PROJECT TEMPLATE: SIMULATION AND FORECASTING

(The attached PDF file has better formatting.)

Take heed: You can do a student project on forecasting. Focus on the following:

- The simulation parameters vs the ordinary least squares estimators.
- Standard errors and confidence intervals for the regression forecasts.
- The skewness of the lognormal distribution: comparing means and medians.
- Estimating dollars of loss from regression forecasts of their logarithms.
- Confidence intervals for dollars of loss

This PDF file explains several concepts you should keep in mind. The effects of skewness seem surprising at first, but they occur in many actuarial tasks.

{Distinguish the simulation parameters from the ordinary least squares estimators: one is a scalar (but unknown) and the other is a random variable.}

Jacob: Let me see if I understand the sequence of steps in the project template.

- We assume a geometric decay (development period) and inflation rate (calendar year).
- We simulate logarithms of paid losses with the assumed decay and inflation.
- We use the simulated values to derive the ordinary least squares estimators.

Won't the ordinary least squares estimators be the same as the simulation assumptions?

Rachel: They differ for several reasons:

- The simulation is stochastic. As σ increases, the ordinary least squares estimators have greater variance and may differ from the simulation assumptions.
- The geometric decay and the inflation rate may not be constant. Even if $\sigma = 0$, the constant ordinary least squares estimator does not equal the changing parameter.
- The student project shows how we test for stability of parameters, determine how parameters are changing, and revise the regression equation.
- For the project template on parameter stability, higher stochasticity obscures the residual plots and makes it hard to estimate the inflation rates and geometric decay.

For the student project, you select regression assumptions, run the regression analyses, formulate hypotheses and test them, make forecasts, and derive conclusions.

In practice, we are given data, which reflect all the changes and uncertainties in the real world. We formulate and test hypotheses, but we never know if we are correct.

The simulation keeps the student project clear. We know the true parameters, but the scenarios (non-constant coefficients or heteroscedastic data) make it hard to estimate them. The student project shows:

- ~ How we detect the problem. Residual plots show the problem graphically.
- ~ How to correct the problem with dummy variables or squares of independent variables.

LOGARITHMS AND DOLLARS

{Many candidates assume that if we forecast K = the logarithm of paid losses, then the forecast of the dollars of paid losses = e^K. This is not correct.}

Jacob: What are the simulated Y values? Do we simulate dollars of paid loss?

Rachel: Real-life reserving uses dollars of paid loss. The student project uses logarithms of the paid losses, so that the relations are linear. If we are given dollars of paid losses Z, the logarithm of these dollars is *ln*(Z). From the regression line, we forecast future logarithms of paid losses, from which we derive the estimated losses.

Jacob: If the logarithms of the paid losses are Z, aren't the dollars e^Z?

Rachel: The relation between the logarithm and dollars differs for forecasts vs simulations.

- For the historical simulation, we take logarithms of dollars.
- For the forecasts, the relation is more complex. The expected dollars of paid loss depends on the expected logarithm and its volatility (standard deviation).

If the expected logarithms of paid losses is Z and the volatility is σ, the expected dollars of paid losses are $e^{(Z+\frac{\sigma^2}{2})}$

NORMAL AND LOGNORMAL MEANS

{*Note:* This section deals with the skewness of the lognormal distribution. It is used for forecasts of loss dollars, for estimating the regression coefficients.

Actuaries use these concepts many places, such as

- size-of-loss distributions for liability lines
- options pricing with the Black-Scholes model

Learn the concept with the illustrative worksheets. You can simulate data and see why the $\frac{1}{2}\sigma^2$ term is needed for an unbiased estimator.

If you use this regression analysis for reserve estimates, you must use the technique in this section.}

Jacob: Once we forecast the lower right-hand triangle, are we done?

Rachel: The regression equation gives the *logarithms* of the paid losses. We must convert the logarithms into dollars of paid loss.

Jacob: Don't we just exponentiate? If the logarithm of the paid loss is 20, the paid loss itself is e^{20} = $485,165,195. The reserve estimate would be $485 million.

Rachel: This is true for scalars, not for distributions. If the logarithm of the paid loss is a normal distribution with a mean of μ and a standard deviation of σ, the paid loss itself is a lognormal distribution with a mean of $e^{\left(\mu+\frac{\sigma^2}{2}\right)}$.

Jacob: Does this have a large effect?

Rachel: That depends on the variance of the distribution: the size of σ. The table below shows the effect of various σ's on the mean of the paid losses:

μ	σ	$\left(\mu+\frac{\sigma^2}{2}\right)$	*Paid Losses ($000)*	*Percentage Difference*
20	0	20	$485,165	0%
20	0.1	20.005	$487,597	1%
20	0.2	20.02	$494,966	2%
20	0.5	20.125	$549,764	13%
20	1	20.5	$799,902	65%
20	2	22	$3,584,913	639%

20	3	24.5	$43,673,179	8902%
20	4	28	$1,446,257,064	297996%

Jacob: Does the median paid loss also depend on the variance?

Rachel: No; the median paid loss is $485 million regardless of the variance.

Jacob: May we infer that actuaries should focus on the median, not the mean?

Rachel: On the contrary: focusing on the median under-estimates the expected loss.

Jacob: Why is this important?

Rachel: Most actuarial reserving estimates give a point estimate, which may be biased down because the loss distribution is skewed to the right. A regression-based method gives a confidence interval, which is valuable to an insurer monitoring its solvency. The confidence interval is symmetric in the logarithms of the paid losses. It is not symmetric in the dollars of paid loss, because the distribution of paid losses is skewed to the right.

Jacob: Does regression analysis for reserve estimates has a bias that other methods don't have?

Rachel: The traditional reserve estimates are also biased. Regression analysis is not more or less biased. But it uses a mathematical model, so we can estimate and correct the bias.

CONFIDENCE INTERVALS

Jacob: The distribution of dollars of loss is not symmetric. How do we form a confidence interval for the dollars of loss?

Rachel: We form the confidence interval for the logarithms of paid losses.

- Determine the standard error of the forecast using the formulas in the textbook.
- Determine the confidence interval for the logarithms of paid losses.
- Exponentiate to get the confidence interval for the dollars of paid losses.

Illustration: The forecast for one cell is 8.60 with a standard error of 1.15.

- The 95% confidence interval is 8.60 ± 1.96 × 1.15.
- The 95% confidence interval for dollars is $e^{(8.60 \pm 1.96 \times 1.15)}$.

Take heed: This confidence interval assumes the variance of the error term is known. If the variance of the error term is estimated from the data, use *t* values, not *Z* values.

Jacob: Do we adjust the exponent by ½ × σ^2?

Rachel: No. The confidence interval is two points, not an expected value.

Jacob: How do we get a 95% confidence interval for the sum of all future paid losses (all the cells in the forecast triangle)?

Rachel: The forecasts are partly correlated and partly independent. The required formulas are not given in the textbook, and they are not needed for this student project.