the index. For Medical CPI we took data covering the 1947-1998 period. The historical data on Legal Services CPI is much shorter, covering the period from 1986-1998.

Step 2: Set Targets

From our analysis of historical data we determined the following targets:

	Medical CPI	Legal Services CPI
Standard deviation	1.9 – 2.2%	0.8 – 1.0%
Correlation to CPI	0.6 - .0.7	0.45 – 0.55
Average spread over CPI	0.9	0.7

We express the targets as ranges. These targets depict a blend of historical experience and forward-looking analysis. First, we start with the ranges that are consistent with historical experience. Then we adjust for historical trends. For example, for the last 11 years medical CPI has outpaced CPI by 2.5%. Can we reasonably expect this trend to continue? Over a long-term horizon, we might expect the growth in medical costs to be closer to CPI. This issue must be solved by the model developers so that proper targets can be set.

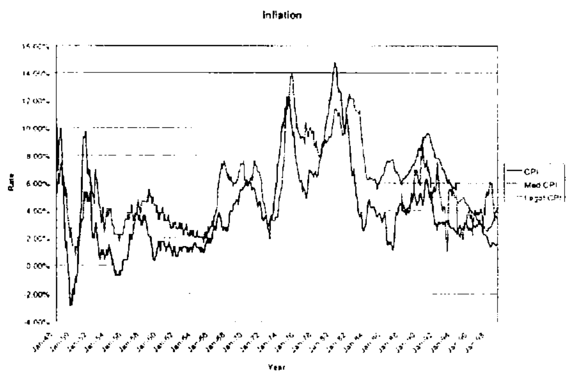


Figure 5
Historical Data for Target Inflation Series

Step 3: Use the Calibration Tool

The calibration system solves the IPE optimization model presented in section 2.2 (equation 2.5). We set up the calibration tool to run 100 scenarios per iteration. We have found through experience that 100 scenarios is a small enough number to run quickly, yet 100 scenarios produce a large enough sample to be representative of the 500 or 1,000 simulations we typically run. We have also discovered through experience that it is best to base calibrations on pure normative conditions. Calibrating to normative conditions removes the effect of trends, as initial conditions move toward their normative states. Depending on the differences between initial and normative conditions, these trends may be significant. If the trends are significant, then they may prevent the calibration tool from being able to meet the targets. Since the targets are set by essentially "normalizing" history, it is best to base calibrations on a normalized environment.

Step 4: Review Model Output

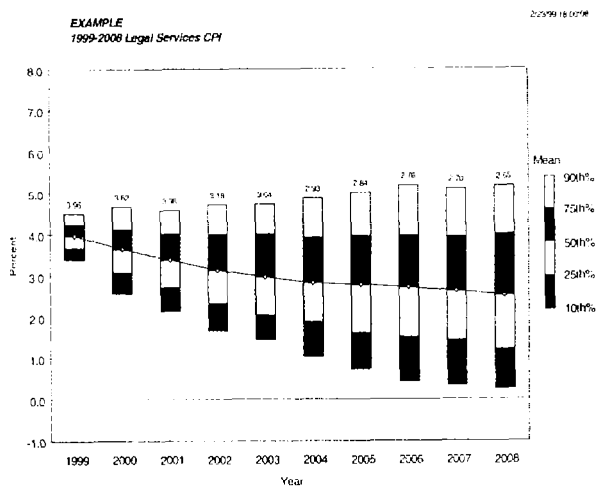
The final step is to take the optimum set of calibration parameters and use them to generate a 500-scenario CAP:Link projection. In this projection we have started with initial conditions so that we can evaluate the effect of the trends. Now we are able to fully evaluate the effect of initial conditions on the optimized parameters. These results must be fully reviewed by an experienced asset simulation expert to determine the reasonableness of the results. In the end, any calibration is only as good as the credibility of the results.

4.2 Linking Assets and Liabilities in DFA Simulations

Next, we consider a hypothetical insurance line of automobile policies. We assume that these liabilities are driven by an equally weighted combination of medical inflation and legal services inflation. Using a starting liability value of $80 million, combined with the stochastic liability growth rates, we can project future liabilities. The initial asset value of $100 million can likewise be combined with the stochastic asset growth rates to project future asset values. For our analysis, we focus on the difference between the assets and liabilities – dollar surplus. The simulation renders investment and business decisions each month over the 10 year horizon.

4.2.1 Generating the Asset/Liability Efficient Frontier

We can use an asset/liability optimizer to generate an efficient frontier. The efficient frontier tells us the combination of assets that produce portfolios with the highest expected reward for a given level of risk at the end of the multi-period horizon. In this case, we have defined reward to be ending dollar surplus and risk to be the standard deviation of dollar surplus. To generate the surplus efficient frontier requires a proper multi-period DFA system. These results show the benefits of calibrating the assets and liabilities to a common set of economic factors.



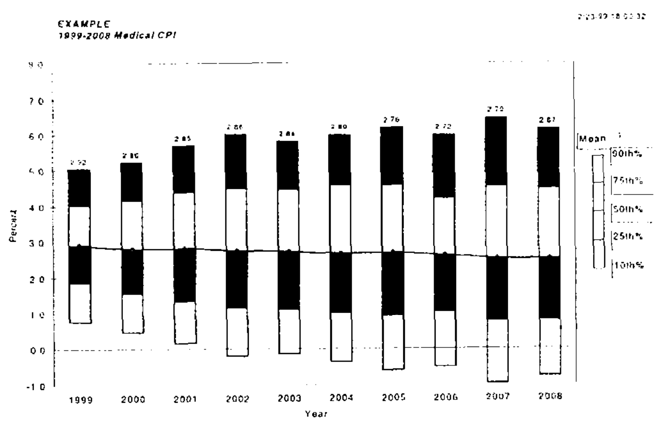
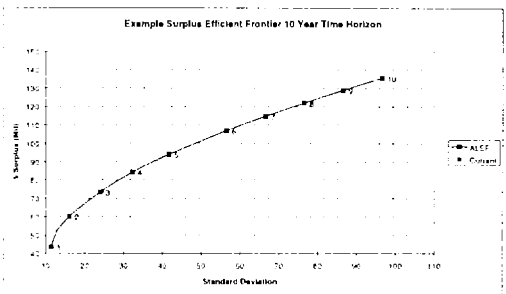


Figure 6
Projected Distributions of Medical and Legal Inflation over the Planning Horizon

4.2.2 Asset Classes and Constraints

For our analysis we have included the following asset classes and constraints. The DFA model resets the asset proportions to these value at each time step. Rebalancing the portfolio is conducted by following a fixed mix decision rule.

Asset class	Min %	Max %	Current Portfolio %
Cash	0	100	5
US Large Cap Equity	0	100	20
Bond Index	0	100	75



Asset Mix %	1	2	3	4	5	6	7	8	9	10	Current
Cash U.S.A.	74.0	38.4	12.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.0
Eqty U.S.A.	1.9	13.3	22.1	32.6	45.8	62.9	73.2	82.7	91.6	100.0	20.0
Bond Index	24.1	48.3	65.4	67.4	54.2	37.1	26.8	17.3	8.4	0.0	75.0
Reward	43.5	60.2	72.2	84.2	91.9	106.8	114.7	122.5	128.9	135.4	73.7
Risk	11.4	16.1	24.0	32.3	41.7	56.5	66.6	76.5	86.0	90.6	24.5

Asset/Liability Efficient Frontier (ALEF^{sm})
Figure 7

5.0 Concluding Remarks

This paper described a systematic method for calibrating a stochastic scenario generator for DFA based on an optimal fitting problem over a set of sampled scenarios. As shown by the example, the resulting calibration tool can be solved by Tabu search and related meta-heuristic approaches. Assets and liabilities should be calibrated together since there are underlying driving factors that affect the company's surplus. To properly calculate risk, we must consider both sides of the balance sheet within a DFA system. The integrated parameter estimation provides a practical method for solving this problem.

Two lines of research merit attention. The first requires the development of better ways to address the non-convex optimization model. We are currently investigating an adaptive algorithm that takes into consideration sampling errors. The goal is to solve the optimization model with the greatest degree of confidence and the least amount of sampling error. The second avenue for research is to extend the procedure to the selection of the forecasting model structure itself. Certain well-defined structural changes could possibly improve greatly our ability to generate scenarios exhibiting the desired behavior. Here, we are taking up the difficult issue of model structure error.

Notwithstanding these issues, we have shown that employing an optimization model for calibration is a practical procedure. We have illustrated the approach on a forecasting model for financial planning -- CAP:Link. We believe the approach holds promise for forecasting systems in other planning domains.

References

Berger, A. and C. Madsen (1999), "A Comprehensive System for Selecting and Evaluating DFA Model Parameters," CAS DFA Forum.

Bradford, David F. (1998) (ed.), **The Economics of Property-Casualty Insurance**, University of Chicago Press.

Broze, L., Scaillet, O. and Zakoian, J-M. (1995) "Testing for Continuous-Time Models of the Short-Term Interest Rate," *Journal of Empirical Finance*, **2**, 199-223.

Campbell, J., A. Lo, A. C. MacKinlay (1997) **The Econometrics of Financial Markets**, Princeton University Press.

Cariño, D.R. and W. T. Ziemba (1998) "Formulation of the Russell-Yasuda Kasai Financial Planning Model," *Operations Research*, July-August.

Correnti, S., S. Sonlin, and D. Isaac (1998) "Applying a DFA Model to Improve Strategic Business Decisions," *Casualty Actuarial Society Forum*.

Duffie, D. and Singleton, K. (1993) "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, **61**, 929-952.

Froot, K. and J. Stein (1998) "Risk Management, Capital Budgeting, and Capital Structure Policy for Financial Institutions: An Integrated Approach," *Journal of Financial Economics*, **47**, 55-82.

Glover, F. (1990) "Tabu Search: A Tutorial," *Interfaces*, **20**, 74-94.

Glover, F. (1995) "Tabu Search Fundamentals and Uses," Technical Report, University of Colorado, Boulder, Colorado, 80309.

Glover, F., Mulvey J.M., and Hoyland K. (1995) "Solving Dynamic Stochastic Control Problems in Finance Using Tabu Search with Variable Scaling," in *Proceedings of the Meta-Heuristics International Conference MIC-95*, Kluwer Academic Publishers, 429-448.

Gourieroux, C., Monfort, A. and Renault, E. (1993) "Indirect Inference," *Journal of Applied Econometrics*, **8**, S85-S118.

Gregory, A. and Smith, G. (1990) "Calibration as Estimation", *Econometric Reviews*, **9**, 57-89.

Hansen, L. (1982) "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, **50**, 1029-1054.

Hansen, L. and Singleton, K. (1982) "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, **50**, 1269-1286.

Holmer, M. (1998) "Integrated Asset-Liability Management: An Implementation Case Study," in **Worldwide Asset and Liability Modeling**, (eds., W.T. Ziemba and J. M. Mulvey), Cambridge University Press.

Lee, B. and Ingram, B. (1991) "Simulation Estimation of Time Series Models," *Journal of Econometrics*, **47**, 197-205.

Lowe, S.P., and J. Stanard (1996) "An Integrated Dynamic Financial Analysis and Decision Support System for a Property Catastrophe Reinsurer," *Casualty Actuarial Society Forum*.

McFadden, D. (1989) "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration," *Econometrica*, **57**, 995-1026.

Mulvey, J. (1996) "Generating Scenarios for the Towers Perrin Investment Systems," *Interfaces*, March 1996.

Mulvey, J., R. Rush, J. Mitchell, and T. Willemain (1997) "Stratified Filtered Sampling in Stochastic Optimization," Princeton University report, SOR-97-7.

Mulvey, J., D. Rosenbaum, and B. Shetty (1996) "Parameter Estimation in Stochastic Scenario Generation Systems," Princeton University Report, SOR-96-15.

Mulvey, J. and A. Thorlacius (1998) "The Towers Perrin Global Capital Market Scenario Generation System: CAP:Link," in **World Wide Asset and Liability Modeling** by W. T. Ziemba and J. M. Mulvey (eds.), Cambridge University Press.

Mulvey, J.M., S. Correnti, and J. Lummis (1997) "Total Integrated Risk Management: Insurance Elements," Princeton University Report SOR-97-2.

Mulvey, J.M. and C.K. Madsen (1999) "Strategic Decision Making for Pension Plans Under Stress," Princeton University SOR Report.

Mulvey, J.M. and W.T. Ziemba (1998) "Asset and Liability Management Systems for long-term Investors: Discussion of the Issues," in **Worldwide Asset and Liability Modeling**, (eds., W. T. Ziemba and J. M. Mulvey), Cambridge University Press.

Pakes, A. and Pollard, D. (1989) "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, **57**, 1027-1057.

Customizing the Public Access Model Using Publicly Available Data

Robert J. Walling III, ACAS, MAAA,
Thomas E. Hettinger, ACAS, MAAA,
Charles C. Emma, FCAS, MAAA, and
Shawna Ackerman, FCAS, MAAA

Abstract

Dynamic Financial Analysis is an extremely powerful tool for all aspects of the insurance operation. With the constantly increasing amounts of information available to the public, DFA models can be better customized to fit the needs of the end user. This paper will examine several areas in which a publicly available model can be customized to fit a company's specific management structure and risk management priorities. Specific approaches to these customizations will be provided along with possible data sources, reasonableness checks, and potential advantages and disadvantages of each approach. Where possible the paper will use publicly available data in order to provide the reader with available sources for developing DFA applications like this one.

Introduction

This paper will discuss specific areas to consider for customization in a DFA model, alternative approaches to take in performing such a customization, available sources of data to aid in the changing of the parameters, and advantages and disadvantages of the tactics presented. We will provide general commentary on the area of customization and then specific examples using the workers compensation line as an example.

We will discuss four general areas of model parameterization: Interest Rate and Economic Condition Modeling, Premium Modeling, Loss Modeling, and Other Modeling Considerations. First, we will briefly describe the model.

About DynaMo

The model used in this analysis is DynaMo by MRH&T[1]. Dynamo is a publicly available model, which allows DFA users to learn about DFA in a forum which proprietary systems do not allow. It is intended to be a learning tool for the public and to help generate ideas on DFA. It has been developed using Excel to facilitate real-time run times and ease of use. DynaMo is completely open so as to help in the understanding of the intricacies in developing and running a DFA model. This includes the formulas for assets, liabilities, and interest rate models. All parameters are readily accessible and can be easily changed. Since every company is different and some parameters may not be appropriate, it is recommended that the users review these parameters prior to using the model.

The model can be thought of as a combination of interactive asset and underwriting cashflow generators. As new money becomes available, either from investments or premiums, underwriting and tax cashflows are generated and any remaining monies are reinvested. Should the outflows exceed the inflows, assets are sold to cover the difference. These cashflow generators are tied together by the workhorse variable -- the interest rate. Exhibit 1 displays a general schematic of the data flows within the model.

The model contains a number of inputs, including company specific historical data and model parameters. Much of the historical data inputs can be taken directly from the company's year-end actuarial report and Annual Statement. In addition to these inputs, economic and underwriting cycle parameters are required. These parameters, combined with some of the company specific input, are used to stochastically generate the following variables:

[1] DynaMo can be downloaded free at www.mrht.com.

1. Underwriting Frequencies	6. Yield Curve
2. Underwriting Severities	7. Claims Inflation by Line of Business
3. Loss and LAE Payment Patterns	8. Equity Returns
4. Catastrophic Losses	9. Underwriting Cycle Positions
5. Short Term Interest Rates	

These variables are used to quantify the following risk categories to which companies are exposed:

1. Pricing
2. Loss Reserve Development
3. Catastrophe
4. Investment

The model generates cashflows at an exposure level basis to aid in the quantification of the impact of the variables listed above. In particular the loss ratio is not modeled in total but calculated as the result of its components.

Future premiums are generated by the following two step process: 1) adjust the previous periods average rate per exposure to reflect inflation, company rate changes, jurisdictional, and underwriting cycle (competitive) impacts, and 2) multiply the adjusted average rate per exposure by the future exposures. For example, the starting average rate may be $100, the modeled rate change 6%, and estimated exposures of 1,000. This would lead to written premium of $106,000.

A-priori ultimate losses for future years are generated by multiplying the exposures by the stochastically generated frequencies and severities. These frequencies and severities are adjusted to reflect inflation and underwriting cycle impacts. For example, inflation may force the average severity upwards and the underwriting cycle may indicate that the market is softening thus bringing riskier business into the company and higher frequency of loss. By breaking the loss ratio into its pieces, we are able to adjust each of its components to reflect the changing economic and competitive environment. It is particularly useful to model the components of the loss ratio when considering the impact of inflation and unemployment.

Two previous papers by this DFA research team provide additional information about the development and application of DFA models generally and this model specifically. The general approach used in this model, the key risks of U.S. property-liability insurers subject to modeling, the parameters incorporated in the financial aspects of the model and examples of the output are described in D'Arcy, Gorvett, et al. (1997)[2]. An application of an enhanced version of the original model to a multiline, multistate primary insurance company is described in D'Arcy, Gorvett, et al. (1998)[3]. This paper includes a case study examining several of the key features of the model, the process of parameterizing the model and refining the results, and the communication process with a company's management team.

---

[2] D'Arcy, Stephen P., Richard W. Gorvett, Joseph A. Herbers, Thomas E. Hettinger, Steven G. Lehmann, and Michael J. Miller (1997) "Building a Public Access PC-Based DFA Model," Casualty Actuarial Society *Forum*, Summer 1997, Volume 2, pp. 1-40.
[3] D'Arcy, Stephen P., Richard W. Gorvett, Thomas E. Hettinger, Robert J. Walling III (1998) "Using the Public Access DFA Model: A Case Study," Casualty Actuarial Society *Forum*, Summer 1998 Edition, pp. 53-118.

INTEREST RATE AND ECONOMIC CONDITION GENERATION

Before discussing the modeling of the fundamental insurance variables, it is best to review the key economic drivers involved in the model. Particular discussion should be provided about the workhorse variable -- interest rates. The model utilizes generated interest rates to affect other relevant economic variables.

Cox-Ingersoll-Ross Interest Rate Generator

Recognizing that an interest rate model requires definition as to precisely what type of rate will be modeled, we chose short-term treasury rates as the base rate resulting from model generations. In particular, we will model 90-day treasury rates on an annual basis.

As discussed in D'Arcy et al. (1997), Cox-Ingersoll-Ross (CIR) provides a workable process for modeling interest rates. CIR offers a mean-reverting random walk, where interest rates are projected by modeling incremental movements in interest rates. These increments are the sum of mean-ward and purely random generated movements. We provide the formula on Exhibit 2. This process is advantageous in that it balances flexibility, simplicity, and intuitive appeal. CIR, by itself, is merely a parameter driven formula concept; it is not intended to be a completely comprehensive or universally accurate system of projection methodologies. Nonetheless, it appears to suit most DFA modeling purposes quite well.

Appropriate parameterization of interest rates demands that one study historical interest rate data as a method for assuring reasonableness. From links to the CAS DFA Web Site[4], a monthly time series was available as shown in Exhibit 3. Observing a graph of several decades of data, our parameter analysis ultimately focused on T-Bill rates observed since 1983. This choice was made to avoid reliance on the unusual economic conditions prevalent early in the 1980's, combined with the belief that future interest rates may remain relatively low in future years given the recent emergence of a balanced federal budget. The long term mean, b, we ultimately selected for the subject model was 6.0%.

CIR also demands that the user provide a mean reversion parameter, a. This was selected based on our judgment in consideration of the historical movements observed about the long-term mean. We selected .25 as the frequency of reversion parameter, a, indicating that we believe the rate should revert around b approximately every four years.

The random element discussed above is the last parameter to select. The standard deviation of the generated normal variate, s1, represents the volatility parameter of CIR. It is projected by observing the standard deviation of prior annual incremental movements in T-Bill rates. We have selected 1.40 as, s1, the volatility parameter.

How do we assure CIR is providing us with a reasonable interest rate result? We use two techniques to accomplish parameter validation: 1) descriptive statistic analysis, and 2) graphical validation. First, we observe the basic descriptive statistics of the historical data in comparison to the same measurements of the projected interest rates. For example, over the process of 100 CIR trials, the mean of the projected data should approximate b (adjusted to consider the impact of low initial rates), and the standard deviation of incremental movements should also approximate s1. Second, we utilized basic spreadsheet graphing

[4] http://www.stls.frb.org/fred/data/irates/gs3m is "hot-linked" to the CAS's website at http://www.casact.org/research/dfa/appendix1.htm

processes to analyze the graphical behavior of historical rates versus projected rates. This was accomplished by recalculating the random generation process several times and illustrating to our own eyes the graphical reasonableness of the projection range. Exhibit 3 shows a single iteration of this process.

Finally, CIR creates a term structure for longer-term treasury maturities. Due to the relatively long duration of assets and liabilities, we felt this property of the yield curve was a variable we should model directly. Therefore, using a slight departure from CIR's original term structure formula; we separately modeled a stochastic spread variable, p. Defined as the difference between 90-day T-Bills and 30 Years T-Bonds, p is projected by a normal random process, using selected mean and standard deviations based on historical spread observations. To project T-Bill rates at points between 90 day and 30 years we utilized an arctangent curve. This provided the proper first (increasing) and second (concave down) derivatives of a typical yield curve. We found that this form also accommodated an inverted yield curve. A graphical validation similar to the 90-day validation process is shown in Exhibit 4.

Inflation Models

Based on the expectation of a positive correlation between interest rates and general price inflation, we utilized a simple linear modeling process shown on Exhibit 5. The critical parameters to be analyzed, therefore, are the slope, m, and intercept, b, of the line as well as the volatility parameter, s2. CPI data[5] was obtained from the CAS DFA Web Site, and a linear regression was run between the 90-day T-Bill rates and the CPI data. We present the regression results on Exhibit 5. The graphical illustration of the fitted general inflation is shown on Exhibit 6.

General inflation should be distinguished from the inflation components affecting workers compensation premiums and loss. These components include wage inflation and medical inflation. Wage inflation[6] was also retrieved from public sources and was compared via its basic statistical properties to CPI data. Our basic observation was that wage inflation and general inflation rates did not differ materially. As a result we used the general inflation variable as representative of wage inflation rates.

Medical inflation rates, by contrast, have exhibited very unique historical behavior relative to general price inflation. Specifically, medical inflation has historically tended to be higher and more volatile. This is particularly evident for workers compensation medical costs during the early 1990's, which were unprotected from deductibles, limits, or benefit coordination. Workers compensation medical losses over these years often exhibited annual inflation levels in excess of 10%. More recently, however, major legislative reforms, combined with the impact of managed care initiatives, have reduced workers compensation medical inflation to levels lower than the CPI. Observing the graph on Exhibit 7 we can see the illustration of these historical rate movements.

As we did for inflation rates, we matched descriptive statistics between historical and projected data as well as the graphical validation of stochastic projections on Exhibit 7.

---

[5] http://www.stls.frb.org/fred/data/cpi/cpiaucsl is "hot-linked" to the CAS's http://www.casact.org/research/dfa/appendix1.htm

[6] http://146.142.4.24/cgi-bin/surveymost?ee is a data page at the Bureau of Labor Statistics site at http://stats.bls.gov/blshome.htm

Unemployment Rate

Workers compensation loss costs are widely thought to be positively correlated with unemployment rates. Previously written CAS papers[7] have offered and supported that when unemployment (particularly involuntary unemployment) increases the average frequency of claims increases. This is apparently due, for the most part, to the lack of return to work prospects for an injured worker. Therefore, the unemployment rate is an important variable to be considered in a workers compensation DFA model as an indicator of general economic conditions and a specific driver of loss result trends.

One possible approach that can be considered for modeling unemployment rates is to use data from the Bureau of Labor Statistics Web Site[8]. This source provided data specifically from the single state where the subject company in this example writes its workers compensation business. When observing the graph of historical interest and unemployment rates, a correlation is not immediately evident. However, upon deeper analysis, we considered that a lagged effect of interest rates on unemployment rates was possible. We ran linear correlations on historical data using lagged unemployment rates as the dependent (affected) variable and 90 day T-bills as the independent (causal) variable. Specifically we ran correlations against unemployment rates with zero, one, two, and three year lags. The best R-Squared measures occurred using the two and three year lags. We further used the average of two and three year lagged unemployment and found the best fit. Therefore, a two-and-a-half year lag on unemployment rates appeared optimal. The lag concept also offers intuitive appeal in that observed higher interest rates generally lead to poorer economic conditions over a span of several months, which later lead to workforce reductions.

The results of our linear regression are shown on Exhibit 8. A linear slope, intercept, and error term were observed and ultimately selected in the same manner that we used to project medical inflation. To validate these selected parameters we again used the tools of descriptive statistical matching and graphical simulation. An example of the graphical validation can be seen in Exhibit 9.

PREMIUM MODELING

Jurisdictional Risk

We will define jurisdictional risk as the risk associated with judicial, legislative and/or regulatory actions that impact the operations of an insurance company. While it is clear that no DFA model could simulate all possible governmental interventions (nor should an efficient model need to), many states have jurisdictional climates that significantly influence operating results. The element of jurisdictional risk that we have chosen to focus on first in The model is in the area of underwriting. Specifically, jurisdictional risk's influence on underwriting results is modeled in two ways: rate change constraints (capping) and implementation lags.

First, proposed rate changes produced by a combination of prior underwriting results and future growth goals are required to stay within an "allowable range". This capping does not mean that rate level changes outside the reasonable range aren't possible. Rather, changes outside the reasonable range will require additional time and/or expense (additional analysis and filing preparation, consultants' fees, insurance department trips, etc.) for approval. Second, states have regulatory structures that range from allowing relatively rapid implementation of desired rates (e.g. open competition, use & file statutes) to structures that

[7] The reader is referred to Lommele, Jan A. and Sturgis, Robert W. (1977) "An Econometric Model of Workers Compensation," Proceedings of the Casualty Actuarial Society and Butler, Richard J. and Worrall, John D. (1982) "Workers' Compensation: Benefit and Injury Claim Rates in the Seventies," Review of Economics and Statistics for two relevant statistical analyses of this relationship between unemployment and workers compensation loss results.
[8] http://146.142.4.24/cgi-bin/surveymost?r5 is a data page at the Bureau of Labor Statistics site at http://stats.bls.gov/blshome.htm

almost assure a lengthy delay (prior approval statutes with lengthy waiting periods). This implementation lag phenomenon and its impact have been evaluated by a number of sources, including research done by the Virginia Bureau of Insurance in their study on alternate methods of rate regulation[9]. It should also be noted that a certain amount of lag in rate implementation exists purely due to data collection and analysis[10]. Intuitively, the capping and implementation lag factors create a maximum and minimum rate change that can be reasonably implemented and impose a delay on how quickly the capped rate change can be implemented[11].

The reason for customizing the jurisdictional risk parameters of This model is that for a given line of business, a number of factors may substantially increase or decrease the jurisdictional risk for an individual company. These factors include the size (e.g. large market share), target market (e.g. non-standard programs), state of domicile (e.g. domestic companies), and regulatory history (e.g. several previous filings going to hearing) of the company. The parameterization of the jurisdictional risk element of a DFA model should use actual company rate filing experience to the extent that the information is credible. The broadest use of company data would be to analyze historical rate levels filed versus those finally approved and delays in the effective dates of those filings to parameterize the rate caps and lags. However, a company's own filing experience may not have enough filings, particularly enough large increases and decreases, to be fully credible. Furthermore, a state can change its regulatory structure (e.g. a "use and file" state converting to prior approval or a change from an appointed commissioner to an elected one) thereby making a company's filing history less relevant.

As a proxy for meaningful filing history, the public access version of The model has been parameterized to represent a "typical" insurance company's jurisdictional risk based on the "1994 Property-Casualty Regulatory Survey" from Conning & Company. This report surveys insurance company executives for their assessment of each state's regulatory restrictiveness as related to reduced business writings, rate suppression, and freedom to manage personal and commercial lines business. The parameterization of the public access model also considers the type of filing statute that exists in an individual state (use & file, file & use, prior approval, state mandated rates), the type of insurance commissioner (appointed or elected), as well as any state specific requirements (Georgia's rate hearing requirement for filings over +9.9%). Data such as the Conning study, the filing statute, and the type of commissioner can serve as a valuable way to extrapolate a company's experience into new states and/or lines. For example, assume a company writes in State X and is considering expanding into State Y. If State X has a prior approval filing statute and an elected commissioner and State Y has a file and use statute, an appointed commissioner, and a more preferable ranking in the Conning study, a looser set of caps and a shorter jurisdictional lag may be appropriate for State Y.

The key to parameterizing the jurisdiction risk component of the underwriting cycle is the reasonableness check. Regardless of the blend of company data and industry experience that is used to parameterize the impact of jurisdictional risk, two questions need to be answered in the reasonableness assessment: "Do the factors seem reasonable to practitioners?" and "Do the jurisdictional risk parameters change the underwriting results in an intuitive way?" The answer to the first question depends on the skill and judgment of the practitioners. We used a number of actuaries and underwriters with filing experience in all states and a variety of backgrounds (different company sizes and a former regulator) to give our selections a

[9] Competition in the Property and Casualty Insurance Industry: An Evaluation of Alternative Methods of Rate Regulation. Bureau of Insurance, State Corporation Commission, January 1978.

[10] Daykin, C.D.; Pentikäinen, T.; and Pesonen, M., *Practical Risk Theory for Actuaries* (First Edition), 1994, p. 340. The combination of the rate review lag and the jurisdictional lag are described as follows: "Profitability and other relevant factors can only be ascertained after a certain delay and further time is required to implement corrective measures. If tariff bureaus and regulatory approval is involved, the process may take even longer. The total time delay is usually 1.5-2.5 years."

[11] It should be pointed out that the selected rates are capped first and then subjected to the lag. This approximates a realistic situation where the company prepares their filing proposing a capped rate change that is then subjected to jurisdictional lag.

peer review. To assess the impact of jurisdictional risk, we expected underwriting results to be impacted in two ways: 1) more disparity between indicated and implemented rate changes and 2) more variance in simulated loss ratios. Intuitively, if a company's ability to respond to rate inadequacies and redundancies is capped and lagged, loss ratios above a company's permissible loss ratio cannot be reduced completely (in severe circumstances) or immediately. Similarly, loss results better than permissible will not worsen to the permissible level as quickly, due to caps and lags on rate decreases. Exhibit 10 shows an example of what the differences in the implemented rate changes for a sample company might look like with and without jurisdictional risk. This example takes a typical selected rate level (a blend of market demand and indicated rate need) and subjects it to jurisdictional capping and lagging. As can be seen, the capping component limits any possibility for large rate changes and the lag component forces a portion of the rate level change to not be realized until the following calendar year. Exhibit 11 then demonstrates the impact on loss ratios for the next accident year. The model's random number reseed feature allows the user to run simulations with all randomly generated elements identical to a previous set of simulations. This allowed us to test the impact on loss ratios of introducing jurisdictional risk with otherwise identical parameters and simulated values. As you can see, there is both a higher variance in the simulated loss ratios and the mean loss ratio has increased.

Advantages and disadvantages of these methodologies are as follows:

Advantages

1. Adding jurisdictional components allows simulated premiums to more closely model reality
2. Allows the testing of changes in environment including:
   - Rate freezes
   - Changes in regulatory system
3. Increases accuracy of testing state entrance or exit implications
4. Takes advantage of a company's own filing experience to the extent that it is credible

Disadvantages

1. Tough to parameterize in a jurisdiction or line where the company has little or no experience
2. Modeler needs to know historical relationship between company and jurisdiction
3. Commissioners and regulatory systems change in sometimes unexpected ways

Impact of Rate Adequacy on Future Rate Levels

There are a number of ways a model can handle changes in rate adequacy[12]. We will propose five methods that can be used to parameterize the model to handle the issue of rate level adequacy. The first one is the simplest approach. It assumes the company's rates are adequate to begin with and only impacted by inflation. Method 2 assumes the company is only concerned about the competitiveness of its rates. Depending on the market position a supply/demand curve is used to determine the required rate change needed to obtain the desired exposure growth.

Method 3 allows the company to look at actual experience when developing the rate change. This becomes more complex as management intervention may result. The basis for this approach is to build into the model techniques similar to the company's actual rate review process. Past loss, premium, inflation, and investment experience are reviewed to determine the rate adequacy. Loss ratios are developed for the preceding time periods by using the a priori ultimate losses adjusted to reflect inflation as of time period t-1. These losses are then trended to the midpoint of period t using an average of claims inflation over the past three years. Premiums are adjusted to bring them to current level and to reflect inflation. The average loss

[12] Daykin, C.D; Pentikäinen, T.; and Pesonen, M., *Practical Risk Theory for Actuaries* (First Edition), 1994, p. 315-319

ratio adjusted to period t cost levels is compared to the company's permissible loss ratio, with an investment income offset (similar to the NAIC Calendar Year Investment Income Offset Approach[13]) to generate an indicated rate level change[14]. This rate level change would need to be capped based on management rules.

The next two methods are hybrids of preceding ones. Method 4 is a weighting between methods 1 and 2. Method 5 is a combination of 2 and 3. The combinations are heavily dependent upon management's views of how the company would handle each of these situations. The mixing of the different methods is intended to help approximate the reality that a company will not always follow the indicated trends but will go with competitive forces in some cases. At this point an example will be helpful.

Example 1

1. Claims inflation (CI) = +6%
2. Trended and adjusted loss ratio (ALR) = 0.75
3. Permissible loss ratio (PLR) = 0.75
4. Investment income offset (IO) = 0.05
5. Growth objective (G) = 10% exposures
6. Simplified supply/demand curve of RC = Gx+y, where RC is indicated rate change and G is growth objective.
7. Soft Market with x = -0.05 and y = -0.05
8. Assumes 50/50 weights are given in weighting together methods

Method 1	Method 2	Method 3	Method 4	Method 5
RC = CI	RC = Gx+y	RC = 1-ALR/(PLR + IO)	RC = .06(.5) + -.055(.5)	RC = -.0625(.5) + -.055(.5)
RC = 6%	RC = 10(-0.05)-0.05	RC = 1 - .75/.80	RC = 0.25%	RC = -5.9%
	RC = -5.5%	RC = -6.25%		

This same example can also be thought of in a graphical sense. The comparison of the implemented rate change to the actuarially indicated change for each method is shown as Exhibit 12.

Advantages and disadvantages of these methodologies are as follows:

<u>Advantages</u>

1. Allows pricing to be dynamic
2. Reflects inflationary pressures also put on losses
3. Method 1 is simple to implement and understand
4. Method 2 recognizes impact of the market conditions
5. Method 3 is consistent with company's current actuarial process
6. Methods 4 & 5 provide a way to balance these impacts on a more realistic way

<u>Disadvantages</u>

1. Requires management intervention to be built in, which may not always be predictable and which is not consistent within or between companies
2. Method 1 is an over simplification and may not be realistic

---

[13] The model contains all of the necessary information to compute a provision for investment income from insurance operations using the NAIC calendar year investment income offset approach. The advantages and disadvantages of calculating investment income using this approach are beyond the scope of this paper. Other methods of calculating investment income and profit provisions (e.g. Discounted Cash Flows) are also easily computed using the information available in the model. The reader is referred to Robbin, Ira, "The Underwriting Profit Provision", 1992 for a detailed discussion of alternatives in this area.

[14] The approach to calculating indicated rate need is provided as an example. Advantages of this methodology and alternative methods to calculating rate need are beyond the scope of the paper. It should be pointed out however that given the data available in the model, a number of different approaches to indicated rate need could be customized into the underwriting module.

3. The supply/demand curves in Method 2 vary between companies, lines of business, and states.
4. Method 3 requires the user to select an actuarial methodology for adjusting rates, including trend selection, credibility issues and catastrophe loads

Once the method of rate change is chosen, it should be tested for reasonableness. This test of reasonableness should look at the following items over a number of simulations:

a. Inflation
b. Trended and developed loss ratio
c. Permissible loss ratio
d. Investment income offset
e. Rate change allowed by competition (This inherently means the supply demand curves have been checked for reasonableness)
f. Actual modeled change

If item f goes against management intuition given a through e, the weightings should be modified.

Impact of Exposure Trend on Premium Level

One of the fundamental properties of this model is that premiums are simulated based on projected exposures and average rates. This premise creates a need for care to be exercised when estimating projected exposure growth so that real exposure growth and inflationary pressure are both reflected in the exposure growth estimate. Several commonly used exposure bases are inflation sensitive. These include property value (used in homeowners), sales (used in general liability), and payroll (used in workers compensation). We have used wage inflation for this workers compensation application; however, the approaches presented could easily be applied to other inflation sensitive exposure bases.

For workers compensation, wage inflation affects premiums through the payroll exposure base. Wage inflation is projected through the random process described earlier and the effect on payroll is calculated. Normally, this is thought to be a fairly instantaneous relationship. Careful consideration should be given to the impact of unionization involving long-term wage agreements and their potential to delay the impact on payroll inflation. For a recent customization project, it appeared from our analysis of the company's own data that such a lag was not material. Therefore, we chose not to build in a wage inflation lag.

Payroll data was projected using audited payroll estimates in order to avoid the concern of estimating subsequent premiums due to audits.

LOSS MODEL PARAMETERIZATION

Impact of Wage and Medical Inflation

Workers compensation benefits include indemnity and medical payments. Loss adjustment expenses (LAE) will also be modeled as a percentage of the sum of the two benefit components. Indemnity losses are typically a direct function of injured worker wages. Therefore, wage inflation is a natural and direct driver of indemnity inflation through its influence on the average replaced wages under the workers compensation statute. However, in addition to the amount of the payment, the average time duration of disability payments should also be considered in the modeling process. Thus, a duration trend element was also necessary to project indemnity inflation.

To develop an indemnity duration trend parameter, in a recent customization project, we analyzed a company's actual indemnity loss experience relative to actual wage inflation. A fairly constant additive increment of 2.0% over wage inflation appeared evident through most statistical indications. Therefore, the formula for indemnity inflation was set at wage inflation + 2.0%.

Having previously modeled medical inflation, we used a percentage mix of benefits to develop a total loss inflation. Historical data for the subject company and others in its market indicate a fairly steady observation of two-thirds indemnity to one-third medical. By calculating annual loss costs through the projection period we could rebalance these weights. Through this apportionment of benefits, a total loss trend can be modeled which offers an analytical basis of inflation through its components.

Unemployment's Effect on Frequency

As discussed earlier, changes in unemployment rates are thought to have an effect on claim frequencies. For the subject company in a recent customization and other companies writing in its jurisdiction, we have analyzed the historical unemployment time series we used above in comparison to the change in reported claims per unit payroll for these companies. We ran a linear regression on these frequency measures versus unemployment rates and found the relationship to be nearly direct. That is, for each point (1.0%) change in the unemployment rate, the claim frequency changed approximately one point as well. As a result, we utilized a formula that increased the frequency per \$100 payroll, one point for each point the modeled annual unemployment rate changed.

OTHER PARAMETERIZATION ISSUES

Collateralized Mortgage Obligations

The model has the ability to model different types of bonds. Bonds are segregated based upon their class and maturity. The maturity groupings are 1) Less than 1 Year, 2) Over 1 Year through 5 Years, 3) Over 5 Years through 10 Years, 4) Over 10 Years through 20 Years, and 5) Over 20 Years. The model then uses the same underlying methodology to develop the appropriate cashflows. This methodology is as follows:

1. Start with face values and coupon rates
2. Model coupon payments by multiplying the face value by the coupon rates
3. Determine end of year statutory book values using straight line amortization
4. Determine end of year market value according to the following formula:

   $MV = FV \times \Sigma\, CF_t / (1+i)^t$ where CF is the Cash flow ratioed to the face value

5. Mature bonds between maturity buckets assuming uniform distribution. Thus 20% of the market values in the maturity grouping "Over 5 years through 10 years" are assumed to migrate into maturity grouping "Over 1 Year through 5 Years"
6. Coupon rates are adjusted for each maturity group to reflect bonds maturing in and out and the purchase of new bonds

This model can be re-parameterized fairly easily to model collateralized mortgage obligations (CMO's) on a simplified basis. The inclusion of CMO's involves two additional steps. The first step is the modeling of the expected percentage of mortgage prepayments. The prepayment percentage is based upon the Public Securities Association (PSA) model, which assumes that the proportion of mortgages prepaid increases linearly by 0.2% annually for the first thirty months, then levels off at 6% per year thereafter. These assumptions are then indexed to represent greater or lesser prepayment activity due to change in interest rates. For example, if the interest rate were to increase by 100 basis points we would expect a decrease in the prepayment activity. Thus the PSA model would be adjusted down to reflect fewer mortgage prepayments and accordingly fewer prepayments of CMO's. The CMO model can be set up to handle a number of interest rate change ranges. Currently it is set up according to the following:

Interest Rate Change From Starting Point	% of PSA
+1.5%	50%
+1.5% to +0.5%	75%
+0.5% to –0.5%	100%
-0.5% to –1.5%	125%
-1.5%	150%

Once the percentages of prepayments are known, we assume the CMO's are prepaid in the same proportion according to the maturity of the bond. Using the same steps as outlined above we offer three additional steps to include in the process:

2a. Face value redemption would be calculated as the prepayment percentage times the face value. This will generate a cashflow available for claims or reinvestment.
3a. Book values are recalculated assuming a decrease according to the modeled percentages.
4a. Market values are also decreased in proportion to the modeled prepayment percentages.

Checks for reasonableness are best performed using historical result. Past prepayment levels can be compared to interest rate level changes in determining the factor adjustment to the PSA study.

Advantages and disadvantages of these methodologies are as follows:

Advantages

1. Simple to understand
2. Allows the user to test the impact of CMO's on the company's returns and cashflows
3. Models the correlation between change of interest rates and prepayment of CMO's in an understandable manner

Disadvantages

1. Does not take into consideration impacts on different traunch holdings
2. May be an over simplification of the real world

Underwriting Expense Modeling

In DFA and general actuarial literature, underwriting expenses have historically taken a back seat to research on losses (in terms of their impact on rates and reserves) and assets. The reason for this lower priority in the development of DFA research is that underwriting expenses have less variability and therefore have a smaller impact on the mean and variability of future company results. However, as more companies focus on operational efficiency, the need for more sophisticated expense modeling has grown. We will examine two added levels of complexity that some insurers may wish to consider adding to a general DFA model if their company's situation warrants a more detailed parameterization: fixed versus variable expenses and step-wise incremental fixed expenses.

For the purpose of this discussion we will define other underwriting expenses (OUE) as the sum of the other acquisition expense and general expense items. The easiest approach that can be taken for parameterizing and simulating other underwriting expense ratios is to assume a constant percentage of direct written premium will be used for underwriting expenses regardless of increases or decreases in premium level, rate adequacy, or any other operational change. This approach works exceptionally well for commissions and taxes that are almost completely variable with written premium. For companies with stable expense ratios, this fixed percentage approach also provides a reasonable approximation of reality for other underwriting expenses that can be programmed and modeled easily. In fact, the public access version of The model uses this approach for simplicity and the broadest possible applicability. However, companies can be faced with many situations where this approach is not reasonable. For example, a start-up organization whose premiums are growing rapidly may see substantial decreases in their expense ratios as fixed costs (office space, computer systems, etc.) are spread over a larger premium base. Companies going through premium reductions, down-sizings, changes in distribution channels, or acquisitions of other companies or additional blocks of business may also be in situations where the underwriting expense ratio is a moving target rather than a fixed one.

The first parameterization alternative is to recognize some other underwriting expenses as fixed. Any other underwriting expense that remains completely unchanged regardless of premium level can be viewed as fixed. Typical fixed expenses are such items as computers (especially large mainframe computers), rent and other overhead items. A common assumption about fixed expenses is that about half of all current OUE is fixed. This approach is intuitively appealing and is commonly used in the development of expense constants. For a company that feels that their expenses are materially different from this general assumption, an analysis of the "Acquisition, Field Supervision and Collection Expenses" column of Part I of the Insurance Expense Exhibit may be appropriate. We did such an analysis (see Exhibit 13) for a recent client and found the results not substantially different from the 50/50 split.

Another level of sophistication that can be added to projecting other underwriting expenses is the addition of incremental fixed expenses at specific levels of premium growth and needs a larger computer or more space. This modification reflects the realistic situation of additional fixed expenses being incurred as a company experiences significant growth. Situations that might give rise to this situation would include computer upgrades and renting additional office space. It should be noted that several of these items impact assets as well as liabilities and the DFA model needs to be customized on the asset side to reflect these additional non-invested assets. One simple approach to approximating this step-wise fixed expense behavior is to select a premium growth amount at which a fixed expense amount (either a dollar amount of incurred fixed expense or a percentage increase of the other underwriting expense ratio) is incurred. Note that when premium is declining this modeling approach has the effect of making the expense ratio increase until a fixed expense item can be eliminated. This parameterization causes the expense ratio to decrease less rapidly than a simple fixed expense approach and may create a more realistic projection of expense levels in models predicting substantial growth or decline.

Another expense modeling alternative is reflecting expenses that vary by unit cost. Items in this category would include loss control surveys, policy forms and jackets, identification card issuance and loss reporting kits. These items behave like variable expenses but are sensitive to rate adequacy per exposure and changes in average policy size.

A simple reasonableness check for the parameterization of the other underwriting expense generator is a graph comparing the other underwriting expense ratio (to direct written premium) to the change in direct written premium. As you can see in Exhibit 14, an all-variable expense model creates a horizontal line. A partially fixed expense model implies a line with some recognition of economies of scale. A partial fixed expense model with a recognition of additional fixed expenses after sufficient premium growth, decreases in a somewhat jagged fashion and at a slower rate than the partial fixed expense without the step-wise adjustment.

Advantages and disadvantages of these methodologies are as follows:

<u>Advantages</u>

1. Companies focused on operational efficiency as a style will want the split
2. Allows companies to incorporate staffing models into DFA analysis
3. Allows much better forecasts of U/W results under growth scenarios
4. Allows more accurate measurement of the expense component of the new business penalty[15]

<u>Disadvantages</u>

1. Future expense levels and management decisions difficult to parameterize
2. Could be an over-parameterization of the model for the subject company that could distract from more significant risks

Policyholder Dividends

Another expense related issue that may not be directly related to premiums is policyholder dividends. Many workers compensation writers, for example, have a wide variety of policyholder dividend plans that pay either a flat percent of premiums (flat dividend plans) or a percent of premium that varies depending on the insureds size and loss results (variable dividend plans). Neither the variable expense approach used for commissions, nor the fixed expense approach presented for other underwriting expenses works well for dividend plans. There are two reasons for this: 1) the market influences the type and number of dividend plans extended to a company's insureds, and 2) loss results, not premium, dictate how much of a dividend is paid out[16]. Furthermore, dividends are generally paid out six to nine months after policy expiration and so lag behind the earned premium and incurred losses with which they are associated.

The public access version of The model assumes policyholder dividends to be a minimal issue and is initially parameterized with a fixed percentage of premium approach. This accommodated our desire for the public access model to be as widely applicable and straightforward as possible. However, any company with a sufficient amount of written premium subject to dividend plans needs a more sophisticated approach. Two basic issues need to be parameterized in a more sophisticated dividend model: 1) the percentage of the

[15] Traditionally, the new business penalty has been thought of as a quantification of the inferior loss ratio results of new business. There is a similar penalty to the expense ratio for lines of business with substantial fixed costs associated with the first policy (e.g. MVRs, loss control surveys, policy file set up).

[16] It should be noted that the payout from flat dividend plans do not vary with loss results, except to the extent that by law, no dividend disbursement can be guaranteed so even a flat dividend could not be paid if loss results were poor enough. Flat dividend programs are currently used almost exclusively in states where the rate regulatory environment precludes deviation of rates from bureau levels (e.g. Wisconsin and New Jersey); therefore our discussion focuses on the more commonly used variable plans. If a company used predominately flat dividend plans a percentage of premium approach or an approach that varied the dividend according to market position might be more appropriate.

book of business that are offered each kind of dividend plan in a given phase of the market, and 2) the expected payout for each plan given a known loss result.

Exhibit 15 provides an example of how this model could be parameterized in the case of a company with 2 variable dividend plans. The modeler could develop an expected distribution of written premium in each dividend plan at each point in the cycle based on actual company experience and discussions with company personnel concerning their expected behavior. Information estimating dividend payouts at different loss ratios should be available for each plan or can be fairly easily approximated. Once this parameterization is accomplished, future dividend payouts are computed as the weighted average of the expected payouts for the two prior accident years as is shown in Exhibit 15. Net loss ratios can be used to approximate loss capping that occurs in some dividend plans, if retention levels are similar. A straightforward reasonableness check for this customization is a graph comparing loss ratios (net or direct as selected above) from a two year period versus the policyholder dividend ratio (to direct earned premium) paid in the first subsequent year.

This technique of modeling items as a percentage of premium based on loss results and market position has two other significant uses: 1) contingent commissions, and 2) residual market burdens. Contingent commissions are in many respects simply dividends paid to the agent instead of the policyholder. Multiple agency incentive plans with different payouts which can be extended to different numbers of agents depending on market conditions can be parameterized using an approach almost identical to the one shown in Exhibit 15. Residual market burdens can be viewed as a cost of doing business (literally a percentage of earned premium) in certain lines, most notably workers compensation, automobile and property lines in certain states. This cost of doing business varies by market position and jurisdiction. An approach that incorporates some elements of a jurisdictional risk assessment and is designed similarly to the dividend approach provides a reasonable approximation to future residual market loads. NCCI and AIPSO both provide data to member companies by line and state that assists greatly in parameterizing this customization. An example of a straightforward parameterization of residual market burdens is shown as Exhibit 16.

Advantages and disadvantages of these methodologies are as follows:

Advantages

1. Intuitively more reasonable
2. Easy to program
3. Recognizes the impact dividends, contingent commissions, and residual market burdens can have on operating results
4. Recognizes the loss and/or market sensitivity of these items

Disadvantages

1. Difficult to validate some parameters
2. May overcompensate
3. Increases impact underwriting cycle position has on underwriting results

AREAS OF CONTINUED RESEARCH

There are a number of areas of research in the area of model parameterization that the DynaMo research team is continuing to develop. Some of these include the following:

Enterprise-Wide Modeling – How are foundational risk factors that are common to many industries but with sometimes different impacts, like catastrophes, inflation, and interest rates, used to build an enterprise-wide DFA model for an organization that includes property/casualty insurance companies and other entities like banks and life insurance companies? What kinds of metrics are needed? How are the unique risk factors for these other industries parameterized and modeled?

Managed Care Impacts – How are the impacts of managed care penetration and network strength incorporated into estimated frequency and severity for a workers compensation writer? How will managed care impact loss payment patterns? How should network access and network management fees, especially contingent fee structures, be parameterized and modeled?

Securitization – How are the bond modeling and catastrophe modeling capabilities of a DFA model best blended to estimate the price of catastrophe bonds? How can a DFA model be used to test the loss payout risk in an apparent financial reinsurance agreement?

Ratemaking – What is the best approach to using a DFA model to simulate a range of possible indicated rate needs? Can this approach bring something akin to risk margins into ratemaking as an alternative method for computing a profit provision?

Demutualization, Mergers, and Acquisitions – How can a DFA model be customized to assist an insurance company deciding whether to demutualize? How can a company combine their own data with one or more merger or acquisition candidates in a DFA model to assess and potentially rank possible candidates? How can this information be used to estimate dilution value?

ACKNOWLEDGEMENTS

The authors would like to thank a number of people for their assistance in the development of The public access model and this paper. Thanks to Steve D'Arcy and Rick Gorvett, integral members of our DFA research team from the beginning; their assistance is invaluable. Thanks to the other members of the team from MRH&T: Steve Lehman, Mike Miller, Joe Herbers, and Bob Anker for their continued efforts in the areas of DFA research and development. Thanks to Elizabeth Wiesner and Anthony G. Phillips who provided a wealth of expertise and experience in the specifics of customizing the parameterization of a DFA model for a mid-sized regional workers compensation insurer. Thanks to Elizabeth Wiesner, Joe Herbers, and Dave Otto for providing peer reviews. Thanks to Anne Walling for providing proof reading.

Exhibit 1 – Operational Schematic

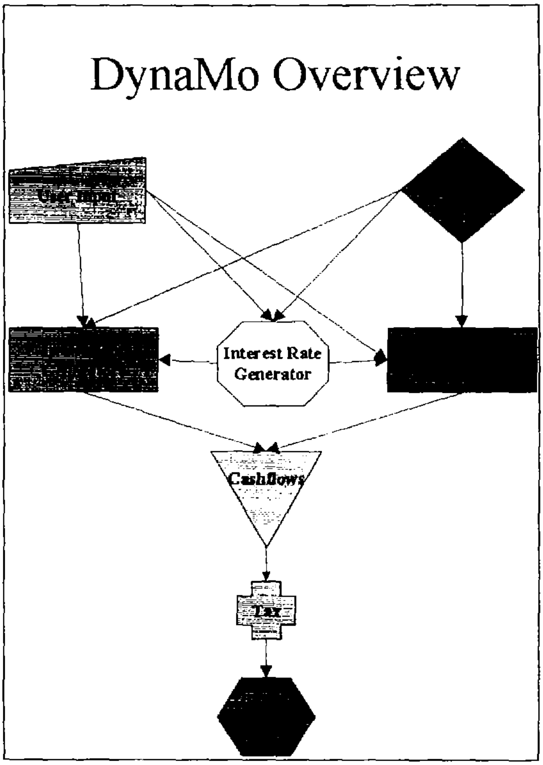


Exhibit 2

Cox Ingersoll Ross

Interest Rate Generator Formula

General Formula: $r_i = a \times (b - r_{i-1}) + s_i \times z_i$

Selected Formula: $r_i = 0.25 \times (0.06 - r_{i-1}) + 1.40 \times z_i$

where

where	r_i	=	90 day rate for year i
	a	=	reversion frequency parameter
	b	=	long-term mean for 90 day rates
	s_i	=	volatility parameter
	z_i	=	standard normal variate

Exhibit 3

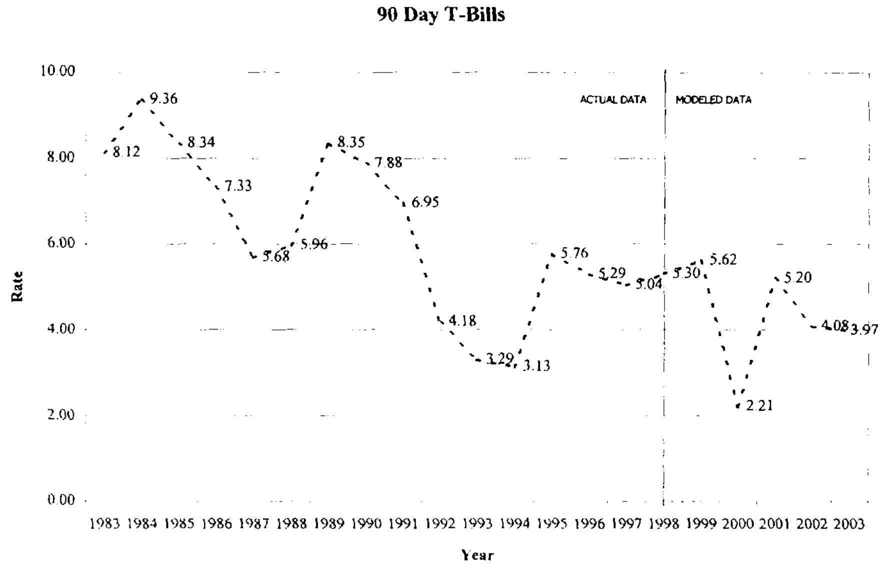


Exhibit 4

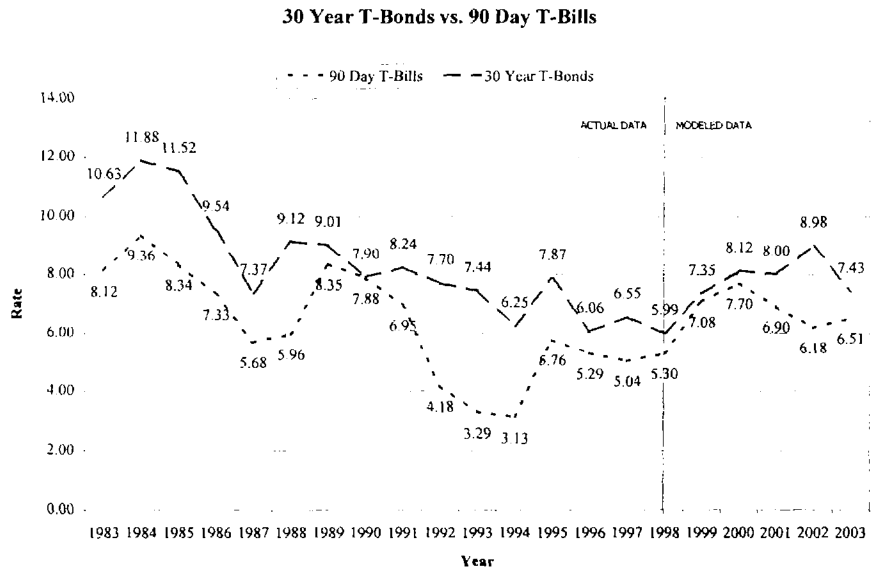


Exhibit 5

$CPI = m\,(interest) + b + s2 \times z2$

Date	Interest	CPI
1984	9.36	3.58
1985	8.34	4.04
1986	7.33	3.79
1987	5.68	1.19
1988	5.96	4.42
1989	8.35	4.41
1990	7.88	4.64
1991	6.95	6.25
1992	4.18	2.98
1993	3.29	2.96
1994	3.13	2.81
1995	5.76	2.60
1996	5.29	2.60
1997	5.04	3.31
1998	5.30	1.70
1999	*6.66*	*3.56*
2000	*4.50*	*2.88*
2001	*4.85*	*3.48*
2002	*7.40*	*3.87*
2003	*7.15*	*3.33*

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.494962134
R Square	0.244987514
Adjusted R Square	0.186909631
Standard Error	1.133478755
Observations	15

	Coefficients	*Standard Error*
Intercept	1.386254038	1.031405168
X Variable 1	0.331762727	0.161532906

Exhibit 6

Inflation vs. 90 Day T-Bills

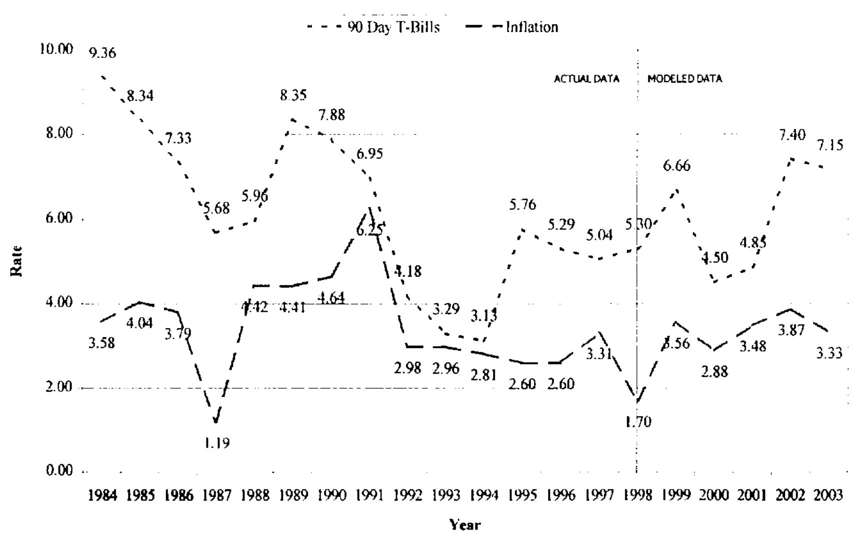
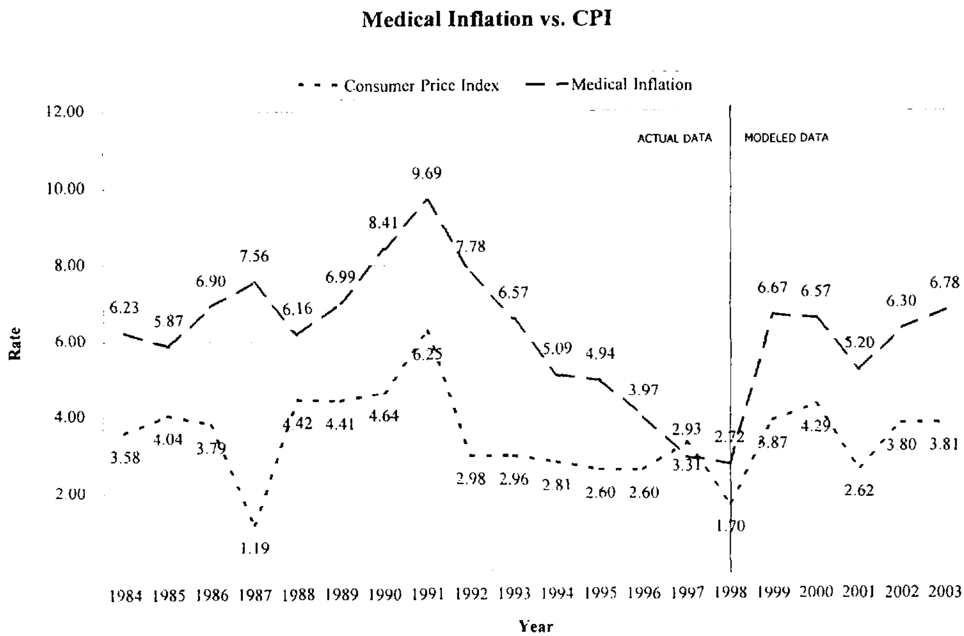


Exhibit 7



Medical Inflation vs. CPI
Consumer Price Index
Medical Inflation
ACTUAL DATA
MODELED DATA
Rate
12.00
10.00
8.00
6.00
4.00
2.00
6.23
5.87
6.90
7.56
6.16
6.99
8.41
9.69
7.78
6.57
5.09
4.94
3.97
2.93
2.72
6.67
6.57
5.20
6.30
6.78
3.58
4.04
3.79
1.19
4.42
4.41
4.64
6.25
2.98
2.96
2.81
2.60
2.60
3.31
1.70
3.87
4.29
2.62
3.80
3.81
1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003
Year

Exhibit 8

CPI = m (interest) + b + s2 x z2

Date	TB 3M	UE+2.5
1983	8.12	
1984	9.36	
1985	8.34	14.00
1986	7.33	11.10
1987	5.68	9.70
1988	5.96	8.65
1989	8.35	8.20
1990	7.88	7.50
1991	6.95	7.50
1992	4.18	8.05
1993	3.29	8.70
1994	3.13	8.45
1995	5.76	7.00
1996	5.29	6.00
1997	5.04	5.20
1998	5.30	4.80
1999	*2.63*	*6.08*
2000	*3.31*	*5.99*
2001	*4.34*	*5.92*
2002	*3.01*	*5.18*
2003	*2.75*	*5.91*

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.385638451
R Square	0.148717015
Adjusted R Square	0.077776766
Standard Error	2.270314616
Observations	14

	Coefficients	*Standard Error*
Intercept	5.062381825	2.252755614
X Variable 1	0.533179613	0.368247289

Exhibit 9

Unemployment vs. 90 Day T-Bills

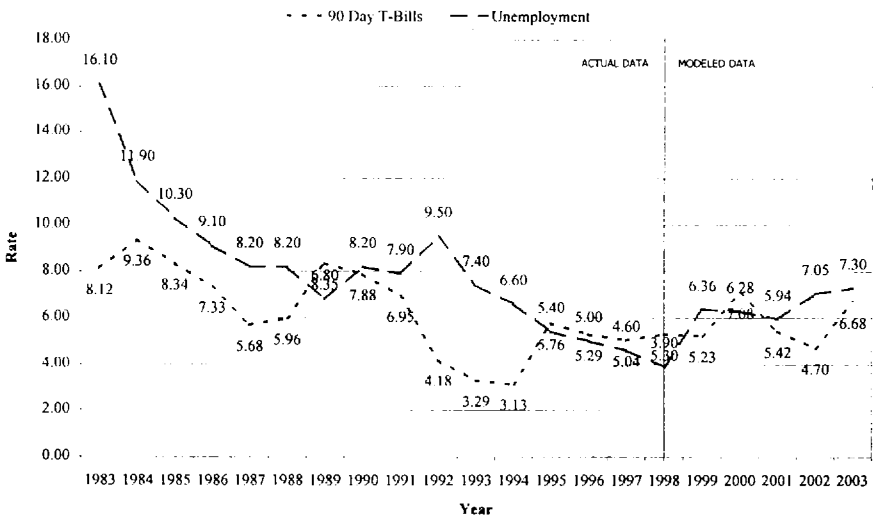


Exhibit 10 – Impact of Jurisdictional Risk on Selected Rate Level

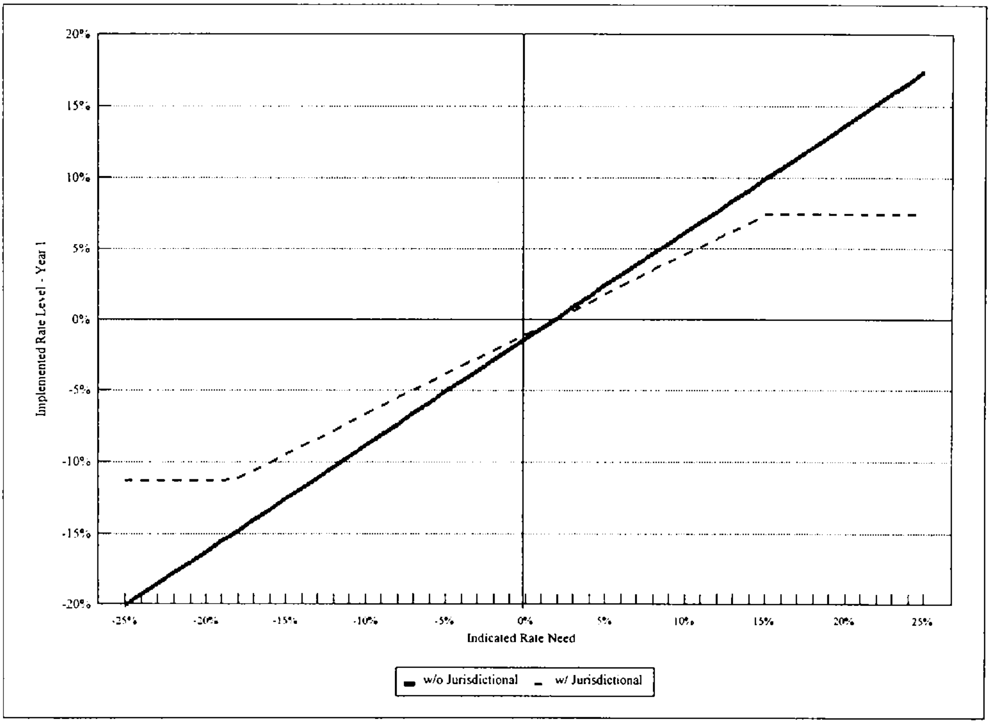


Exhibit 11 – Impact of Jurisdictional Risk on Direct Loss Results

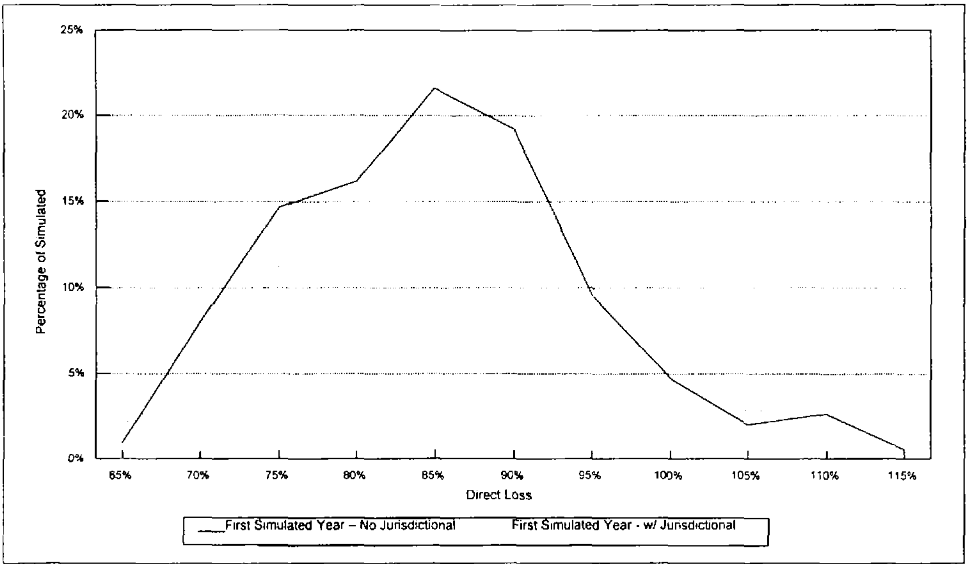


Exhibit 12 – Selected Rate Level Alternatives

Assumptions:
Loss Inflation – 4.0%
Change required for desired premium growth at existing point in cycle. – 5.0%
Method 4 weight assigned to inflation – 50%
Method 4 weight assigned to indicated rate level – 75%
No jurisdictional effects

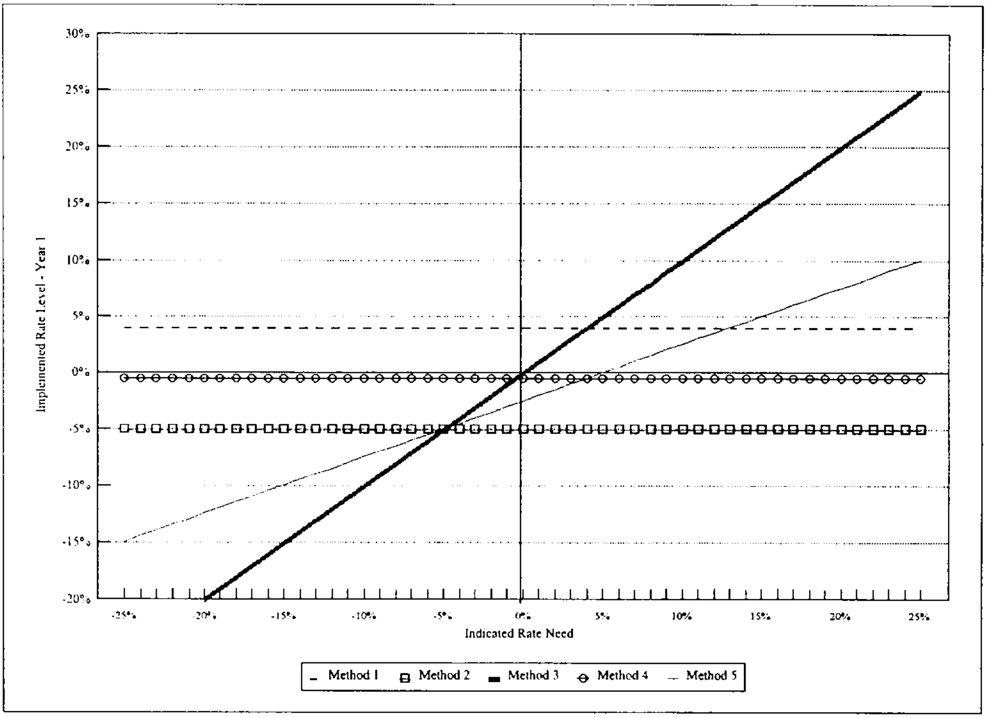


Exhibit 13 – Insurance Expense Exhibit Analysis of Fixed versus Variable Expenses

Category	Expense Dollars	Percent Fixed
Allowances to Managers	350	50%
Advertising	750	80%
Boards & Bureaus	-	0%
Surveys	-	0%
Audits	-	0%
Salaries	2,675	40%
Payroll Taxes	200	40%
Employee Relations	500	50%
Insurance	-	0%
Directors' Fees	-	100%
Travel	125	75%
Rent	175	100%
Equipment	425	100%
Printing	125	0%
Postage & Telephone	200	0%
Legal & Auditing	700	100%
TOTAL	6,225	57%

Exhibit 14 – Graphical Representation of Various Other Underwriting Expense Models

Assumes a current other underwriting expense ratio (to Direct Written Premium) of 18% and the ability/need to incrementally reduce/increase fixed expenses by 2% of DWP for every 15% decrease/increase in DWP.

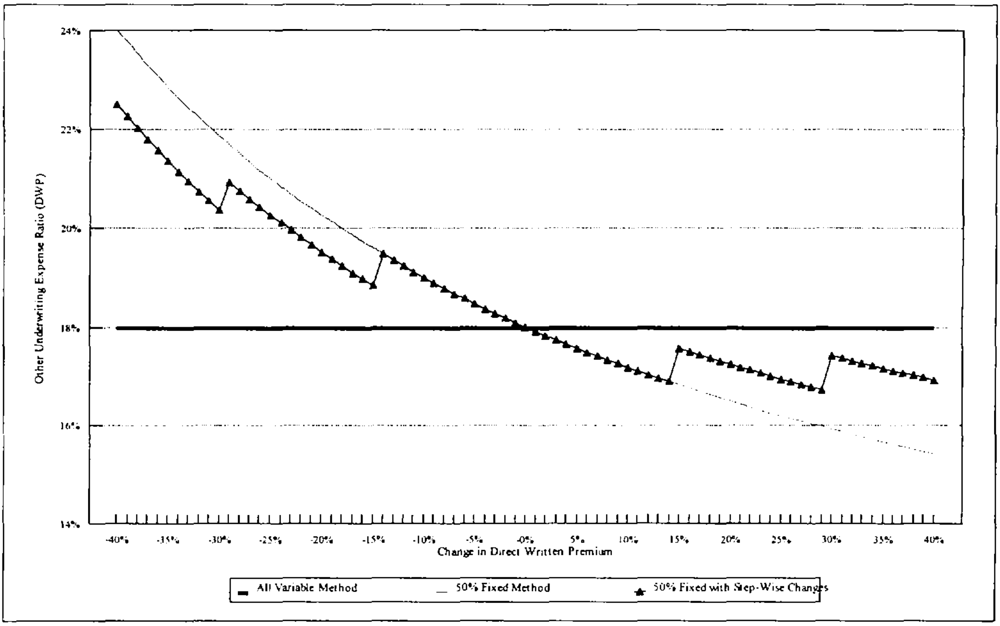


Exhibit 15 – Policyholders Dividend Ratio Parameterization

Expected Dividend Distribution			
	% of DWP by Dividend Plan		
Phase	No Plan	Plan 1	Plan 2
Mature Hard	40%	50%	10%
Immature Soft	25%	40%	35%
Mature Soft	10%	35%	55%
Immature Hard	25%	40%	35%

Dividend Payout Estimate		
Loss Ratio	Plan 1	Plan 2
20%	24%	37%
22%	23%	35%
24%	22%	34%
26%	20%	32%
28%	19%	30%
30%	18%	29%
32%	17%	27%
34%	16%	26%
36%	14%	24%
38%	13%	22%
40%	12%	21%
42%	11%	19%
44%	10%	18%
46%	8%	16%
48%	7%	14%
50%	6%	13%
52%	5%	11%
54%	4%	10%
56%	2%	8%
58%	1%	6%
60%	0%	5%
62%	0%	3%
64%	0%	2%
66%	0%	0%
68%	0%	0%
70%	0%	0%

Dividend Computation

Assume:
Mature Soft 2 years ago, with 56% loss ratio, $24 M DWP
Immature Hard last year, with 54% loss ratio, $30 M DWP

Expected Dividend = [Year 1 DWP * (% DWP in each plan * payout) +
Year 2 DWP * (% DWP in each plan * payout)] / (Total DWP)

Expected Dividend = [24* (0.35*0.02 + 0.55*0.08) + 30 * (0.40*0.04 + 0.35*0.10)] / (24 + 30)
= 5.1%

Exhibit 16 - Sample Residual Market Burden Parameterization

Residual Market Burdens as a Percentage of Direct Earned Premium

	Year in Market Condition			
Market Condition	1*	2	3	4th and subsequent
Mature Hard	5.0%	6.5%	7.0%	Increase 0.3 points per year (no maximum)
Immature Soft	1/3 of gap to mature soft*			1.0%
Mature Soft	1.0%	0.8%	0.6%	Decrease 0.2 points per year (minimum 0)
Immature Hard	1/3 of gap to mature hard*			5.0%

* Module is programmed to calculate the difference between the last observed mature market burden and the next logical mature market burden and . For example, assume a 3rd year mature hard market was simulated to change to immature soft. The difference between the 3rd year mature hard residual market burden (7.0%) and the first year mature soft burden (1.0%) which equals 6.0% (7.0% - 1.0%) would be divided by 3 to reflect a selection that generally it takes 3 years for a residual market burden to change from mature hard to mature soft. This 2.0 point reduction (0.06/3) would be subtracted from the prior year burden of 7.0% to compute a burden of 5.0%. If the market stayed in the immature soft state for a second year, the burden would be 3.0% (5.0% - 2.0%). The immature burdens are capped at the appropriate first year mature market burdens.

Surviving Price Deregulation

Emily Gilde, Ph.D., ACAS, and
Charles Bryan, FCAS, MAAA

Abstract

Theory and evidence from the past two decades demonstrate that price deregulation increases efficiency and lowers costs and prices. The impact of deregulation on profit, however, is ambiguous and depends in part on the industry's market structure. Theory predicts that, in competitive industries, price deregulation tends to reduce prices by about as much as costs, producing little change in profit. In industries with monopolistic characteristics, however, price deregulation may permit higher profits for the surviving firms. This paper argues that price deregulation itself can have a profound impact on an industry's market structure. Understanding how this change in market structure may occur is crucial in predicting the impact of price deregulation on an industry's profitability.

This paper focuses on how price deregulation is likely to impact the U.S. auto insurance industry. At present, the industry is competitive. Unlike the transportation industries, existing regulation has not seriously impeded entry into or exit from the market. In this competitive market environment, price deregulation may exert only a minimal impact on profits. On the other hand, increased pricing freedom is likely to stimulate development of new technologies for varying rates and segmenting markets, similar to those developed by the deregulated airline industry. Specifically, price deregulation will lead to more sophisticated class plans, more frequent rate changes, and more consumer shopping. To exploit these changes, insurers must integrate computer systems, increase employee skills in gathering and analyzing customer data, and offer high quality, individualized service. Price deregulation, thereby, may create new profit opportunities for the largest existing insurers, who possess the data and expertise for sophisticated analysis. Actuaries need to be prepared for these changes.

Introduction

In the late 1970's and early 1980's several industries in the energy, transportation, and financial sectors of the U.S. economy experienced significant deregulation. Price restrictions and restrictions on entry and exit were lifted for airlines, trucking, natural gas, petroleum, and brokerage. Rates for railroads and telecommunications were partially deregulated. Many banking industry restrictions on prices and entry were eliminated. The chart below lists recent major regulatory reform initiatives by industry (Winston, 1993, Table 1):

Deregulation Time Line	
Industry	Major Initiative
Brokerage	Securities Acts Amendments (1975)
Airlines	Airline Deregulation Act (1978)
Natural Gas	Natural Gas Policy Act (1978)
Petroleum	Decontrol of crude oil and refined petroleum products (executive orders beginning in 1979)
Trucking	Motor Carrier Reform Act (1980)
Railroads	Staggers Rail Act (1980)
Banking	Depository Institution Deregulation and Monetary Control Act (1980), Garn-St. Germain Depository Institutions Act (1982)
Telecommunications	AT&T Settlement (1982)
Cable Television	Cable Television Deregulation Act (1984)

Although this wave of deregulation had little impact on the insurance industry, the industry and several state legislatures have begun to show an increased interest in deregulating insurance. In 1998 Pennsylvania passed new legislation exempting carriers from rate and policy form filings involving large commercial risks. Other states plan to follow Pennsylvania's lead in 1999. Some analysts believe that deregulation of personal lines will follow. Several trade organizations, including the American Insurance Association, the Alliance of American Insurers, the National Association of Independent Insurers, and the National Association of Mutual Insurance Companies, have appealed to the National Association of Insurance Commissioners to hold a hearing on the issue of complete free market pricing. The groups argue that price controls have distorted markets, are political, deny choices to customers, and are an "artifact of industry practices and a relic of an economic theory discredited domestically and globally" (The Insurance Regulator, November 30, 1998 and December 14, 1998).

This paper examines how deregulation would likely affect the property and casualty insurance industry. Considering experiences in airlines and trucking, this paper draws implications for the impact of deregulation on auto insurance companies. We focus on airlines and trucking for two reasons. First of all, both of these industries have undergone swift and nearly complete price deregulation. Although banking deregulation is an obvious candidate for comparison with insurance, deregulation in that industry has occurred rather slowly and is not complete. Secondly, the underlying rationale for regulating the airline industry is similar to that of insurance: concern for public safety. The Federal government originally regulated the airlines to promote air safety. Airline deregulation in the late 1970's generated concern that competitive price wars would cause flying to become less safe. Similarly, one of the main justifications for insurance regulation has been to prevent reckless competition that could lead to insurer insolvency.

The paper has four sections. In Section I the paper describes the impact of deregulation on costs. The second section argues that the effect of deregulation on industry profits depends in part on market structure. In Section III the paper compares and contrasts the experiences of the top five airlines before and after deregulation. The aim is to better understand how auto insurance companies might survive deregulation. The fourth section is a conclusion.

I. Inefficiencies of Regulation

Regulation causes inefficiency by limiting competition and weakening the incentive to minimize costs. Entry and exit barriers prevent development of optimal networks and make it more difficult to shed excess capacity. Price regulation discourages efficient marketing and prevents firms from responding effectively to external disturbances. Empirical studies conclude that regulation has resulted in higher costs and prices in several industries, including airlines, trucking, railroads, telecommunications, cable television, brokerage, and natural gas (Winston, 1993, Table 3; Conference Strategy Board, July 1998). We discuss these issues below for airlines and property-casualty insurance.

Airline Industry [1]

In 1938 the Civil Aeronautics Board (CAB) began to regulate the airline industry. It immediately restricted entry. To begin servicing a new route, an airline first needed to obtain a certificate from the CAB showing that the presence of another carrier was required by "public convenience and necessity." One of the first acts of the CAB in 1938 was to grandfather in the 16 existing trunk carriers. Over the next 40 years, more than 150 applications were submitted to the CAB to add long distance routes, but not a single entry was allowed! According to critics, the CAB specifically rewarded inefficiency through its practice of awarding monopoly routes to airlines in troubled financial condition in order to "maintain competitive balance and prevent bankruptcies." The CAB also strictly controlled air fares. For a particular route, airlines could charge only coach or first class, with the fare based primarily on miles traveled. Since costs involved a heavy fixed component that did not vary with miles traveled, the long distance routes generated excess profit, while the short routes were unprofitable. The airlines responded by competing intensely for the profitable, long distance routes. Since regulation prevented competition based on price, airlines engaged in non-price competition based on flight frequency, meal quality, width of seats, and friendliness of staff. This behavior greatly increased operating costs.

Airline deregulation began in 1976 when the CAB began to allow airlines to offer discount fares ("super savers") and to make route awards to all applicants "fit, able, and willing" to compete. The Airline Deregulation Act of 1978 codified these changes. Deregulation led to several important changes in the airline industry. First, airlines developed a computerized pricing and reservation system that allowed them to vary prices according to marginal cost and differences in customer price sensitivity. Deregulation also dramatically altered airlines' route networks. Prior to deregulation, airlines traveled in a linear fashion between particular cities as required by CAB rules. This system was inefficient, often resulting in planes being flown half empty. After deregulation, airlines developed the much more efficient "hub and spoke" system for carrying passengers between cities. Third, after deregulation, airlines turned to non-unionized labor in order to reduce costs and began to use equipment more intensively. Wages of pilots fell dramatically, while time spent in the air increased. Similarly, planes were flown more hours each day. Over the past 20 years, the net impact of these changes has been a dramatic reduction in

[1] This discussion draws on Williams (1993).

operating costs. After adjusting for inflation, the average airfare has dropped about 40 percent (Aviation Week and Space Technology, November 9, 1998).

Property and Casualty Insurance

Regulation also contributes to higher production costs and prices in the insurance industry. First of all, regulated rates cannot be changed rapidly, making it more difficult for insurers to respond to cost changes or competitive changes. Since the regulatory process required for approval of rate increases can be especially time consuming, insurers are hesitant to decrease prices for fear of difficulty in increasing them later. Restrictions on classifying risks for pricing purposes also lead to higher costs. Studies show consistently that "class plan" restrictions result in an increase in the size of the "involuntary" market that insurers must support (Grabowski et al, 1989; Tennyson, 1998).

One interesting question is whether insurance regulation has created barriers to entry and/or exit akin to those present in the airline industry. If so, this is another important source of inefficiency. Although most economists assume that the insurance industry has low "natural" barriers to entry, the National Association of Independent Insurers (NAII) argues that the bureaucratic requirements that must be satisfied to enter a new state market are excessive (Harrington, 1984; NAII, undated). For example, licensing requirements for new companies can create delays of a year or more. Insurance regulation also includes exit barriers: the insurance commissioner in a state has the authority to deny a company's request to withdraw from a product line or a market. Despite these costs, many insurance companies have entered and exited the industry over the past few decades.[2]

II. Will Deregulation Increase Industry Profits?

The effect of deregulation on profits depends on the competitive characteristics of the industry. Some industries operate competitively, even under regulation. In these industries, firms enter and leave the market freely. There are no significant economies of scale. If regulators fix prices above the competitive level, non-price competition between firms eliminates any excess profits.

In this type of industry, deregulation tends to have little impact on profits. Although operating costs may fall, competition ensures that prices fall along with costs. Economists viewed the airline industry as competitive and predicted that deregulation would not increase industry profits by much (Winston, 1993). In other industries, regulation insulates firms from competition so that excess profits are made. When these industries deregulate, profits of existing firms can fall. The trucking industry falls into this category. Finally, industries with production technologies that involve economies of scale are the most likely to experience an increase in profits following deregulation. Deregulation frees these firms to exert market power and price discriminate. For example, the railroad industry, which is a natural monopoly, experienced a significant increase in profits following deregulation.

Predicting the effect of deregulation on profits is complicated by the dependence of market structure on existing technology. Deregulation of the past two decades shows that the sources of an industry's competitiveness can change rapidly because of new technology. In predicting whether deregulation will increase or decrease profits, it is important to consider how deregulation is likely to influence an industry's technology, and how the change in technology will influence market structure. The following examines these questions in the context of the airline, trucking and property-casualty insurance industries.

Airline Industry[3]

Airline industry profits suffered under CAB regulation. Despite the absence of price competition, the regulated airlines made no monopoly profits. These were dissipated through extensive non-price competition. On the eve of deregulation in the 1970's, economists generally predicted that deregulation would not increase airline profits by much, because the industry seemed so naturally competitive. Analysts assumed that deregulation would allow many new, small airlines to enter the market and challenge the established players. They believed that falling costs and fierce price wars would lead to a less concentrated, more competitive market structure.

At first, these predictions held. Initially, many new carriers did enter the industry. Most,

---

[2] Between 1980 and 1993, 613 new property-casualty companies were formed and 320 left the industry voluntarily or because of merger (National Association of Independent Insurers, 1989-1993).
[3] This discussion draws on Williams (1993).

however, have not been able to survive the ensuing price wars. Instead, a few large existing carriers have held on to and strengthened their market positions. They accomplished this in several ways. First, existing airlines monopolized take-off and landing slots. Consequently, new entrants could only fly at less attractive times. Second, incumbents already had national networks and could offer attractive frequent flyer packages, which regional carriers could not match. Third, incumbents had code sharing alliances with each other that put new entrants at a competitive disadvantage. Fourth, existing carriers, particularly American and United Airlines, increased their market dominance through the development of in-house computer reservation systems (CRS's). Airlines negotiated deals to have their CRS's installed exclusively in large travel agencies. Airlines then manipulated the presentation of CRS data to their advantage and paid agents extra commission to book flights with them. Finally, incumbents protected and expanded their markets by developing "hub and spoke" networks. In so doing, they were able to drive out small companies. For example, before deregulation, one small carrier, Frontier Air, had developed a modest "hub and spoke" system out of Denver. As a small commuter airline, Frontier did not have to follow all of the CAB regulations enforced for the trunk carriers and was highly competitive and innovative. Because Frontier was small, the large airlines were not threatened. After deregulation, however, the big airlines began rapidly developing new route systems. Continental and United began operating "hub and spoke" systems out of Denver. Unable to compete with these brand names, Frontier went out of business.

Overall, airline deregulation led to a substantial increase in growth and profitability for the industry and for several of the large, incumbent carriers in particular. No one believes any longer that the ideal airline is "small, quick, and nimble." The experience of the last two decades shows that the airline industry now involves significant economies of scale, and that "big is better." This fact has led to more and more calls for re-regulation of the airline industry in order to end "monopolistic abuse" by the large carriers.

Trucking

In the 1930's, the Interstate Commerce Commission (ICC) began to regulate the trucking industry, largely in order to protect the railroads from competition. The ICC kept trucking rates artificially high and restricted entry into the industry. These regulations allowed existing truck companies to make monopoly profits. In the early 1980's, the trucking industry was deregulated.

From 1980 to 1985, trucking firms faced massive market entry and falling prices. Although 4,500 trucking companies went out of business, there were 40 percent more trucking firms at the end of 1983 than before deregulation (Zingales, 1998). In this case, deregulation reduced industry profits (Winston, 1993).

Especially hard hit by deregulation were large trucking companies specializing in carrying many small loads for different customers (Fortune, April 27, 1998). These companies had developed extensive networks of hundreds of warehouses where the many partial loads were consolidated into full loads, which could then be transported more efficiently. All the consolidation was expensive and time consuming. Trucking firms could profit despite such inefficiency because regulation restricted entry. With deregulation, small, independent trucking firms emerged. These firms were willing to work for less and could deliver small loads more quickly. The small firms drove many of the large, partial-load specialists out of business. Thus, in this case, competitive advantage due to size and networks became obsolete under deregulation.

In conclusion, the airlines and trucking industries had very different experiences under deregulation. Although analysts predicted that airline deregulation would lead to more competitiveness and less market concentration, this prediction turned out to be wrong. Deregulation encouraged the development of new technologies, such as the "hub and spoke" system, which could only be exploited by the largest carriers. These new technologies led to increased profits and greater market concentration. By contrast, deregulation of the trucking industry led to reduced profits and less market concentration. The load and delivery networks developed by large trucking companies under regulation were no longer profitable. These strikingly opposite experiences show the danger in drawing conclusions about how deregulation will affect an industry, based solely on the industry's current technology and market structure.

Property and Casualty Industry

Joskow's seminal 1973 paper on the property and casualty regulation describes the industry as including a large number of firms with low market share, entering and exiting the market relatively freely and producing nearly identical products using constant returns to scale technology. As explained above, theory predicts that deregulation of a competitive industry such as this results in efficiency gains and customer benefits but has little impact on industry profits. Empirical studies comparing states with different degrees of auto insurance regulation generally supports this prediction. There is no consistent evidence that prices, profits, or loss ratios are consistently higher or lower in states requiring "prior approval" for rate changes. The results vary, depending on the time period and particular states being considered (Ippolito, 1979; Harrington, 1984; Tennyson, 1997; Bajtelsmit, 1998).

There is evidence, however, that the property and casualty industry does not fit the competitive industry paradigm completely. Particular insurance distribution systems do involve significant economies of scale.[4] Economies of scale clearly are important in a direct response arrangement, as most of the acquisition expenses are fixed costs associated with the start-up period. Several analysts have also suggested that economies of scale exist in production of insurance by direct writers since developing an exclusive agent distribution system entails high fixed costs (Joskow, 1973; Harrington, 1984). Once these fixed costs have been made, direct writers can produce more cheaply than independent agents. Thus, the cost of establishing an exclusive agency potentially creates a barrier to entry by small firms. Over the last few decades, concentration of the industry has increased due to faster growth by direct writers (Tennyson, 1997). Theory suggests that deregulation would most benefit the sectors of the insurance industry with economies of scale and market power.

The key question is how pricing deregulation would affect the profitability of different production and distribution strategies – and whether these strategies would benefit large established companies or new "upstarts." For example, after deregulation of the airlines, it became both possible and profitable to develop "hub and spoke" networks. Only the largest airlines could manage this on a national scale. Thus, a new economy of scale emerged after deregulation. On the other hand, the networks developed by the trucking industry became obsolete after deregulation, eliminating a source of competitive advantage to large firms. In both cases, these changes reflect the increased emphasis on competitive pricing following deregulation. The "hub and spoke" airline system allowed prices to fall – even though the system is less convenient for air travelers. A similar shift in emphasis may occur following deregulation of auto insurance. Considering the experience of the airline industry, the production and distribution strategies most likely to succeed in auto insurance will be those that permit the lowest prices, even at the expense of customer convenience.

Deregulation is likely to have a profound effect on insurance pricing. Deregulation created incentives for the airlines to vary prices. This in turn made complex pricing technologies profitable. Since the pricing technology was so expensive, it provided an opportunity for the largest airlines to exploit new economies of scale. A similar situation could occur if insurance is deregulated. With more pricing freedom, it will become more profitable to invest in risk assessment knowledge and systems. The class plans of personal auto insurers already reflect a wide variation in risk assessment capabilities. At present, those class plans are subject to regulatory filing and approval in most states. With deregulation, such plans would become proprietary information and thus potentially more important. This proprietary information will become a crucial new source of competitive advantage for companies large enough to make the investment.

Even if presented with new growth opportunities, the incumbents in an industry must move quickly to take advantage of their insider status. In Britain, traditional insurers failed to respond vigorously to the opportunities presented by price deregulation in the 1970's and 80's (Westall, 1997). At the end of the 1980's, a new company, Direct Line, began to sell insurance directly to policyholders over the telephone. This direct marketing pioneer has become extremely successful. Although the initial cost of advertising was very high, Direct Line has reduced costs dramatically by eliminating agents and branch expenses. Database management allows fine discrimination between risks and rapid premium rate adjustments. As Direct Line's market share grew, it began to operate its own body shops. These shops, which reduce claims costs through better management of the repair process, are only feasible for a company with a large market concentration. Currently, Direct Line's market share is 13 percent, about the same as Allstate's.

---

[4] Economies of scale also appear to be important in handling of claims, particularly in controlling use of body shops and local defense attorneys.

III. Surviving Deregulation: Experiences of the Airline Leaders

The table below shows market share leaders in the airline industry, before and after deregulation in 1978 (Aviation Week and Space Technology, 22 December 1986; Williams, 1993). Market share is measured as a percentage of revenue passenger miles (RPM):

Top Five Airlines' Domestic Market Share Comparison (% RPM)					
Market Share	1970	1978	1983	1993	1998
1	United	United	United	American	United
2	TWA	American	American	United	American
3	American	Delta	Delta	Delta	Delta
4	Pan Am	Eastern	Eastern	Northwest	Northwest
5	Eastern	TWA	TWA	US Air	US Air

As the table shows, the two market share leaders in 1978, United and American, are still the top two airlines today. The combined market share of the two airlines has increased from 34.6 percent in 1978 to about 40 percent in 1993, and 36.9 percent in 1998. The market share of Delta, the third largest airline, rose from 12 percent to 15.7 percent. TWA's market share has dropped from 14 percent to under 5 percent. Pan Am and Eastern Airlines went out of business in 1991 (Aviation Week and Space Technology, January 28, 1991). We discuss TWA, Pan Am, Eastern, and Delta in more detail below.

TWA and Pan Am [5]

Both airlines specialized in domestic long distance and international travel, which are the most competitive segments of the market. On the eve of deregulation, the market shares of TWA and Pan Am were falling. The airlines were especially hurt by the recession and fuel crisis in the 1970's. During the 1980's, they continued to lose market share because they did not have a good domestic system to feed their international routes. Unlike United, American, and Delta, TWA had only a small hub in St. Louis, while Pan Am had none. Pan Am sold its international routes in a desperate effort to survive. TWA sold its profitable London routes to raise money and allowed its air fleet to deteriorate. By 1990, TWA had diversified into real estate, fast foods, and hotels. The non-airline portion of the business continues to absorb capital, making it difficult for TWA to upgrade its planes. By 1992, TWA had survived two bankruptcies and Pan Am was gone.

[5] This discussion comes from The Times (July 19, 1996) and Business Week (May 19, 1980).

Eastern Airline[6]

By 1975, Eastern was already in deep trouble. At the time, analysts predicted that the airline could not survive deregulation. Eastern's problems were many. It operated many short flights between small cities, requiring too many planes and too many people. The planes used for these routes were too big. Eastern's balance sheet included considerable long term debt. Its corporate offices were split between Miami and New York, leading to confused management. Eastern had a reputation for poor quality service. The airline ceased flying in 1991.

Delta[7]

Delta is a profitable airline with a strong balance sheet. The source of its strength is its conservative approach to finance and management. Delta has always engaged in consistent capital spending on its fleets. It buys steadily and carefully, in good times as well as bad. In this way, it keeps its fleets relatively young, and avoids excessive spending in any year. During the 1980's, Delta had the highest retained cash flow as a percentage of long term debt of any U.S. carrier. Another strength is Delta's hub in Atlanta. At its hub, Delta consolidates all of its operations, which generates economies of scale for the airline. Delta also has a reputation for its loyal, well paid, but non-union workforce. Delta, however, has been slow to seize new opportunities. Following deregulation, Delta hesitated to buy international routes and was slow to develop CRS's. The airline allowed American to beat it out in developing a second major hub in Dallas. Finally, in 1986, eight years after deregulation, Delta became a transcontinental airline when it merged with Western Airlines. At that point, its market share (including international passenger miles) shot from sixth place to third. In 1992, when other carriers were advertising price cuts, Delta increased spending to promote its good service! Delta remains strong because it stays out of too much debt and focuses on long term strategies.

Analogies Between the Airlines and Private Passenger Auto Insurers

It is instructive to draw an analogy between these experiences and what might happen to the major players in the personal auto insurance industry should deregulation occur. The chart below

---

[6] See Business Week (Dec. 22, 1975) and Aviation Week and Space Technology (January 28, 1991).
[7] Forbes (Sept. 15, 1980), Aviation Week and Space Technology (Oct. 14, 1991), Brandweek (May 18, 1992), Business Week (Nov. 8, 1982), and Air Transport World (June, 1993).

shows market shares for personal auto for the five largest insurers (One Source Information Services, Inc. Market Share Application):

Market Share: Personal Auto (%)			
Insurance Group	1993	1995	1997
State Farm	21.8 %	21.7 %	20.8 %
Allstate	11.7	11.9	12.2
Farmers	0.0	5.8	5.9
Nationwide	3.5	3.7	3.9
Progressive	1.4	2.4	3.7

The two largest insurers, State Farm and Allstate, can be compared to United and American, the two largest airlines. Like United and American, State Farm and Allstate have dominated the industry for years. Both are national insurers with a long history of excellence in risk assessment. The companies are large and financially stable. They are both in a strong position to take advantage of new opportunities presented by deregulation. The third and fourth market share players, Farmers and Nationwide, are both regional companies who have attempted in recent years to become more national. These companies resemble Delta, also a regional company prior to deregulation. Deregulation may well lead to a battle between these companies to become the third largest auto insurer.

Who will be the Pan Am and Eastern of auto insurance after deregulation?

IV. Final Comments

Although deregulation clearly increases efficiency, the impact on the profits of existing firms in the industry depends on market structure and competitiveness:

- If an industry behaves competitively, consumers are the main beneficiaries of deregulation, as efficiency gains are passed on to them in lower prices. The effect on existing firm profits may be minimal.
- To the extent that deregulation protects industry profits by fixing prices above the competitive level or by restricting entry into the market, existing firms may be worse off after deregulation.

Faced with new competition, the existing firms may lose profits or go out of business.

- Existing firms in industries with barriers to entry tend to be more profitable after deregulation. Essentially, deregulation allows the firms to exploit monopolistic power.

The wild card in this analysis is technological change. The experience of the last two decades shows that deregulation stimulates technological change. This paper has tried to show how deregulation, by freeing firms to pursue new pricing and distribution strategies, suddenly makes new technologies much more profitable than before. These technologies may give rise to new economies of scale and may make traditional economies of scale obsolete. Thus, in predicting how deregulation may influence profits of existing firms in an industry, it is important to consider how deregulation may influence the sources of competitive advantage.

How would deregulation influence the property and casualty industry? The property and casualty insurance industry has many of the characteristics of a competitive market. There is no consistent evidence showing that regulation allows insurers to make excessive profits or that it seriously restricts entry and exit from the industry. Ignoring technological change, this suggests that deregulation might not affect the profits of existing insurers very much. Deregulation, however, would give insurers the freedom to develop new market segments and rate relativities and to respond quickly to external shocks. Technologies that permit this will become much more profitable. Insurers with the capital to develop and implement new pricing capabilities will experience new competitive advantages. Insurers who do not respond quickly may find that traditional sources of competitive advantage, such as branch offices, agent networks, and relationships with regulators, are no longer profitable. One piece of evidence suggesting that these changes may be coming: in the last 26 months, Progressive Insurance reduced its rates in Texas, a state allowing flexible rating, on seven separate occasions (PR Newswire, May 14, 1998). By contrast, in Illinois, the state with perhaps the most pricing freedom, State Farm, Allstate, and Nationwide changed their rates only five or six times in the past seven years. Actuaries need to prepare for these changes.

BIBLIOGRAPHY

"Airline Woes Catch Up With Delta," Business Week, 8 November 1982, LEXIS-NEXIS Academic Universe 12/29/98.

Bajtelsmit, Vickie and Raja Bouzouita. "Market Structure and Performance in Private Passenger Automobile Insurance," The Journal of Risk and Insurance, 1998, 65:3, pp. 503-514.

"Can Frank Borman Make Eastern Take Off?" Business Week, 22 December 1975, LEXIS-NEXIS Academic Universe 12/22/98.

"Commercial Lines Dereg," The Insurance Regulator, 30 November 1998, LEXIS-NEXIS Academic Universe 12/22/98.

The Corporate Strategy Board, "Weathering the Storm: Successful Strategies for Surviving Deregulation," July 1998.

"Demand for Prompt P/C Dereg Irks NAIC," The Insurance Regulator, 14 December 1998, LEXIS-NEXIS Academic Universe 12/22/98.

Feldman, Joan. "Straining the Family Ties: Delta Air Lines, Inc.," Air Transport World, June 1993, LEXIS-NEXIS Academic Universe 12/29/98.

Grabowski, Henry; Kip Viscusi; and William Evans. "Price and Availability Tradeoffs of Automobile Insurance Regulation," The Journal of Risk and Insurance, June 1989, 56:2, LEXIS-NEXIS Academic Universe 12/29/98.

Harrington, Scott. "The Impact of Rate Regulation on Prices and Underwriting Results in the Property-Liability Insurance Industry: A Survey," The Journal of Risk and Insurance, 1984, 51, pp.577-617.

Ippolito, R. "The Effects of Price Regulation in the Automobile Insurance Industry," The Journal of Law and Economics, April 1979, 22, pp. 55-89.

Joskow, Paul. "Cartels, Competition, and Regulation in the Property-Liability Insurance Industry," Bell Journal of Economics and Management, Autumn 1973, 4, pp. 375-427.

Kolcum, Edward. "Airline Mainstay Eastern Stops Flying After 62 Years," Aviation Week and Space Technology, 28 January 1991, LEXIS-NEXIS Academic Universe 12/22/98.

National Association of Independent Insurers. "Improving Market Competition: Recommendations for "Streamlining" State Insurance Laws and Regulations," undated.

National Association of Independent Insurers. Greenbook, annual volumes, 1989-1993.

One Source Information Services, Inc., Market Share Application, 1998.

O'Reilly, Brian and Alicia Moore. "On the Road," Fortune, 27 April 1998, LEXIS-NEXIS Academic Universe, 12/29/98.

Reguly, Eric. "Former Lord of the Skies Faces Grim Battle to Keep Up Revival," The Times, 19 July 1996, LEXIS-NEXIS Academic Universe 12/22/98.

Seneker, Harold. "Delta is Ready," Forbes, 15 September 1980, LEXIS-NEXIS Academic Universe 12/22/98.

"Seventh Rate Decrease In 26 Months For Progressive, State's Fourth Largest Auto Insurer," PR Newswire, 14 May 1998, LEXIS-NEXIS Academic Universe 12/29/98.

Shifrin, Carole. "Tough Policy Questions: A Deregulation Legacy," Aviation Week and Space Technology, 9 November 1998, LEXIS-NEXIS Academic Universe 12/22/98.

"The Strategy Squeeze on the Airline," Business Week, 19 May 1980, LEXIS-NEXIS Academic Universe 12/22/98.

Tennyson, Sharon. "The Impact of Rate Regulation on State Automobile Insurance Markets," Journal of Insurance Regulation, Summer 1997, 15:4, pp. 502-523.

"Top Ten Airlines' Market Share Comparison." Aviation Week and Space Technology, 22 December 1986, LEXIS-NEXIS Academic Universe 12/29/98.

Underwood, Elaine. "Delta Puts a Polish on Its Image," Brandweek, 18 May 1992, LEXIS-NEXIS Academic Universe 12/29/98.

Velocci, Jr., Anthony. "Delta's Fiscal Caution to Remain After Pan Am Acquisitions," Aviation Week and Space Technology, 14 October 1991, LEXIS-NEXIS Academic Universe 12/29/98.

Westall, Oliver. "Invisible, Visible, and 'Direct' Hands: An Institutional Interpretation of Organizational Structure and Change in British General Insurance," Business History, October 1997, 39:4, LEXIS-NEXIS Academic Universe 12/29/98.

Williams, George. The Airline Industry and the Impact of Deregulation. Ashgate Publishing Co: Brookfield, Vermont, 1993.

Winston, Clifford. "Economic Deregulation: Days of Reckoning for Microeconomists," Journal of Economic Literature, September 1993, 31, pp. 1263-1289.

Zingales, Luigi. "Survival of the Fittest or the Fattest? Exit and Financing in the Trucking Industry," Journal of Finance, June 1998, 53:3, pp. 905-938.