Student Project Design on Claim Frequency and Severity

Several candidates have done independent student projects on actuarial data sets from their current jobs: life insurance medical underwriting, health insurance disability rates, and property-casualty claim costs. Using data from your actuarial work may make the student project quicker and more useful to your company.

Claim frequency and severity are particularly good data for student projects. You can do many innovative statistical projects. If you examine the data well, your work will be valuable to your company, in addition to fulfilling the VEE requirements. You learn much about pricing insurance policies, and you see the use of regression analysis and time series for actuarial work.

Potential student projects are discussed in various postings on the discussion boards. This synopsis brings together several topics and outlines project designs for claim frequency and severity.

This posting mentions personal auto frequency and severity. The line of business is not restrictive; similar projects can be done on other lines of business or practice areas.

Objective

The objective is to improve your company's current procedures for projecting frequency and severity trends. Your student project should include a proposed out-of-sample test to compare your company's current procedures with the regression models or ARIMA models in your analysis. Readers of your report may not be persuaded unless you show that you can forecast more accurately.

Current Procedure

Check how your company now projects claim frequency and severity trends. Many insurers use simple procedures that you can improve. Potential topics are:

- Internal trends vs external indices
- State vs countrywide trends
- Company vs industry trends
- Simple vs stochastic trend models
- Seasonality

Internal Trends vs External Indices

Some insurers use internal claim severity or claim frequency numbers, which they fit to an exponential curve. Your student project may other ways to project claim costs.

- *Regression analysis:* Relate claim severity or claim frequency to inflation indices or

other macroeconomic variables (such as gas prices and GDP).
- *Time series:* Form an ARIMA process to model claim severity or frequency.

Some insurers have extensively analyzed their claim severity and frequency trends. If your employer has analyzed its claim severity and claim frequency trends with regression and time series analyses, it probably has several data sets or time series that you can use for the student project, and you may get additional ideas for new analysis from the work already done.

Picking the inflation index the regression analysis is not always easy. The CPI comes in several flavors: medical, wage, urban, rural, seasonally adjusted or not seasonally adjusted, and dozens of sub-varieties. Rely on intuition and the fit of separate regressions.

- Use medical inflation for workers' compensation medical benefits and wage inflation for workers' compensation indemnity benefits.
- Use a combination of medical inflation and overall CPI for personal auto bodily injury and a combination of medical inflation and wage inflation for personal auto no-fault.
- You may try several combinations of inflation indices and see which explanatory variables give the highest adjusted R^2.

Do not use two inflation indices as explanatory variables in a single regression. Some candidates think: "Let me use both wage inflation and medical inflation as explanatory variables. If the ordinary least squares estimators are 40% for wage inflation and 60% for medical inflation, I can use a 40% - 60% weighting of these two inflation indices to project claim severity trends.

The high multicollinearity distorts the ordinary least squares estimators. Random fluctuations in the inflation indices may cause large changes in the ordinary least squares estimators. The betas are not good indicators of the proper weight for each inflation index.

COMPANY VS INDUSTRY TRENDS

Industry claim frequency and severity trends are available from rating bureaus and other statistical agencies. Personal auto insurance has Fast Track industry data. With your own company's claim severity time series, Fast Track industry severities, and CPI inflation indices, you can do several student projects.

Industry trends differ from company trends for three reasons, with different implications for forecasting.

- Reason 1: The books of business differ and the trends differ.
- Reason 2: The true trends do not differ, but the most recent severity or frequency figure is distorted by random loss fluctuations.
- Reason 3: The true trends do not differ, and the most recent severity and frequency figures are not distorted by random loss fluctuations, but previous figures are distorted.

All three of these reasons may contribute to the observed differences. Your student project determines the optimal forecasting method for future claim frequency and severity.

Many pricing actuaries give too much weight to internal (company) loss cost trends in comparison to external (industry) trends. These weights are often justified as "Our underwriting is stricter than that of other insurers, so we have lower loss cost trends." Strict underwriting does not necessarily cause lower loss cost trends, and most insurers say they have stricter underwriting.

If you say: "the industry loss cost trend is 8% per annum," your manager might say: "Yes, but what do our data show?" if the company's own loss cost trend differs, the manager might give full weight to the company's trend. The optimal weights of company vs industry trend depend on many items. Your student project uses the techniques in the on-line courses to improve the trend estimates.

The techniques depend on the course for which you do the student project. You can use ARIMA processes, regression analyses, and structural models.

Set up three models to forecast your insurer's future claim severity or frequency.

(1) One model uses your own company's claim severity data. Some insurers use a simple random walk. They fit their past claim severities to an exponential curve and apply the trend rate to the experience period severities. Some insurers uses four quarter moving averages; others use seasonally adjusted quarterly figures.

For the student project, you may form two other types of models:

- An ARIMA process, using past claim severities or frequencies. You might take logarithms, first differences, or second differences to form a stationary time series.
- A structural model, using claim severities regressed on an inflation index or claim frequencies regressed on (deflated) gasoline prices or real GDP.

Some actuaries use separate models from the average date of loss to the end of the inflation index and from the end of the inflation index to the average date of loss under the new rates.

For the first period (the historical period), use the change in the inflation index, adjusted by the parameters in the regression model. For example, if the regression model says that the claim severity trend is 20% greater than the inflation rate, and the inflation index increases 15% from the average date of loss in the experience period to the end of the inflation index, the average claim severity should increase 15% × 1.2 = 18%.

For the second period (the forecast period), you can use econometric forecasts of inflation or you can fit an ARIMA process to the inflation index. After forecasting inflation, you use your regression model to forecast average claim severity.

(2) A second model uses industry Fast Track data. Use the same type of models, though the parameters may differ. You may use an F test to see if the different parameters are just random fluctuation or a truly different model.

The important test is which model has the better out-of-sample goodness-of-fit for your own company's average claim severities. It may be that the internal model has the better in-sample goodness-of-fit and the external data has the better out-of-sample goodness-of-fit.

(3) A third model uses the residuals of your company's claim severity data regressed on industry Fast Track data. This model may be especially good at offsetting random loss fluctuations in your own data. Your hypothesis might be the following:

> If the industry claim severity trend from 2007 to 2008 is 8% and my company's trend is 10%, I presume the higher trend for my company reflects random loss fluctuations, such as more than expected severe claims in 2008. My insurer's 2008 average claim severity is over-stated by two percentage points, so its 2008 to 2009 claim severity trend should be two percentage points below the industry average. If we project the industry trend to be 8%, we should project the company's trend to be 6%.

This is moving average MA(1) model, which you can easily test. Use both in-sample and out-of-sample goodness-of-fit tests.

In truth, we expect the moving average parameter to be no higher than 50%. The random loss fluctuations may be in 2007 or 2008 or a mix of both.

An autoregressive AR(1) parameter of 100% says: "If the company's trend from 2007 to 2008 is two percentage points above the industry's, we expect the same two percentage point difference next year as well."

The optimal ARIMA process depends on the insurer's size and the stochasticity of the claim severities. A small insurer looking at uninsured motorist claims might have an MA(1) parameter of 50%. A large insurer looking at collision claims might have an MA(1) parameter of zero and a significant AR(1) parameter. You might compare two coverages or two lines of business.

For your ARIMA model, make sure you have a stationary process. For claim severity, take logarithms and first differences. For the student project, examine the graph of the data and the correlogram, and explain why logarithms and first differences are needed.

Your conclusions have strong implications for insurance pricing. Your company's current procedure may indicate a trend factor of 10% and your model indicates a trend factor of 15% (or vice versa). A 5% change in rate levels affects market share, profitability, and corporate strategy.

If your student project suggests that the current trend procedure is not optimal and your

proposed model does better, spend extra time on the text write-up. The ARIMA process fit to the residuals may show the value of a moving average or autoregressive model, but readers of your report want to understand why the model makes sense.

Some pricing actuaries point to differing past claim severity trends between their company and the industry to justify an internal model. A difference in observed trends *proves* that industry figures are not applicable to the company's business. This argument sounds reasonable, but it is often misleading. The on-line courses give you the tools to test this justification. Test whether the industry trend vs your company's trend is statistically significant. For many companies, the differences are not statistically significant, and your actuarial pricing is more accurate if you use a larger data base. Even if the difference is significant, you must test if the differences have positive or negative autocorrelations.

State vs Countrywide trends

You can compare the models for different states, coverages, and time periods. Use a variety of states, coverages, and time periods to see if your results are consistent or are just random loss fluctuations.

An examination of state vs countrywide trends is similar to insurer vs industry trends. Use the same statistical procedures and hypotheses. Even states with different compensation systems, such as no-fault vs tort liability, may have the same claim severity trends. Test this empirically using the methods in the on-line courses.

Trends by class group are like trends by state. Examine each class group relative to the state (or countrywide) total.

Trends by coverage may differ, depending on the type of injury, such as bodily injury vs property damage.

Claim Frequency

Claim frequency has similar themes. Industry data may show a clearer relation to other macroeconomic variables, such as the cost of gas or weather conditions. You might use industry data to select autoregressive parameters and the residuals of your company's data on the industry data to select a moving average parameter. It is harder to fit claim frequency to an ARIMA process, and improvements in trend procedures are valuable.

Seasonality

If you have quarterly data, you can use quarterly claim frequency and severity figures. Claim frequency and severity are often seasonal. You can de-seasonalize the data or you can use an AR(4) term. Examine the effects of the alternative methods and select the one that seems best.

Out-of-Sample Tests

Examine several models, such as ARIMA(1,1,0), ARIMA(2,1,0), ARIMA(0,1,1), and ARIMA(1,1,1). Use the in-sample goodness-of-fit tests to pick the optimal ARIMA process. If two models seem equally good, use out-of-sample tests to choose between them. Often a more complex model has a better in-sample fit, but a simpler model has better out-of-sample forecasts.

Set up an out-of-sample test between your ARIMA model and your company's current ratemaking method. Make predictions for the next four quarters from each model, show them to your manager, and see which model forecasts better over the next year. It is likely – but not certain – that your model will do better.

In many cases (including claim frequency and severity projections), the ARIMA processes are better applied to the residuals. For a structural model, determine the optimal inflation index to use for claim severity and the optimal macroeconomic variables to use for claim frequency.

Develop a structural model, using inflation forecasts published by consulting firms. If you don't have access to these forecasts, use your own forecasts of inflation by ARIMA modeling on the inflation indices. Regress your claim severity time series on the inflation index and apply an ARIMA process to the residuals.

Set up an out-of-sample test for the next four quarters. If your model does better than the company's current procedures, you will get immediate credit (and perhaps a raise). If your model does not do better, you lose nothing. You will have received VEE credit and you learned much about claim frequency and severity.

Don't wait four quarters to write up your student project. Send in your student project when you complete the modeling. A year later, you can write back to NEAS reporting the results of the comparison between your model and the company's current procedures.

READINGS

The following papers may give you more ideas for a student project on claim severities.

Masterson, Norton E., "Economic Factors in Liability and Property Insurance Claims Costs," ASTIN Bulletin International Actuarial Association - Brussels, Belgium, 1977: Volume 9, No 3, pages 278-280;
http://www.casact.org/library/astin/vol9no3/278.pdf

Masterson, Norton E., "Economic Factors in Liability and Property Insurance Claims Costs, 1935-1967," Proceedings of the Casualty Actuarial Society Casualty Actuarial Society - Arlington, Virginia, 1968: Volume LV, pages 61-89;
http://www.casact.org/pubs/proceed/proceed68/68061.pdf

Evans, Jonathan P., and Schmid, Frank, "Forecasting Workers Compensation Severities and Frequency Using the Kalman Filter," Casualty Actuarial Society Forum Casualty

Actuarial Society - Arlington, Virginia, 2007: Winter, pages 43-66
http://www.casact.org/pubs/forum/07wforum/07w49.pdf