

PROJECT TEMPLATES FOR REGRESSION ANALYSIS APPLIED TO LOSS RESERVING

BACKGROUND ON PAID LOSS TRIANGLES

(The attached PDF file has better formatting.)

{The paid loss triangle helps you

- distinguish between observed data and forecasts
- visualize the relations among accident year, development period, and calendar year
- see the correlation between development period and calendar year.

Casualty and health actuaries use paid loss triangles; pension and life actuaries do not. The background information below helps you with the student project.}

Jacob: What data do we use? Do we compile paid loss triangles for a line of business?

Rachel: You *simulate* paid loss triangles. This posting is background information about loss triangles. It explains the paid loss triangles used by reserving actuaries and the triangles used for the student project.

Jacob: I have never simulated paid losses or other actuarial figures. Can I do a student project using simulation?

Rachel: We provide Excel spread-sheets that simulate the paid loss triangles in the format needed for the student project. We explain all the steps needed for the project.

- ~ If you are comfortable with regression analysis, use the project templates as guides. Modify the scenarios to create analyses that interest you. We provide concise project templates on other topics, such as life insurance mortality and auto insurance pricing.
- ~ If you feel less certain about your mastery of the course, follow the steps in the project templates. Use low values for σ so that the relations are not obscured by stochasticity. Replicating the illustrative worksheet gives the confidence needed for a student project.

Jacob: The original paper has examples of paid loss triangles. Should we review those examples?

Rachel: The paper focuses on reserving theory; the student project deals with statistics.

- Nothing from the paper is needed for the student project.
- The student project helps you understand the ideas in the actuarial paper. Many CAS candidates studying this paper have trouble with the regression concepts, such as the use of residual plots to examine parameter stability. This student project focuses on the statistical concepts.

Jacob: Does the paper use the same methods as we use in the student project?

Rachel: The paper uses advanced statistical methods. They are not explained in the paper and they are not used in the student project. We use the methods in the course textbook.

Jacob: If I have no experience with actuarial reserving, can I do this student project?

Rachel: No prior experience is needed.

- The project template uses a regression method developed by an Australian firm, not traditional actuarial reserving.
- All items needed for the student projects are explained in the postings.

Jacob: My company uses the Equifas reserving software produced by this firm. Will this software help me with the student project?

Rachel: The student projects focus on the statistical methods. The software does the statistics in the background and gives reserve estimates; it will not help you with the student project. It is easier to do the statistics in Excel than to find the statistics in Equifas.

BACKGROUND: PAID LOSS TRIANGLES

This background explains paid loss triangles. It is easier to perform the regression analysis if you understand the data.

- If you apply the statistical techniques in this student project to real data, such as your company's loss reserves or industry loss reserves, you must understand how the loss triangles in the project template relate to standard actuarial loss triangles.
- For the student project itself, we simulate paid loss triangles in the needed format. You don't need to convert loss triangles from one format to another.

Understand the types of trends and the multicollinearity among the explanatory variables. The documentation here is the quickest way to become familiar with the trends.

Jacob: What are paid loss triangles?

Rachel: We illustrate with the 10 year triangle in Table 1. We have 10 rows of accident years and 10 columns of calendar years.

Table 1: Cumulative Paid Benefits (\$000,000)

	20X0	20X1	20X2	20X3	20X4	20X5	20X6	20X7	20X8	20X9
20X0	103	226	294	334	363	384	398	412	422	433
20X1		111	238	309	356	387	409	428	442	454
20X2			108	221	286	328	354	375	391	403
20X3				111	238	311	357	392	416	434
20X4					135	299	394	458	504	534
20X5						146	314	418	490	542
20X6							159	343	463	546
20X7								146	353	485
20X8									152	406
20X9										156

- The rows are the years in which the insured accident occurs, such as an auto accident for auto insurance or a disability for health insurance.
- The column is the valuation date, such as December 31, 20X0. The valuation date is the date at which the cumulative paid losses are evaluated.
- The cells of the table are the cumulative paid losses: losses (or benefits) for accident year AY paid through the valuation date.

Illustration: The first row in Table 1 shows cumulative paid losses for accidents occurring in 20X0 evaluated at December 31, 20X0, 20X1, 20X2, and so forth. Losses for accident year 20X0 paid in 1/1/20X0 – 12/31/20X5 are \$384 million: row 20X0 and column 20X5.

The table is a rectangle. If the number of rows equals the number of columns, the table is a square.

Jacob: Are the axes of the loss triangle the accident year and the calendar year?

Rachel: Those are the axes of Table 1, and they are the axes of loss triangles in Annual Statements of insurance companies. They are simplest axes to understand.

We convert the triangle to 10 rows of accident years and 10 columns of *development years* in Table 2. Calendar years are shown by *diagonals* in the converted table.

Table 2: Cumulative Paid Benefits (\$000,000)

	1 yr	2 yrs	3 yrs	4 yrs	5 yrs	6 yrs	7 yrs	8 yrs	9 yrs	10 yrs
20X0	103	226	294	334	363	384	398	412	422	433
20X1	111	238	309	356	387	409	428	442	454	
20X2	108	221	286	328	354	375	391	403		
20X3	111	238	311	357	392	416	434			
20X4	135	299	394	458	504	534				
20X5	146	314	418	490	542					
20X6	159	343	463	546						
20X7	146	353	485							
20X8	152	406								
20X9	156									

- The accident years are 20X0 to 20X9; for the regression analysis, these are 0 to 9.
- The development years are 0-1, 1-2, ..., 9-10; for the regression, these are 0 to 9.
- The calendar years are 20X0 to 20X9; for the regression analysis, these are 0 to 9.

We use a base of zero, not a base of one. The first year is 0 (not 1), and the last year is 9 (not 10). This simplifies the multicollinearity relation and the Excel formulas.

We have data for calendar years 0 to 9, the first 10 calendar years. We forecast values for calendar years 10 to 18, the next 9 calendar years.

Jacob: What does development year mean?

Rachel: Development year 0 to 1 means from 0 years after the inception of the accident year to 1 year after inception of the accident year.

Illustration: If accident year = 20X0, development year 3 – 4 = 1/1/20X3 to 12/31/20X3.

Take heed: Development year is also called development period or maturity.

Development year is used in property-casualty insurance and health insurance.

Take heed: Tables 1 and 2 have the same data formatted differently.

- Development year is a downward sloping diagonal in Table 1 and a column in Table 2.
- Calendar year is a column in Table 1 and an upward sloping diagonal in Table 2.

Jacob: What is the current date of this table?

Rachel: The current date is December 31, 20X9. The table has 100 cells: 10 rows (accident years) by 10 columns (development years). 55 cells have data which we use to form the regression equation. We forecast the paid losses for the remaining 45 cells.

- The row for accident year 20X0 is complete: we have 10 valuations. The regression equation has no ending valuation: losses continue to be paid forever, though the size of paid losses might be small after 10 years.
- The row for accident year 20X1 has only 9 valuations, the last occurring at 12/31/20X9. The next valuation is in the future; we forecast this valuation.
- The row for accident year 20X9 has only 1 valuation, occurring at 12/31/20X9. The next nine valuations are in the future; we forecast them.

From the 55 cells with data, we forecast 45 future cells by regression analysis.

Jacob: Let me see if I understand this. We have a square with two dimensions:

- ~ X_1 is the accident year (AY), ranging from 20X0 to 20X9, or 0 to 9.
- ~ X_2 is the development year, ranging from 0 (0 to 1 year) to 9 (9 to 10 years).

Y is the cumulative paid loss in the cell. We write

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$

We use ordinary least squares estimators for the regression coefficients and we forecast the Y values in the bottom triangle of the square.

Rachel: Yes, that is the concept. We need several adjustments to use regression analysis.

- We use a third independent variable: calendar year.
- We use incremental loss payments, not cumulative loss payments.
- We use logarithms of paid losses, not dollars of paid losses.

CALENDAR YEAR

The loss dollars depend on the calendar year in which they are paid.

Illustration: Loss costs rise with inflation: If inflation is 10% per annum, a doctor's visit that costs \$100 in 20X0 costs \$110 in 20X1.

Jacob: Where is the calendar year shown?

Rachel: The calendar year (CY) in which the loss is paid is a *diagonal* running from the lower left to the upper right. We have data for 10 calendar years, 20X0 through 20X9, labeled 0 to 9. The calendar year equals the accident year plus the development year.

Illustration: Accident year 20X4 (AY = 4) at development year 3 to 4 (DY = 3) is calendar year $4 + 3 = 20X7$ (CY = 7).

Important: Make sure the indices are clear. The exercises below help clarify the indices:

- ~ Pick a cell and identify its accident year, development year, and calendar year.
- ~ For a calendar year, accident year, or development year, identify the cells in that year.

Illustration: Consider a 10×10 loss triangle for accident years 20X0 – 20X9. We use indices of 0 – 9 for each dimension.

Accident years are rows of the loss triangle:

- Accident year 20X0 (index = 0) has 10 cells with observed data.
- Accident year 20X4 (index = 4) has 6 cells with observed data; the next 4 cells are future loss payments.
- Accident year 20X9 (index = 9) has 1 cell with observed data; the next 9 cells are future loss payments.

Development years are columns of the loss triangle:

- Development year 0 to 1 (index = 0) has 10 cells with observed data.
- Development year 4 to 5 (index = 4) has 6 cells with observed data; the next 4 cells are future loss payments.
- Development year 9 to 10 (index = 9) has 1 cell with observed data; the next 9 cells are future loss payments.

Calendar years are diagonals of the loss triangle:

- Calendar year 20X0 (index = 0) has 1 cell with observed data: upper left corner. The other 9 cells are for accident years before 20X0.
- Calendar year 20X4 (index = 4) has 5 cells with observed data; the other 5 cells are for accident years before 20X0.
- Accident year 20X9 (index = 9) has 10 cells with observed data.

Indices: For an $N \times N$ loss triangle:

- If the indices are 0 to $N-1$, accident year + development year = calendar year.
- If the indices are 1 to N , accident year + development year – 1 = calendar year.

To avoid subtracting 1, we use indices of 0 to $N-1$.

Dimensions: Any two dimensions specify the cell.

Illustration: Consider a 10×10 loss triangle for accident years 20X0 – 20X9. We use indices of 0 – 9 for each dimension. The paid losses in calendar year 20X8 for accident year 20X5 are

- Accident year = 5; calendar year = 8
- Accident year = 5; development year = 3
- Calendar year = 8; development year = 3

We use dimensions with an intrinsic relation. We assume two relations for this project:

- Payments in *real dollars* have a geometric decay by development period.
- Payments at a given maturity have an inflationary increase by calendar year.

⇒ We use development period and calendar year as the dimensions of the loss triangle.

Geometric decay: Suppose the geometric decay is $Z\%$. If the expected paid loss in accident year AY and development period DP is $\$Y$, the expected paid loss in accident year AY and development period $DP + 1$ is $\$Y \times (1 - Z\%)$.

Illustration: Suppose the geometric decay is 20% and the expected paid losses for accident year 20X4 and development period 2 are $\$100,000$. We deduce the expected paid losses for other development periods:

<i>Expected Paid Losses</i>		
<i>Development Period</i>	<i>Formula</i>	<i>Value</i>
0	$\$100,000 \times 0.80^{(0-2)} =$	\$156,250
1	$\$100,000 \times 0.80^{(1-2)} =$	\$125,000
2	$\$100,000 \times 0.80^{(2-2)} =$	\$100,000
3	$\$100,000 \times 0.80^{(3-2)} =$	\$80,000
4	$\$100,000 \times 0.80^{(4-2)} =$	\$64,000
5	$\$100,000 \times 0.80^{(5-2)} =$	\$51,200

Inflation: Suppose all accident years have the same volume of business: e.g., same real exposures, such as N cars for auto insurance, or same expected claim count, such as M expected indemnity claims for workers' compensation. The inflation rate is $Z\%$ *per annum*. If the expected paid loss in calendar year CY and development period DP is $\$Y$, the expected paid loss in calendar year $CY + 1$ and development period DP is $\$Y \times (1 + Z\%)$.

Illustration: Suppose the inflation rate is 10% and the expected paid losses for calendar year 20X4 and development period 2 are $\$100,000$. We deduce the expected paid losses for other calendar years:

Expected Paid Losses		
Calendar Year	Formula	Value
2	$\$100,000 \times 1.10^{(2-4)} =$	\$82,645
3	$\$100,000 \times 1.10^{(3-4)} =$	\$90,909
4	$\$100,000 \times 1.10^{(4-4)} =$	\$100,000
5	$\$100,000 \times 1.10^{(5-4)} =$	\$110,000
6	$\$100,000 \times 1.10^{(6-4)} =$	\$121,000
7	$\$100,000 \times 1.10^{(7-4)} =$	\$133,100

ORTHOGONAL DIMENSIONS

Actuaries generally use orthogonal (uncorrelated) dimensions.

Illustration: Suppose the volume of business increases 5% each accident year and the paid losses decrease 20% each development period. Accident year and development period are orthogonal dimensions: increasing the accident year doesn't affect the development period and increasing the development period doesn't affect the accident year.

Calendar year and development period are correlated with a ρ of 50%.

For any two independent variables, the set of cells have one common member.

- ~ Accident year 20XN and development year 20XN have one cell in common.
- ~ Accident year 20XN and calendar year 20XN have one cell in common.
- ~ Calendar year 20XN and development year 20XN have one cell in common.

The one cell in common differs for each pair (except for $N = 0$).

Jacob: How large should the loss triangle be?

Rachel: An $N \times N$ loss triangle has $\frac{1}{2} \times N \times (N+1)$ cells with data. A common rule of thumb is that we need about 30 to 40 observations \times the number of independent variables. For changing regression coefficients, we have three independent variables, so we want 90 to 120 observations. If $N = 15$, $\frac{1}{2} \times N \times (N+1) = \frac{1}{2} \times 15 \times 16 = 120$.

Take heed: Do not feel bound by this rule of thumb. Your student project might have 40 observations or 40,000 observations. The time series daily temperature project template has 50,000+ observations.

Jacob: How does each dimension affect the paid losses?

Rachel: Three items affect the paid losses: exposure growth, inflation, and payment pattern

Volume of business or exposure growth is an *accident year* phenomenon. If the insurer covers 1,000 cars in 20X0 and 2,000 cars in 20X1, loss costs (before inflation) are twice as high for accident year 20X1 as for accident year 20X0.

Take heed: The proper measure of exposure for loss reserving is expected claims, if the size-of-loss distribution does not change. If the insurer expects 1,000 claims in 20X0 and an increase in the price of oil or a regulatory restriction on filing claims cause 800 expected claims in 20X1, the loss costs (before inflation) are 20% lower for accident year 20X1 as for accident year 20X0.

Inflation is a *calendar year* phenomenon: If inflation is 10% per annum, the losses for accident year 20X1 paid in 20X1 are 10% higher than the losses for accident year 20X0 paid in 20X0, assuming the volume of business does not change (no exposure growth).

If patients have 100 doctor visits in development year 0 to 1 and another 100 doctor visits in development year 1 to 2, and paid losses are \$10,000 in development year 0 to 1, the paid losses are \$11,000 in development year 1 to 2.

The *loss payment pattern* is a *development year* phenomenon. Suppose the loss payment pattern extends over four years in the ratio 40% - 30% - 20% - 10%. The losses paid in the year following the accident (1 to 2 years) are 30% / 40%, or $\frac{3}{4}$ of the losses paid in the year of the accident (0 to 1 year), assuming there is no inflation.

Three explanatory variables affect the paid losses: the exposures (in real dollars), the payment pattern (in real dollars), and inflation.

Jacob: Do we use all three dimensions simultaneously?

Rachel: If paid losses are \$10,000 in accident year 20X0 and development year 0 to 1, and all three trends above apply, the paid losses in accident year 20X1 and development year 1 to 2 are $\$10,000 \times 2 \times 1.10^2 \times 75\% = \$18,150$.

Jacob: Why did we square the trend factor of 1.10?

Rachel: Accident year 20X0 and development year 0 to 1 is calendar year 20X0 (CY = 0). Accident year 20X1 and development year 1 to 2 is calendar year 20X2 (CY = 2). We have two years of inflation.

Exercise: Apply Rachel's calculation to accident year 20X7 and development year 5 - 6. This cell is in the forecast section of the loss triangle. If we determine the three trends, we can forecast the paid losses in this cell.

Jacob: This sound simple; why do we need regression analysis?

Rachel: We focus on three reasons in the student project:

- We don't know the values of these trends; we estimate them from the sample data.

Jacob: Does this table just take first differences?

Rachel: Yes. The rows are the years in which the insured accident occurs, such as an auto accident (for auto insurance) or a disability (for health insurance). The column is period during which benefits are paid, such as a court award or disability benefits.

Table 3 says that for accidents occurring in 20X0, \$103 million was paid in 20X0, \$123 million was paid in 20X1, ..., and \$11 million was paid in 20X9. For accidents occurring in 20X1, \$111 million was paid in 20X1, \$111 was paid in 20X2, \$127 million was paid in 20X3, ..., and \$14 million was paid in 20X9. We have converted the cumulative paid losses of the original loss triangle into incremental paid losses.

Take heed: The illustrative worksheet starts with logarithms of incremental paid losses in the needed format. You don't need any adjustments. To apply the techniques to your own company's data, adjust cumulative paid losses to logarithms of incremental paid losses.

This material is *background* for the student project. It orients you to the data being used, but you do not perform these calculations. If you grasp the concepts but are hazy about the details, that's fine. The illustrative worksheets teach the details for this project template.

ADDITIVE VS MULTIPLICATIVE

Each independent variable is a multiplicative relation; we need additive relations for the regression equation.

- If inflation is 10% per annum, we expect calendar year 20X1 payments to be 10% higher than calendar year 20X0 payments (assuming no exposure growth).
- If the insurer's business is growing 5% per annum, we expect accident year 20X1 payments to be 5% higher than accident year 20X0 payments (assuming no inflation).
- Given a 40% - 30% - 20% - 10% loss payment pattern, we expect the inflation-adjusted losses paid in the year after the accident (development year 1 - 2) to be 25% lower than the inflation-adjusted losses paid in the year of the accident (development year 0 - 1).

The model for this relation is $Y = \alpha \beta_1^{CY} \beta_2^{DY} \beta_3^{AY} \epsilon$.

Jacob: This model is not linear. We can not use ordinary least squares estimators.

Rachel: The model is not linear, but it is inherently linear. To transform a multiplicative relation into an additive relation, we take logarithms:

$$\ln(Y) = \ln(\alpha) + CY \ln(\beta_1) + DY \ln(\beta_2) + AY \ln(\beta_3) + \ln(\epsilon)$$

We re-define the estimators as $\beta_1' = \ln(\beta_1)$, $\beta_2' = \ln(\beta_2)$, and $\beta_3' = \ln(\beta_3)$.

Note: Most actuarial and financial relations are multiplicative or exponential. For regression analysis, we use logarithms of the independent and dependent variables.

Jacob: What change do we make to the loss triangle?

Rachel: The calendar year, development year, and accident year do not change. Instead of the paid loss, we use the logarithm of the paid loss; that is the only change.

Illustration: Instead of \$103 million in accident year 20X0 and development year 0 to 1, we have $\ln(103 \text{ million}) = 18.450$. We convert dollars of loss to logarithms of dollars of loss.