AN ECONOMETRIC MODEL OF PRIVATE PASSENGER LIABILITY UNDERWRITING RESULTS

Richard M. Jaeger and Christopher J. Wachter

Richard Jaeger is currently Actuarial Director in the Actuarial Development Department of Insurance Services Office. He is a Fellow in the Casualty Actuarial Society and a Member of the American Academy of Actuaries. He is serving on the CAS Syllabus and Examination Committees. He holds a BA degree in Mathematics from Boston College and an MS degree in Mathematics from the University of Notre Dame.

Christopher Wachter is currently Senior Actuarial Assistant in the Actuarial Development Department of Insurance Services Offices. His responsibilities include econometric modeling. He received his BS degree in Actuarial Science from Lebanon Valley College in 1983 and is currently pursuing associateship.

This paper presents an econometric model of private passenger liability underwriting results. The model, fitted on data from 1954 to 1983, is used to forecast results from 1984, 1985 and 1986. Premiums, losses, and expenses are modelled separately, with the loss model based on two sub-models (severity and traffic accidents). The paper covers the process of model building from initial *a priori* analysis, through forecasting. The paper also attempts to provide a general framework useful in the modeling of other lines.

PURPOSE

Our objective in undertaking the research presented here was to forecast industry combined ratios for private passenger automobile liability. While actuaries have always been concerned with trending, projecting and forecasting there is little in the actuarial literature on forecasting industry results. Some of the papers in the actuarial literature (Alff and Nikstad, James, Lommele and Sturgis) are listed in the Bibliography of this paper. Although such forecasting may be several steps removed from the actuary's day-to-day work, senior executives, insurance regulators and financial analysts are all interested in what the results will be. The actuary has the training and experience to help. A secondary objective of this paper is to indicate a general approach that can be used to model other lines of insurance as well.

ORGANIZATION OF THE PAPER

This paper follows a chronological format showing the development of the model from initial *a priori* analysis through forecasting. The presented models, therefore, appear after the section on the model fitting process and before the section on *a posteriori* analysis. The paper contains two appendices and a brief bibliography. The appendices contain a glossary of useful econometric terms, a list of data sources, and graphs displaying the fit of the presented models.

A PRIORI ANALYSIS

The importance of *a priori* analysis cannot be overemphasized. In an ideal, perfectly efficient world the researcher will analyze the situation to be modeled or forecasted, determine the relevant forces, select the appropriate variables, specify the form of the mathematical relationship, indicate the signs (and perhaps the magnitudes) of each variable in the relationship, and only then test the hypothesis against the data. In the real world one tries to follow this approach while coping with disappointing test results, new ideas that come to mind after the first results, and the nagging question - where can the model be improved?

It is important to use the data to test the *a priori* hypothesis, rather than to search for a model which fits the data well, and then derive an *a posteriori* "hypothesis" from the model. We want assurance that it is a good model, not just that a good fit results from much trial and error. We, of course, want the model to fit well *in addition to* agreeing with the *a priori* hypothesis.

This is especially important when there is limited data. Everyone is familiar with the inadvisability of explaining the variation in ten data points by using ten independent variables, or even six independent variables. The same effect can occur when the best-fitting model is chosen after testing too many sets of variables using several model forms, even if all of the variables and forms are reasonable.

There are ways to mitigate this problem. One way is to use part of the data for fitting and another part for testing. Any partition that is expected to yield the same model on the subsets could be used. Some possibilities are first and third quarters separate from second and fourth quarters, seasonally adjusted as necessary; a geographical partition, by state or region; and stock companies separate from mutual companies. Another way is *ex post* testing in which we try to forecast the latest points after fitting to the data excluding those points.

The first assurance, however, of having a good model comes from the model's agreement with _a priori_ analysis. _A priori_ analysis provides an intuitively reasonable explanation of the actual situation. Steps to improve the model should be governed by attempts to improve the _a priori_ analysis. This insures that the resultant model will be sound on a conceptual basis.

As stated earlier the purpose of the model is to forecast combined ratios. We decided early, however, not to model the combined ratios directly, but to model the losses and premiums separately and even to attempt to decompose losses and premiums into separate factors. We attempted to model losses as volume times frequency times severity and premiums as expected losses times a "pricing factor."

One reason for this approach is to reduce the problem to relatively bite-size pieces, each with a more manageable number of possible causal factors. Another reason is to make more efficient use of relatively few data points. Separating losses from premiums creates, in effect, twice as many points as using just the combined ratios.

The most important reason, however, for decomposition is to guide efforts to improve the model. If the premium model behaves better than the loss model, then attention can be directed to the loss model. If frequency is the loss factor showing the most unusual behavior, then frequency can be investigated before the other factors. A related reason is that it is easier to confirm whether a proposed improvement has the expected effect on the proper component.

One consideration in the _a priori_ analysis is that the model is primarily intended for forecasting rather than explanation of the changes in the historical period. The independent variables selected should be easily forecastable or already forecasted in a satisfactory way.

Losses

The initially selected form of decomposition for incurred losses was volume times frequency times severity. At this point there were two ways of proceeding. One way was to seek sources of standard insurance data for each element in the decomposition, such as, earned car-years for volume, and incurred or paid claim frequency for frequency. Each element could then be modeled separately. The other way was to build a precise decomposition from a reasonable starting point. An example of such a decomposition is to start with the number of registered vehicles (VRCAR) for volume and then, using the number of traffic accidents (TRAFACC), define frequency as TRAFACC/VRCAR and severity as incurred losses/TRAFACC. Both VRCAR and TRAFACC are forecasted by Data Resources, Inc. (DRI). Severity would have to be modeled.

The advantages of the first way arise from the fact that the elements in the decomposition are standard insurance concepts.

- Prior knowledge of these concepts can be applied directly to the analysis. If the elements have already been modeled, then much of the work is already done.
- If there are strong judgmental reasons to expect particular changes in the elements, these judgmental values can be used directly in the model to obtain a forecast.

The advantages of the second way are:

The decomposition is precise, that is, the product of the factors exactly equals the variable of interest. There is no need to adjust the product for such differences as absolute scale (because of using a subset of industry data) or paid rather than incurred data.

If there is a reasonable starting point that is already modeled or forecasted, then part of the work is already done.

We chose the second approach to decomposition because of the above advantages plus a very practical additional advantage. We wanted to have at least 20 years of data for modeling, so that several underwriting cycles and a variety of economic conditions would be present in the data. TRAFACC was available back to 1950 and VRCAR even farther back. Insurance time series for a decomposition would have been more difficult to obtain for a comparable length of time.

This second approach still leaves open the possibility of comparing elements of the decomposition to available insurance time series for reasonableness.

Early work with the decomposition based on VRCAR and TRAFACC led to the conclusion that TRAFACC showed very anomalous behavior, described more fully later in the paper.

Frequency

A number of factors were identified that might influence frequency. The principle of simplicity and the 80-20 rule were applied. (Keep it simple, and 80% of the effect comes from 20% of the causes.) Factors expected to have considerable effect were demographic shifts (notably changes in the proportion of young drivers), changes in the incidence of reporting traffic accidents (as repair costs go up while reporting thresholds tend to remain fixed), and energy crises (gas shortages). Selection of the first factor was influenced by work that had been done on other automobile insurance frequency data. The ratio of population aged 16-28 to the number of vehicle drivers licenses was selected to represent the demographic shift. A CPI-based measure of automobile repair costs, BODYWORK, was selected as a variable corresponding to the second factor. Eventually, it was decided to represent energy crises by the variable vehicle miles traveled (VMTCAR), and to recast the decomposition as losses equals accidents times severity, where TRAFACC is modeled as VMTCAR times some factor. This is equivalent to substituting VMTCAR for VRCAR in the initial loss decomposition.

Severity

Inflation should be the major force driving the loss severity model. Loss severity is a combination of bodily injury and property damage severities. To represent inflation we created an index that is a weighted sum of various CPI component indices expected to be related to automobile liability severity. The weights were judgmentally selected in the *a priori* stage, with the expectation that the exact weighting would not be critical.

Social inflation, as distinct from the purely economic underlying inflation, may also be a factor. We expected that social inflation would be reflected in the fitted quantitative relationship between severity and economic inflation, and therefore did not represent it by a separate variable.

Small cars are likely to provide less protection to occupants in an accident. They may also tend to be damaged more severely. To represent the proportion of small cars to total cars, we selected the ratio of foreign new car sales to total new cars sales. We realized, however, that a possible future refinement would use this ratio averaged over several years.

It is possible that the introduction and subsequent modification of no-fault laws affected severity. We did not expect a strong effect, however, and did not attempt to represent this factor in the model.

Premiums

Premiums may be influenced by a large number of factors. Some of the factors are exposure volume, inflation, recent loss experience, recent profitability, competition, supply and demand, capacity, investment yields and the future expected value of several of these factors. We finally selected several variables that represent reasonably distinct factors. The number of vehicle registrations was selected to represent exposure volume. The GNP deflator was selected to represent inflation. The incurred losses of the current year and the prior year were selected to represent the ratemaking process, recent profitability, and management's expectation of future losses. Real surplus (using the GNP deflator) was selected to represent supply and capacity and as a proxy for competition. These variables should be positively related to premiums, except for real surplus. A high real surplus should have a downward effect on premiums due to over-capacity and consequent increased competition.

We intended to model written premiums as above and to produce earned premiums by a simple linear model using the current year and prior year written premiums.

Expenses

The expense ratio should be inversely related to deflated written premium, since there are fixed expenses which do not vary with written premiums. We used written premiums deflated by the GNP deflator to model the expense ratio and obtained a reasonable model.

Model Form

The final stage in *a priori* analysis is selection of the model form. The selection of model form is significant, but is somewhat less important than the earlier stages of *a priori* analysis. We selected a logarithmic form for the written premium and severity models for the following reasons:

1. We expected a multiplicative relationship between the component independent variables.

2. The coefficients are elasticities, rather than absolute magnitudes. The effect of a 1% change in an independent variable is the coefficient times 1%. Thus, the relative contribution of each component variable can be easily determined.

3. Inflation-sensitive time-series are transformed from an exponential form to a linear form.

4. Heteroscedasticity is minimized, since inflation will not cause residuals to grow as large with time if a logarithmic form is used.

5. The fit is more robust, since outliers tend to have less of an effect on parameter estimation.

This completes the initial *a priori* analysis and prepares us for fitting and testing actual models.

THE MODEL FITTING PROCESS

The first stage in the model fitting process is the selection of data. Ideally, data should come from recognized and reliable sources, and be available for a significant length of time. We used insurance premium, loss, expense and surplus data from A.M. Best's *Aggregates and Averages*, and various economic time series from DRI's data banks.

The second stage in the model fitting process was to splice data together. Prior to 1974, for example, auto liability was not split into private passenger liability and commercial automobile liability. We applied a splicing factor of 0.89 to the pre-1974 automobile liability data (stock and mutual only) to extend backwards the 1974-83 private passenger liability data (stock, mutual and reciprocal). The splicing factor was based on the observed ratios from 1974 to 1983 which were very stable and averaged 0.89. The advantage to splicing is that we are able to use 30 years of data (1954-83) rather than only 10 years (1974-83).

The third step in our model fitting was to handle a problem variable, TRAFACC. TRAFACC is a time-series which measures reported traffic accidents. It is also a variable whose definition has changed over the historical period. Prior to 1968, the ratio of TRAFACC to highway fatalities is remarkably stable, indicating that TRAFACC for that period may have been defined by multiplying highway fatalities by a constant. We decided, therefore, to model TRAFACC over the period 1968 to 1983, and use the fitted values produced by the model over the full period 1954 to 1983 in place of the original TRAFACC series. We later decided, when a reasonable severity model had been fitted, to create a final series representing traffic accidents called TRAFACC', because of the problems noted above with TRAFACC. We felt that a reasonable model for severity would be better than incurred losses divided by fitted TRAFACC, and therefore would probably create a more accurate decomposition.

The fourth stage in the model fitting process was to decompose the incurred losses into severity and traffic accidents. We defined severity as incurred losses divided by fitted TRAFACC (described in the previous paragraph). We selected a deflator judgmentally, and chose a variable to proxy for small cars (the proportion of imported car sales to total car sales) to model severity. The resulting model fit reasonably well, and we therefore tentatively accepted the severity model.

The fifth stage in our model fitting process was to use the fitted values generated by the severity model to create TRAFACC'. We defined TRAFACC' as incurred losses divided by fitted severity. We modeled TRAFACC' using the same variables that we had used to model TRAFACC, and decided tentatively to accept the traffic accident model. Multiplying the two models (severity and TRAFACC') we were able to construct our incurred loss model.

After fitting the loss series, we fixed our attention on the premiums. We selected several variables based on our a priori analysis corresponding to incurred losses, capacity, volume, inflation, investment yields and recent profitability. We looked at the relationship between changes in written premiums and changes in potential explanatory variables at various lags to determine a reasonable lag structure. We determined, for example, that change in surplus lagged two years was related to change in written premiums more strongly than if it were lagged 0, 1, 3, or 4 years.

We fitted a preliminary model, and decided to discard some variables which did not appear significant (had t-statistics of less than 2). We fitted a second model, checked for and corrected for multicollinearity (the independent variables should behave independently of each other, and not show strong correlation), and tentatively accepted the premium model.

After tentatively accepting the premium and loss models, we tested the models to determine if they were acceptable. There were four final tests:

1. We determined that the final models agreed with our a priori analysis. Specifically, we checked the coefficient of every independent variable and confirmed that each coefficient had the expected sign and had reasonable magnitude.

2. We determined that the model's error was acceptable (standard errors of 2.3% for premiums and 2.9% for losses) and that each dependent variable was statistically significant at the 5% level (as determined by the t-statistic).

3. We performed an ex-post test. An ex-post test involves fitting the model over a shorter historical period (we used 1954-80) and then "forecasting" the latest values (1981-83) using actual values for the independent (input) variables. The ex-post forecast errors were deemed acceptable, and are summarized below:

EX-POST ERRORS*

Year	Premium Model	Loss Model
1981	-2.85%	+1.72%
1982	+1.92%	-0.17%
1983	+0.66%	+2.79%

* Error = (Actual-Forecasted)/Actual

4. We analyzed the residuals (errors) of the models for violations of least square assumptions.

Potential Violation	Correction needed (if any)
1. There are outliers.	A dummy variable was incorporated into the severity model for a 1974 outlier.
2. The residuals are correlated.	Autocorrelation corrections were implemented as deemed appropriate.
3. The variance of the residuals is not constant. (heteroscedasticity)	The residuals were examined, and no indication of heteroscedasticity was found.
4. The independent variables are strongly correlated. (multicollinearity)	Multicolliearity was found in an early premium model and corrected for. Whatever multicollinearity remains does not appear substantial based on the observed correlations between the independent variables.
5. The relationship between dependent and independent variables is unstable.	Ex-post testing shows stable parameter estimation. When the models were refitted over the period 1954-80, the parameters did not substantially change from the model fitted over 1954-83.

On the basis of these four tests, we decided to accept the premium and loss models.

THE MODELS

The private passenger liability models are as follows:

1. Written premiums

$$\text{Log }(NPW_t) = 0.664 \text{ Log }((A_t + A_{t-1})/2) + 1.315 \text{ Log }(VRCAR_t)$$
$$- .115 \text{ Log }(Surplus_{t-2}/PGNP_{t-2}) + .884 \text{ Log }(PGNP_t) + .573 \text{ AR1}$$

Where:

NPW	=	Net premiums written
VRCAR	=	Vehicle Registrations (in thousands)
PGNP	=	GNP deflator (1972 = 1.000)
Surplus	=	Surplus (in thousands)
A	=	Incurred Losses/ (VRCAR x PGNP)
AR1	=	Autoregressive term of order 1.

t-statistics are 6.77, 18.08, -3.02, 27.24 and 3.27, respectively, for the five coefficients.

The model has a normalized standard error of 2.3%.

2. Incurred Losses

Incurred Losses equal TRAFACC' times severity

$$TRAFACC' = TRAFACC_{78} \cdot VMTCAR_t \left(.4575 + .4933 \frac{YOUTH_t}{VDL_t} + .0812\ BODYWORK_t\right)$$

Where:

TRAFACC$_{78}$ = Number of traffic accidents in 1978 (original TRAFACC series)

VMTCAR = Vehicle miles traveled by cars

YOUTH = Population aged 16-28

VDL = Vehicle Drivers Licenses

BODYWORK = CPI for auto bodywork (prior to 1978, CPI for auto repair and maintenance)

Note: all variables are normalized to 1.000 in 1978 except for TRAFACC' and TRAFACC$_{78}$.

t-statistics are 3.131, 3.301 and 5.301 for the three coefficients.

$$\text{Log}(SEVERITY_t) = .7321\ \text{Log}(DEFLATOR_t) + .1008\ \text{Log}(PROP_t) - .0955\ Dummy_t + 6.3108 + .9419\ AR1$$

Where:

DEFLATOR = .35 MEDCARE + .35 WAGE + .20 BODYWORK + .10 PC

MEDCARE = .67 CPIU for Hospitals (CPIU for Hospital Rooms before 1978) + .28 CPIU for Physician's Services + .05 CPIU for Medical Commodities. (All components indexed to 1.000 in 1978)

WAGE = Average Hourly Earning Index for Production Workers (1978 = 1.000)

PC = Implicit Price Deflator for Personal Consumption Expenditures (1978 = 1.000)

PROP = Foreign New Car Sales/Total New Car Sales

DUMMY = 1 in 1974, 0 otherwise

t-statistics are 8.96, 3.94, - 3.70, 64.34 and 12.20 for the five coefficients.

The incurred loss model has a normalized standard error of 2.9%.

3. Expense Ratio and Earned Premium

$$EP_t = .668\ WP_t + .336\ WP_{t-1}$$

$$\text{Log}(ER_t) = 2.458 - .240\ \text{Log}(NPW_t/PGNP_t) + 1.096AR1 - .690AR2$$

A POSTERIORI ANALYSIS

After the model fitting stage is completed, the *a posteriori* analysis stage begins. The function of *a posteriori* analysis is to examine the accepted model and attempt to explain any unusual features of the model. This is useful, because possible refinements to the model are identified for future research. It is important to realize that model building is an on-going process and that models should be monitored and updated as additional data becomes available.

We have identified three unusual features in our model:

1. The elasticity of VRCAR, a volume measure, is greater than one in the premium model.

2. The elasticity of DEFLATOR, an inflation measure, is less than one in the loss model.

3. The lag structure of the incurred losses in the premium model is shorter than might be expected.

Possible explanations for these features (which represent deviations from what might be expected) are:

1. A rise in the proportion of insured vehicles over the historical period (1954-83) would impact the elasticity of VRCAR.

2. The "real" severity may be declining somewhat due to safer automobiles and roads.

3. The incurred loss term may combine two separate components: expected losses (IL_t) and "fast-track" experience (IL_{t-1}). If this is, in fact, the case, a longer distributed lag structure (using IL_{t-2} and even further back) for the "expected loss" component may be more appropriate.

This *a posteriori* analysis could serve as an input to the *a priori* analysis stage of future model-building efforts. We feel that the present models are sound and useful for forecasting, but that the *a posteriori* analysis indicates some areas for future research.

FORECASTS

After the analysis, model fitting and testing, the model can be used. The model has two main applications: explanation and forecasting. To the extent the model explains the mechanisms underlying industry written premiums and incurred losses, alternative "what-if" scenarios can be devised and forecasts made for these scenarios. Three potentially interesting scenarios are:

1. The banks enter the insurance industry injecting significant amounts of capital.

2. The campaign against drunk driving significantly reduces accident frequency.

3. Inflation surges upward again.

The actual scenario design and forecasts based on alternative scenarios are beyond the scope of this paper. Our forecasts, based on insurance data through 1983 and on DRI control scenario forecasts (using the July 1984 forecast) are as follows:

	1984	1985	1986
Written Premiums	24,763,224	26,048,890	28,039,674
Earned Premiums	24,391,568	25,721,102	27,482,929
Incurred Losses	21,080,953	22,606,582	24,082,622
Loss Ratio	0.864	0.875	0.876
Expense Ratio	0.246	0.239	0.233
Dividend Ratio (selected)	0.01	0.01	0.01
Combined Ratio	1.120	1.124	1.119

FUTURE RESEARCH

The research to develop these models has raised some questions for further investigation. Potential topics for research include:

1. Incorporating investment yield into the current premium model. There are many investment yield statistics, and also a variety of time frames to select (current yield, recent yield, expected yield, and embedded yield). In addition, it is possible that investment yield may be significant over a small subset of the historical period.

2. Incorporating changes in the proportion of insured motorists in the total driving population. The chief problem is to locate a source of data over the historical period.

3. Incorporating a variable representing increased safety of roads and automobiles. The major task is to find a valid time series which can proxy

for automobile safety, and which is available over the entire historical period.

4. Selecting other measures to proxy for industry price competition.

The authors will, as time permits, research these areas further, and would welcome the insights, suggestions and research of other people involved in this area of actuarial/econometric research.

APPENDIX A - GLOSSARY

Autocorrelation:	The correlation between residuals and the residuals lagged a certain number of periods, called the order. An assumption of least squares regression is that autocorrelation is not present.
Autoregressive Term:	A term (in a model equation) used to correct for autocorrelation when the analysis of the residuals of a model indicate the presence of autocorrelation. In this paper ARi denotes an autoregressive term of order i.
Decomposition:	The breaking of a problem into smaller, more easily handled problems. The solutions to the small problems are combined to form a solution to the overall problem.
Dummy Variable:	A variable that takes on two values, 0 and 1. The dummy variable is used to account for abnormal real world conditions (energy crises, price controls, wars, etc.) or to remove the effects of obvious outliers.
Elasticity:	A measure of the relationship between two variables.
Heteroscedasticity:	Heteroscedasticity exists when the variance of model's residuals is not constant over the entire range of data. Least squares regression assumes heteroscedasticity does not exist.
Lag:	The length of time between the effect on an independent variable and the effect on the dependent variable. If several lags of an independent variable are combined into are term, we say that the term represents a distributed lag structure.
Multicollinearity:	The degree of correlation between the independent variables. Least squares regression assumes that the independent variables are independent of each other.
Normalized Standard Error:	The standard deviation of the error of a model expressed as a proportion of the dependent variable.
Outlier:	A data point that is questionable due to an abnormally large deviation from its expected value. Outliers bias regression results, sometimes quite substantially.
Proxy:	A variable used as a measure for something that is not readily quantifiable.
Residual:	The difference between an actual observation and the expected value of that observation based upon model.
Robustness:	The degree to which a model is stable and unresponsive to outliers.

Splicing: The combination of two similar time series covering differing time periods into a unified series.

T-statistic: The ratio of a coefficient to the standard error of that coefficient. Generally a T-statistic with absolute value greater than 2 indicates a significant relationship between an independent variable and the dependent variable.

APPENDIX B - Sources of Data

Insurance data was obtained from A.M. Best's Aggregates and Averages.

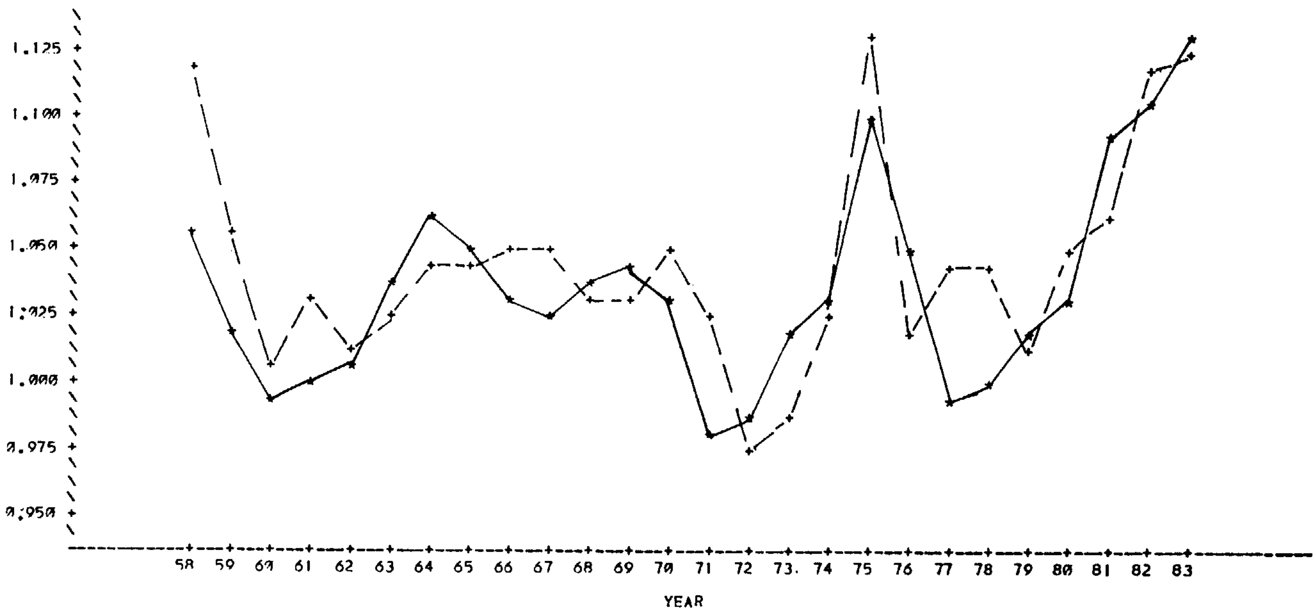
The following time series were obtained from stock, mutual and reciprocal companies combined:

Net premiums written, Net premiums earned, Incurred losses, Expense ratio, Surplus.

Economic data and forecasts were obtained from Data Resources, Inc. The following time series were obtained:

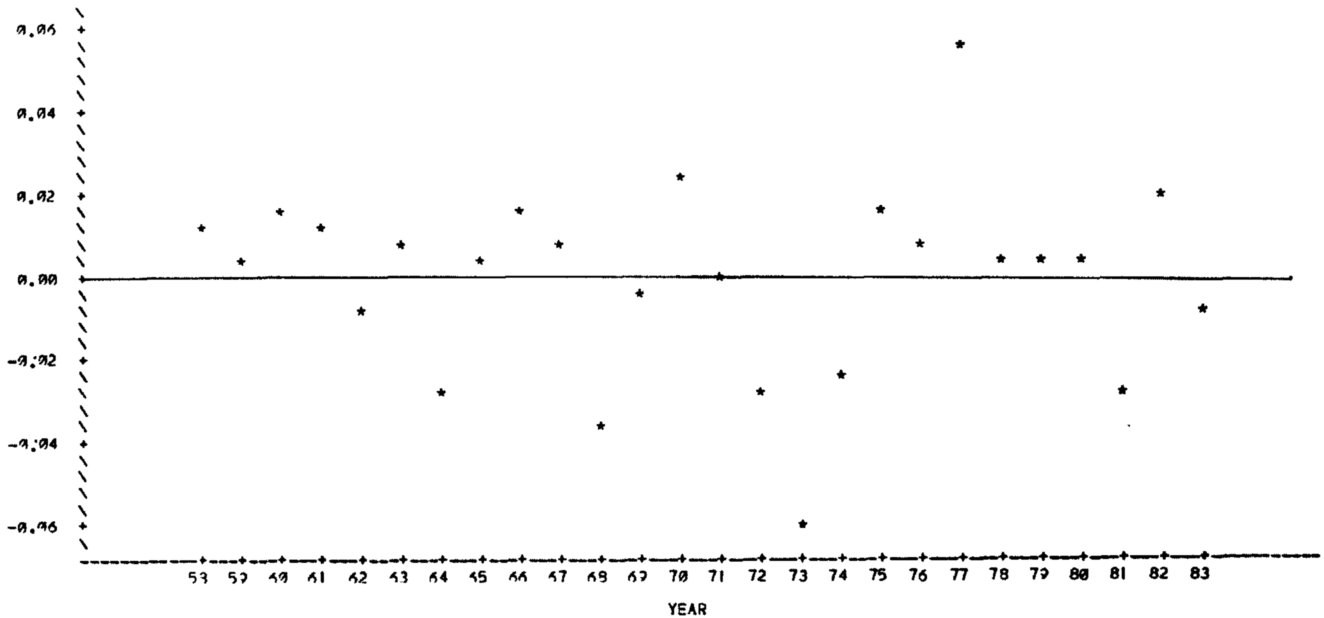
Primary source:	Bureau of the Census Population aged 16 through 28
Primary source:	Bureau of Economic Analysis, Department of Commerce Gross national product deflator, Personal consumption deflator, Retail sales, imported passenger cars, Retail sales, new cars.
Primary source:	Bureau of Labor Statistics, Department of Labor Index of average hourly earnings of non-farm production workers, Consumer Price Indices: Auto bodywork, Auto repair and maintenance, hospital and other medical services, hospital room, medical commodities, physicians services.
Primary source:	Federal Highway Administration, Department of Transportation Vehicle driver licenses (estimated), vehicle miles traveled-passenger cars, vehicle registrations - automobiles.
Primary source:	Insurance Information Institute Traffic accidents.

FITTED VS. ACTUAL COMBINED RATIOS



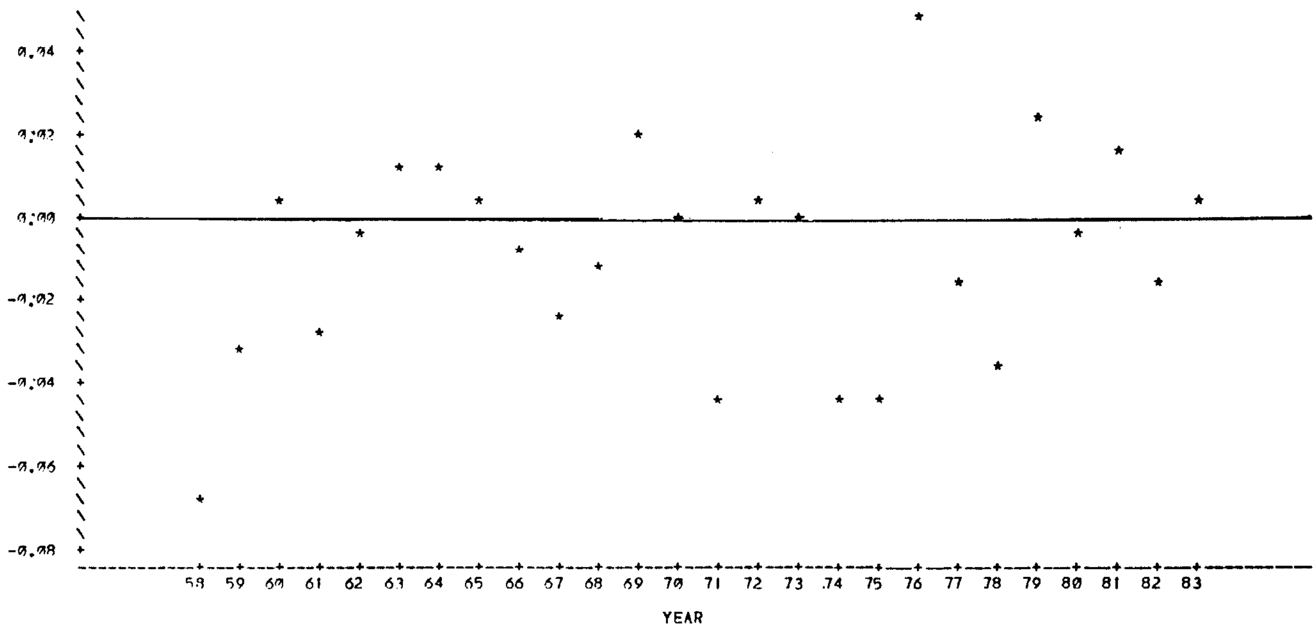
NOTE: Solid lines connect actual combined ratios,
Dashed lines connect fitted combined ratios.

RELATIVE ERRORS - PREMIUM MODEL



$$\text{Relative Error} = \frac{\text{Actual Minus Fitted}}{\text{Actual}}$$

RELATIVE ERRORS-LOSS MODEL



$$\text{Relative Error} = \frac{\text{Actual Minus Fitted}}{\text{Actual}}$$

BIBLIOGRAPHY

Alff, Gregory N. and Nikstad, James R. "A Model of General Liability Net Written Premiums," PCAS LXIX, pp. 30-45.

Armstrong, J. Scott. Long Range Forecasting, New York: John Wiley & Sons, 1978.

James, Kaye D. "Underwriting Cycles in the Property-Casualty Insurance Industry," Inflation Implications For Property-Casualty Insurance, 1981 CAS Discussion Paper Program

Lomele, Jan A. and Sturgis, Robert W. "An Econometric Model of Workmen's Compensation," PCAS LXI, pp. 170-189.