Stochastic Modeling and Error Correlation in Dynamic Financial Analysis

by Son T. Tu, Ph.D., ACAS, MAAA

STOCHASTIC MODELING AND ERROR CORRELATION IN DYNAMIC FINANCIAL ANALYSIS

SON T. TU
SCRUGGS CONSULTING
ARGYLE, TEXAS

ABSTRACT

New treatments of stochastic modeling and error correlation in dynamic financial analysis are introduced. The former refers to the methods for modeling individual insurance operations. The latter refers to the technique for considering the interactions and correlations among those operations. The stochastic chain ladder model, a new technique for loss development, is also introduced and is shown to be an integral part of DFA.

1. INTRODUCTION

Dynamic financial analysis is now segregated into two different philosophies: that of stochastic simulation, and that of scenario testing. Feldblum[1] discussed the strengths and weaknesses of the two approaches. We believe that the two need not be separate and competing, but indeed need to be complementary. With the model proposed in this paper, we hope to narrow the gap between these two approaches.

The construction of this model is motivated principally by four factors. The first two, qualitative in nature, are understandability and flexibility. The other two, quantitative in nature, are the stochastic modeling of individual insurance components and the error correlation among those components.

Many users and interested parties of a dynamic financial analysis model are not actuaries and technical analysts. Therefore, it is important that these users can gain, relatively easily, a good understanding and confidence in the model.

Secondly, the usage to which a model will be applied varies widely. In some studies, the analyst may only be interested in the overall picture; and relatively few insurance operations need be modeled. In another study, a relatively large number of operations needed to be included, because more detailed quantitative analyses are required. A model should have sufficient flexibility to suit both extremes.

To satisfy the above two factors, our model is controlled by a set of governing equations. One or more of these equations describes each operation. The model has understandability, because each equation is usually a readily accepted insurance formula. The model has flexibility, because the number of equations can be expanded or contracted, depending on the needs and objectives of the analyst.

There can be several stochastic models for generating observations of an insurance variable, such as the loss ratio or the investment return. By far, the most popular among actuaries is the averaging technique, where the observations are assumed to be random about some average. In a subsequent section, we present two other alternatives, which we name the current-value and current-change models. We show that they fit the historical data used in this study better than the averaging technique. These two models have analogies in time-series analysis.

An important consideration in any DFA model is the correlation among the variables considered in the analysis. Depending upon the sign of the correlation between two variables, the correlation can be either stabilizing or destabilizing, a concept that we will elucidate in section 4. The correlation coefficients among the variables will be measured. As a natural and necessary by-product, we present a technique for the generation of correlated random numbers.

In this paper, we aim only to demonstrate the concept and potential of the model. We have simplicity as one of the objectives of the paper; therefore, the number of operations has been kept to a relative few. We will study a hypothetical insurer, which is assumed to have written only Workers Compensation for the last ten years. Our study projects five years into the future. At the end of that time frame, among other quantities, we want to examine the probability of ruin. To work with realistic data, all of the relevant data has been taken from the 1997 Best's Aggregates and Averages publication.

In section 2, we present the governing system of equations used in this study. In section 3, we present the hypothetical initial state of the company

In section 4, we present the stochastic modeling of the insurance random variables. In section 5, we model paid losses. For this purpose, we will introduce our research on the stochastic chain ladder and Bornhuetter-Ferguson loss reserving models.

In section 6, a technique for the generation of correlated random numbers will be introduced. In section 7, we pull together the materials in all the preceding sections to generate simulated solutions for the next five years.

In section 8, we show that the simulated results can be assumed normally distributed. In section 9, we outline the many potential extensions to the model.

In the concluding section, we summarize and discuss the criteria by which a user of dynamic financial analysis would evaluate one strategy or decision as being superior to another.

[1] Sholom Feldblum, "Forecasting the Future: Stochastic Simulation and Scenario Testing." *Incorporating Risk Factors in Dynamical Financial Analysis*, 1995 CAS Discussion Paper Program.

2. Governing System of Equations

Following is the list of random variables that we consider for this hypothetical study:

AS = Assets,
LI = Liabilities,
SU = Surplus,
PL = Paid Losses,
IL = Incurred Losses,
II = Investment Income,
EP = Earned Premiums,
PD = Paid Policyholders' Dividends,
PE = Paid Underwriting Expenses,
WP = Written Premiums,
IIR = Investment Income Ratio,
UER = Calendar-Year Underwriting Expense Ratio,
PDR = CY Policyholders' Dividend Ratio,
EPR = CY Earned Premium Ratio,
LR = Accident-Year Loss and LAE Ratio,
F = value of the loss cumulative distribution function, and
ε = process error of the paid loss.

Each of the variables takes the argument of time. It is an understood that a variable refers to the value during that year (such as written premiums and paid losses) or at year-end (such as assets and liabilities). For simplicity, we will consider only yearly intervals.

Consider the following system of nine equations:

(2.1) $$AS(t) = LI(t) + SU(t)$$

(2.2) $$LI(t) = LI(t-1) - PL(t) + IL(t)$$

(2.3) $$AS(t) = AS(t-1) + WP(t) + II(t) - PL(t) - PE(t) - PD(t)$$

(2.4) $$EP(t) = EPR(t) * [WP(t-1) + WP(t)]$$

(2.5) $$II(t) = IIR(t) * AS(t)$$

(2.6) $$PE(t) = UER(t) * WP(t)$$

(2.7) $$PD(t) = PDR(t) * EP(t)$$

(2.8) $$IL(t) = LR(t) * EP(t)$$

(2.9) $$PL(t) = \sum_{all\ AY} LR(i) * EP(i) * [F(k+1) - F(k)] * (1+\varepsilon)$$

Even though there are many variables, many of them are inter-related. In fact, only five of them are independent. They are the investment income ratio (IIR), the U/W expense ratio (UER), the dividend ratio (PDR), the earned premium ratio (EPR), and the loss ratio (LR). In section 4, we will model the stochastic behaviors of these ratios from historical data and calculate their correlations. In section 6, we will simulate correlated random numbers for the ratios.

The most complex equation in the above set is (2.9), which is the sum of the paid losses for all accident years up to the evaluation date. In section 5, we will explain our stochastic loss reserving models and the workings of (2.9).

This is only an example of a set of governing equations. The analyst designs the exact set to meet his own needs. This offers great generality and flexibility.

3. THE INITIAL STATE OF THE COMPANY

Our insurer has been in existence for the last ten years, and writes only Workers Compensation. In the following table, loss ratios, earned premiums, paid losses, and liabilities for the past ten years are listed.

Table 1: The Initial State of the Company

AY	LR	Ear. Prm	Inc. Loss	Paid Loss	Liability
1987	91.19%	5,002	4,562	4,209	353
1988	92.91%	5,403	5,020	4,558	462
1989	93.76%	5,835	5,471	4,865	606
1990	91.72%	6,302	5,780	4,997	783
1991	85.15%	6,806	5,795	4,819	976
1992	74.40%	7,350	5,469	4,299	1,170
1993	72.50%	7,938	5,755	4,150	1,605
1994	72.14%	8,573	6,185	3,859	2,326
1995	74.21%	9,259	6,871	3,232	3,639
1996	75.77%	10,000	7,577	1,593	5,984

The loss ratios were obtained from Schedule P – Part 1D of the Best's Aggregates & Averages publication. The incurred loss is the product of the loss ratio and the earned premium. The paid loss is a function of the incurred loss and the cumulative distribution function, which will be explained in section 5. The liability is the difference between the incurred and paid losses.

We assume the following initial liabilities, surplus and assets, with the initial year being 1996:

$$(3.1) \qquad LI(0) = 17{,}904; \quad SU(0) = 6{,}667; \quad AS(0) = 24{,}570.$$

The total liability is the sum of the last column in Table 1, and (3.1) satisfies (2.1). We assume that the insurer has the following target written premiums for the next five years:

Table 2: Target WP

Year	WP
1997	10,800
1998	11,664
1999	12,597
2000	13,605
2001	14,693

We could as easily assume that the written premium is a product of the premium-to-surplus ratio and the surplus:

$$(3.2) \qquad WP(t) = PSR(t) * SU(t-1).$$

In such case, we would add (3.2) to the set of governing equations in the previous section. We elect not to follow this route, primarily because of a lack of historical data for the Workers' Compensation premium-to-surplus ratios. The analytical treatment of the two cases is similar.

4. STOCHASTIC MODELING OF THE INSURANCE RATIOS

In this section, we present the modeling of the loss and LAE ratio, the U/W expense ratio, the paid dividend ratio, the investment income ratio, and the earned premium ratio. The first is on an accident-year basis; the others on a calendar-year basis. All are taken from Best's Aggregates & Averages – Cumulative by Line Underwriting Experience.

Table 3: The Historical Insurance Ratios

YEAR	LR	UER	PDR	IIR	EPR
1987	91.19%	15.00%	7.20%	10.20%	
1988	92.91%	13.40%	9.40%	10.90%	51.20%
1989	93.76%	13.00%	7.10%	11.40%	51.30%
1990	91.72%	13.40%	5.60%	10.80%	51.40%
1991	85.15%	14.60%	6.00%	11.70%	52.80%
1992	74.40%	16.50%	6.50%	16.60%	50.60%
1993	72.50%	17.20%	6.60%	14.60%	48.10%
1994	72.14%	18.60%	9.20%	13.90%	46.40%
1995	74.21%	20.30%	9.50%	16.70%	46.30%
1996	75.77%	23.30%	9.00%	16.90%	47.60%

The corresponding ratios for future years are, of course, random. The simulation of the random numbers is determined by the historical patterns. There are two things to consider in these patterns: the pattern within each set of ratios, and the correlation between any two sets of ratios.

To determine the pattern within each set of ratios, we consider three models: the average-value model, the current-value model, and the current-change model. For a given set of data, we pick the model that gives the least error deviation.

The average-value, or the averaging, model states that a random number is normally distributed about some average:

$$(4.1) \qquad x_i = \bar{x} + \varepsilon_i .$$

The first term on the right-hand side (RHS) of (4.1) is the average; the second is the uncorrelated errors of mean zero and some standard deviation. If we apply (4.1) to the loss ratios of Table 3, we have:

$$(4.2) \qquad \bar{x} = 82.4\%, \quad \sigma(\varepsilon_i) = \left\{ \frac{1}{n-1} \sum_i \varepsilon_i^2 \right\}^{1/2} = 9.37\% .$$

Therefore, the loss ratios have a mean of 82.4%, and the standard deviation of the errors is 9.37%. Note that n is the number of observations, and the degree of freedom is one less than that value since an average has to be estimated.

There are two sources of error in the average-value model. There is the parameter error, associated with the uncertainty in the estimation of the average. Also, there is the process error, which is associated with the random errors.

If we take a closer look at the loss ratios in Table 3, the average-value model does not seem to be appropriate. In the earlier years, 1987-91, all the ratios are greater than the average. In the later years, 1992-96, they are all smaller. Therefore, we next propose the current-value model:

$$(4.3) \qquad x_{i+1} = x_i + \varepsilon_{i+1} .$$

This model says that a random number tends to stay about its current value. The errors are assumed to be uncorrelated and of mean zero. If we apply (4.3) to the loss ratios, we get:

$$(4.4) \qquad \sigma(\varepsilon_i) = \left\{ \frac{1}{n} \sum_i \varepsilon_i^2 \right\}^{1/2} = 4.44\% .$$

The deviation of the average-value model is much greater than that of the current-value model, indicating that the latter is a much better fit for the observed loss ratios.

Contrary to the average-value model, there is only one source of error in (4.3), the process error, since no parameter has to be estimated in that equation.

If we look even more closely at the loss ratios in Table 3, we notice that an increase in the ratio tends to be followed by another increase, a decrease tends to be followed by another decrease. Therefore, we propose a third model, the current-change model:

$$(4.5) \qquad z_{i+1} = x_{i+1} - x_i, \quad z_{i+1} = z_i + \varepsilon_{i+1}.$$

This model says that the next change tends to be equal to the current change. And, like the current-value model, it only has process error. If we apply (4.5) to the loss ratios, we get:

$$(4.6) \qquad \sigma(\varepsilon_i) = \left\{ \frac{1}{n} \sum_i \varepsilon_i^2 \right\}^{1/2} = 4.09\%.$$

Since the current-change deviation is smallest, it represents the best fit, and we choose it to model the loss ratios in our analysis.

It makes a great deal of difference which model is chosen to represent a set of random variables. For instance, if we choose the average-value model for the loss ratios, then the simulated 1997 loss ratios have a mean of 82.4% and deviation of 9.37%, as shown in (4.2). If we choose the current-value model, they have a mean of 75.8% and deviation of 4.44%. If we choose the current-change model, they have a mean of 77.4% and deviation of 4.09%.

For the other four ratios, we will use the current-change model. The error terms have the following deviations:

Table 4: The Standard Deviations of the Ratios

LR	UER	PDR	IIR	EPR
4.09%	1.57%	1.49%	2.06%	1.49%

We now turn to the calculation of the correlation between any two sets of errors. Let's consider the loss and the dividend ratios. They have the following errors:

Table 5: Values of two Error Sets

YEAR	LR	PDR
1988		.022
1989	-.009	-.023
1990	-.029	-.015
1991	-.045	.004
1992	-.042	.005
1993	.089	.001
1994	.016	.026
1995	.024	.003
1996	-.005	-.005

The correlation coefficient of the two sets in Table 5 equals .185. The correlation coefficient is defined as:

$$(4.7) \qquad \rho(A,B) = \frac{Cov(A,B)}{\sigma(A)\sigma(B)}.$$

The correlation coefficients among the five ratios are found to be:

Table 6: Correlation Coefficients of the Five Insurance Ratios					
	LR	UER	PDR	IIR	EPR
LR	1.000	0.000	0.185	-0.528	-0.486
UER		1.000	0.000	0.000	0.132
PDR			1.000	0.000	-0.429
IIR				1.000	0.000
EPR					1.000

For any coefficient with an absolute value smaller than .1, we assume it to be statistically insignificant and set it equal to zero.

For this study, we use the empirical coefficients. For a genuine study, the analyst should decide whether the observed correlations seem reasonable. He may decide to override them if they do not.

The significance of Table 6 is this: not only should the simulated ratios have the deviations in Table 4, but they should have the correlations shown there. These have real consequences regarding the stability of the insurance process. For instance, that the loss and investment ratios have a negative correlation is destabilizing. The negative correlation means that a higher-than-average loss ratio tends to be coupled with a lower-than-average investment ratio, and vice versa. Taking the former case, the higher-than-average loss ratio means more loss payments, and the lower-than-average investment ratio means less investment income. If two quantities in conjunction tend to have the same effects on the balance sheet, then the correlation is destabilizing. Conversely, if they tend to impart opposite effects, then the correlation is stabilizing.

Every correlation in Table 6 destabilizes, except for the positive correlation between the expense and the earned premium ratios. In this case, if the insurer experiences higher-than-average expenses, then it also experiences higher-than-average earned premiums. The two have opposite impacts on the balance sheet, because the higher outgo (expenses) counteracts the higher income (earned premiums).

We emphasize that there are other reasonable stochastic models for the variables. This aptly demonstrates the tremendous flexibility and variety available to the analyst. The bottom line is that he should have confidence that the underlying model is representative of the future.

We are grateful to a review who pointed out that the current-value and current-change models have analogies in time-series forecasting.

5. Stochastic Modeling of Paid Losses

We have developed two stochastic loss reserving models: one based on the traditional chain ladder method, and the other on the Bornhuetter-Ferguson method. We have written a paper on each of these models.[2] [3]. The interested reader should contact the author for copies of the papers.

Basically, we model the stream of paid losses for an accident year as a function of a cumulative distribution function. The function that we use for Workers Compensation loss payment is the transformed lognormal:

$$(5.1) \qquad F(t;\mu,\sigma,\tau) = \Phi\left\{sign(\ln t)\left|\ln t\right|^{\tau};\mu,\sigma\right\}.$$

In (5.1), Φ is the normal distribution of mean μ and deviation σ. The argument t is measured in years.

Let an accident year have earned premium EP and loss ratio LR. Let Y_t be the incremental loss payment for that accident year between the report years t and $t+1$. Then the stochastic Bornhuetter-Ferguson model gives the following relationship:

[2] Son T. Tu, "The Application of Cumulative Distribution Functions in the Stochastic Chain Ladder Model," Scruggs Consulting Research Paper. This paper is in the process of publication in the *Casualty Actuarial Society Forum*.

[3] Son T. Tu, "The Stochastic Bornhuetter-Ferguson Model," Scruggs Consulting Research Paper.

(5.2) $\qquad Y_{t+1} = EP * LR * [F(t+1) - F(t)] * (1 + \varepsilon_{t+1}),$

where the process error ε_t is a normal distributed random variable of zero mean and some deviation.

In our paper on the stochastic Bornhuetter-Ferguson model, we demonstrate how to fit (5.2) to a triangle of incremental payments and come up with estimates of the function parameters. From an actual Best's Aggregates and Averages paid loss triangle, we derive the following estimates:

Table 7: The Estimates for the Parameters

	Estimate	Deviation
μ	.7840	.0591
σ	.9733	.0360
τ	.9286	.0352

If we use the estimates in Table 7 in (5.1), then we can obtain the following values for the distribution function:

Table 8: The Values of the Cumulative Distribution Function

time	1	2	3	4	5	6	7	8	9	
F	.2103	.4703	.6239	.7211	.7861	.8316	.8646	.8892	.9080	

Table 8 says that, after one year, 21.03% of payments for any accident year has been paid. After ten years, 92.27% has been paid, and therefore 7.73% has yet to be paid.

The function parameters also have the following matrix of correlation coefficients:

Table 9: The Matrix of Parameter Correlation Coefficients

	μ	σ	τ
μ	1.000	.9815	-.7633
σ		1.000	-.8180
τ			1.000

The way that we use (5.2) in the DFA model is as follows. For any calendar year, the loss payments are the sum of the paid losses for all accident years. The paid loss for each accident year is modeled by (5.2).

For the ten accident years in the past, we assume that the earned premiums and loss ratios are fixed, given by the values in Table 1. For the five accident years in the future, the earned premiums and loss ratios are stochastic quantities, given by numerical simulation. For this exercise, the process error in (5.2) has a standard deviation of 10.36%.

6. Generation of Correlated Random Numbers

Section 4 shows the necessity to generate five correlated insurance ratios. Section 5 shows the necessity to generate three correlated function parameters. In this section, we present a general technique to generate correlated random numbers. For instance, from Table 4, the errors of the loss ratios and the investment income ratios have expected deviations of 4.09% and 2.06%, respectively. But additionally, from Table 6, those errors have an expected correlation coefficient of -.528. In this section, we will introduce a technique to generate errors with the desired correlation characteristics. We will present some very technical work, which is needed for the sake of stochastic realism. But the reader may decide to skip this section without fear of losing the continuity among the other sections.

We will work with three variables. The technique can be easily generalized to any number of variables. Let's suppose that we need to generate three normally distributed random numbers X, Y, Z, and together they have the following variance matrix:

$$(6.1) \qquad Var(\mathrm{X}) = \begin{bmatrix} Var(X) & Cov(X,Y) & Cov(X,Z) \\ Cov(X,Y) & Var(Y) & Cov(Y,Z) \\ Cov(X,Z) & Cov(Y,Z) & Var(Z) \end{bmatrix}.$$

Instead of this problem, we are going to generate three uncorrelated normally distributed numbers A, B, C such that:

$$(6.2) \qquad Var(\mathrm{A}) = \begin{bmatrix} Var(A) & 0 & 0 \\ 0 & Var(B) & 0 \\ 0 & 0 & Var(C) \end{bmatrix}$$

Note that the second problem is much simpler than the original one.

We express the two sets of numbers as:

$$(6.3) \qquad \begin{aligned} A &= X, \\ B &= Y + b_1 X, \\ C &= Z + c_1 X + c_2 Y, \end{aligned}$$

where b_1, c_1, c_2 are unknown variables to be found. We will use the condition of no correlation among A, B, C to determine these unknowns. The condition that

$$(6.4) \qquad Cov(A,B) = 0.$$

We apply (6.4) to the first two equations of (6.3) to derive:

$$(6.5) \qquad b_1 Var(X) = -Cov(X,Y).$$

The conditions that

$$(6.6) \qquad Cov(A,C) = Cov(B,C) = 0,$$

yield

$$(6.7) \qquad \begin{aligned} &c_1 Var(X) + c_2 Cov(X,Y) = -Cov(X,Z), \\ &c_2 \left[Var(Y) + b_1 Cov(X,Y)\right] = -Cov(Y,Z) - b_1 Cov(X,Z). \end{aligned}$$

The first equation of (6.6) and the first and third equations of (6.3) give the first of (6.7). The second equation of (6.6) and the last two equations of (6.3) yield the second of (6.7). Equations (6.5) and (6.7) give the values of the unknown b_1, c_1, c_2. Taking the variance of (6.3), we have:

$$(6.8)\qquad \begin{aligned} Var(A) &= Var(X), \\ Var(B) &= Var(Y) + b_1^2 Var(X) + 2b_1 Cov(X,Y), \\ Var(C) &= Var(Z) + c_1^2 Var(X) + c_2^2 Var(Y) + 2[c_1 Cov(X,Z) + c_2 Cov(Y,Z) + c_1 c_2 Cov(X,Y)] \end{aligned}$$

Now we generate three uncorrelated random numbers with the variances in (6.8). Then we can invert (6.3) to obtain:

$$(6.9)\qquad \begin{aligned} X &= A, \\ Y &= B - b_1 X, \\ Z &= C - c_1 X - c_2 Y. \end{aligned}$$

In summary, we generate three uncorrelated random numbers A, B, C. Their variances are given by (6.8). Then we derive X, Y, Z from (6.9). The latter set of random numbers has an expected variance matrix of (6.1).

This technique can be used for any number of correlated variables. The equations for the unknown coefficients, corresponding to (6.5) and (6.7), become quite long and involved, but they fit a very regular and predictable pattern.

7. NUMERICAL SIMULATION

In this section, we outline the numerical simulation scheme to obtain quantitative results of the modeled insurance process. In this scheme, we conduct 200 trials. For a genuine analysis, at least 1000 should be done.

We want to project the study five years into the future. For each year and for each trial, we generate five random numbers for the five insurance ratios discussed in section 4. In the generation of the random ratios, we take into account the correlation coefficients in Table 6. For instance, the 200 loss ratios have an expected deviation of 4.09%, and the 200 dividend ratios have an expected deviation of 1.49%. Moreover, the 200 pairs of loss and dividend ratios have an expected correlation coefficient of .185. As we mentioned earlier, the loss ratios and earned premiums for the past ten accident years are considered non-stochastic, and shown in Table 1.

For each trial, we generate a set of three function parameters, for use in the lognormal cumulative distribution function, having the variances and covariances shown in section 5. This accounts for the parameter errors in the paid losses. For each incremental payment, we also generate the process error in (5.2).

For each trial, we substitute the simulated numbers into equations (2.1)-(2.9). Therefore, for each random variable at each time t, we have a series of 200 realized values. Then we can simply take the mean and deviation of these values, which represent the mean and deviation of the random variable.

8. NORMAL DISTRIBUTION OF NUMERICAL RESULTS

From the numerical simulation, we can obtain the estimate and deviation of any random variable. Ideally, we would want to approximate every random variable as being normally distributed, because then the percentiles for the variable can be readily estimated. In this section, we will use the chi square goodness-of-fit test to show that the variables are approximately normally distributed.

Among the numerical details, in this section we look only at the surplus. The following table gives the means and deviations of the surplus for the next five years. It also includes the probability of ruin, (defined as the insurer having negative surplus), the number of expected ruins, and the number of observed ruins, among the 200 trials.

Table 10: Numerical Results of the Surplus					
Year	Mean	Deviation	% ruin	Expected	Observed
0	6,667	0	-	-	-
1	10,455	830	0.0	0	0
2	15,071	1,871	0.0	0	0
3	20,199	4,372	0.0	0	0
4	26,356	8,595	0.1	.2	0
5	33,770	14,699	1.1	2.2	2

To establish the ruin probabilities in Table 10, we assume the distribution of the surplus to be normal with the given means and deviations. We then compute the probability that the surplus reach zero in any given year. The expected number of ruins is then the product of that probability and 200. The fact that the expected and observed values are very comparable indirectly validates the merit of our approach.

To establish percentiles, we can of course take the distribution found among 200 trials. But a more desirable and convenient way would be to establish that the simulated results are approximately normally distributed. We note that the assumption of normal distribution cannot be taken for granted, since, even though the simulated random numbers are assumed normally distributed, equations (2.1)-(2.9) contain products of normally distributed simulated numbers, which generally do not follow that distribution.

For a numerical example, we take the surplus of the fifth year, and see if the simulated results could be reasonably approximated as being normally distributed. We use the chi-square goodness-of-fit test to either validate or reject this assumption. For the fifth-year surplus of mean 33,770 and deviation 14,699, we divide the whole spectrum of $(-\infty, \infty)$ into ten intervals of equal probability. For instance, the second interval runs from 14,933 to 21,399, representing the 10th and 20th percentiles, respectively. If the distribution is normally distributed, 20, or 10%, of the outcomes would be expected to fall into this interval. The following table presents the observed and expected frequencies for our simulated set:

Table 11: The Chi Square Test for the Fifth-Year Surplus										
	1	2	3	4	5	6	7	8	9	10
Obs.	17	21	25	19	16	28	18	17	19	20
Exp.	20	20	20	20	20	20	20	20	20	20

The chi-square value is:

$$\chi^2 = \sum_{i=1}^{10} \frac{(O_i - E_i)^2}{E_i} = 6.50. \qquad (8.1)$$

χ^2 should follow the chi-square distribution with nine degrees of freedom, giving a probability of 69%. In other words, if the 200 simulated fifth-year surplus values are normally distributed, there is a 69% probability that their chi-square value would be greater than 6.50. Therefore, the normal-distribution assumption is accepted.

We use the chi-square test on many of the random variables, and, by and large, the assumption of normal distribution is reasonably satisfied.

9. Extensions of the Model

In conducting the study, we use historical data. In other words, we assume that our insurer would continue on the same trends as found in the past. But we can also use scenario or assumed data in the model. For instance, after looking at the probabilities of ruin in Table 11, management finds them too great, and decides on two simultaneous changes in operations. First, written premiums could be curtailed. Secondly, underwriting standards could be strengthened, so as to decrease the level and variability of the loss ratios. If the analyst can quantify these changes, they can be built into the model. The model can in turn quantify the degrees of the necessary changes, in order to decrease the probabilities of ruin to acceptable levels.

The model can be used as a tool of scenario testing. For instance, the analyst may discard the historical loss ratios, and decides a future loss ratio of 75% with a deviation of .05 is reasonable. He can then carry out the simulation and analysis with these scenario values.

For the study, we chose a situation as simple as possible. But the model offers great flexibility. As more and more operations are added to the analysis, the number of governing equations would increase. Below we list some of the many other operations that the model can readily accommodate.

Multi-lines insurer: We expect that as more lines are added, the financial results would stabilize. This is especially true if the loss ratios of the various lines have no or negative correlation.

Differing investment strategies: We can allocate the available investment assets into different segments, such as bonds, stocks, and real estate. We can also consider sub-segments within each major category: such as, taxable versus tax-exempt bonds. The model can give us an idea of the optimal investment strategy, given a corporate objective, such as growth versus stability.

Tax liabilities: This item can be readily built into the model.

Interest rates and inflation: These two affect the investment income and the loss payments. There are many theories concerning how inflation affects the stock and bond markets. Once the analyst decides to use a particular theory or model, it can be readily integrated into the framework of our DFA model. Regarding the loss payments, things are not so apparent. There are many ad-hoc techniques to account for inflation. But to our knowledge, there is no mathematically rigorous model that can explain how inflation affects insurance loss payments.

Reinsurance: Two aspects of this item may be considered. One is the default rate of the reinsurers. A default occurrence can be modeled as a Poisson process. Secondly, we can consider different reinsurance strategies, such as excess versus quota-share, and their effects on the balance sheet.

Catastrophes: If the insurer has much property exposures, we have to consider this aspect. An existing software package can be incorporated into this model.

Varying payment patterns: For the same line of business, the payment patterns for the different accident years may vary. We analyze this situation in our loss-reserving papers. For this study, we simplify, and elect not to account for the varying patterns.

SAP/GAAP bases: The model can be used in either basis. In the latter, unrealized capital gains, deferred acquisition costs, etc. have to be considered.

10. Conclusion

We have presented a dynamic financial analysis with two key ingredients: stochastic modeling of the individual operations and the error correlation of the operations in concert. One of its strengths lies in its use of the set of governing equations. This set can be contracted or expanded, depending on whether the actuary wants a simpler or more extensive analysis.

A user of dynamic financial analysis can evaluate the desirability of a strategy over another on several criteria: stability, profitability, and growth, among others. For stability, he should determine that the variability of the results and the probability of ruin are kept to acceptably low levels. For profitability, he should look at the overall income, which in our simplified example is:

$$(10.1) \qquad EP(t) + II(t) - IL(t) - PE(t) - PD(t).$$

In our example, we assume built-in growth. But we can certainly model it as a function of other variables, such as equation (3.2).

We note that the three aforementioned criteria are in many ways conflicting. But with dynamic financial analysis, the user has a better idea of where the best compromise lies, given the objectives and constraints of the company.