Student Project
Regression Analysis
Session: Fall 2013

Stochastic Loss Reserving: Parameter Stability

Wenqiang Shao*

SOA ID 336965

*Email: adam.w.shao@gmail.com

1 Executive Summary

This project investigates the effects of unstable parameters on multi-variate linear regressions, in the context of loss reserving. It is found in this project that failing to capture the changed parameter results in a low regression goodness of fit. The residuals also show patterns, which violates the basic assumptions in linear regressions. A dummy variable is then included to capture the discrete change (jump) in the inflation rate. Results show significant improvement in the goodness of fit and results no longer show undesirable patterns.

2 Introduction

Triangle loss reserving technique is widely used by actuaries in general insurance practice. The underlying rationale behind the technique is that loss claims are affected by the development pattern of claiming processes and the inflation rate. The two factors are respectively captured by the development year and calendar year. Their effects are usually assumed to be multiplicative: the development pattern is assumed to be an exponential decay, and the effect of inflation is obviously multiplicative.

The accident year that represents the time when loss occurs is also an important factor in determining loss claims. For example, as the insurance company grows in business, the exposure is getting larger. This results in a higher loss amount as the compare grows. In addition, it is possible that the accident rate changes over time due to possible catastrophes or some social effects. But include the accident year, as well as the development and calendar years, results in perfect multi-collinearity problem, because the calendar year equals the sum of the accident year and the development period. Multi-collinearity does not bias the estimation but makes the estimated parameters inaccurate.

With this regard, two additional assumptions are made: (1) the risk exposure of the company is constant across the analysis window for 15 years; (2) the underlying accident occurring rate does not change in the 15 years' examining window. The development pattern is held constant as discussed earlier, but the inflation rate can change over time. For simplicity, a jump of inflation rate in year 10 is assumed in simulating a second panel of loss claims data. The original regression is again performed to show the effects of unstable inflation rate. A dummy variable is then included in account for the impact of the jump in the inflation rate.

The rest of the project is arranged as follows. The next section describes how the data are simulated and the major model framework. Section 4 assesses the impact of stochasticity on regression results. The stochastic term is set to zero in the first step and then the standard deviation of the error term is changed to 0.5 and 2 for scenario analyses. Section 5 investigates the impact of unstable parameters and provides remedies - including a dummy variable in the regression. Section 6 concludes.

3 Data Simulation

The data employed in the project are simulated in excel and are shown in the first sheet of attached excel file. The logarithm of increment loss Y happened in year X_1 and ended in development year X_2 is determined only by the two factors X_1 and X_2. In mathematical formula, the regression equation is:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \varepsilon, \tag{3.1}$$

where α is the intercept, β_1 is the slope parameter for development year that evaluates the rate of exponential decay of the development pattern, β_2 is the slope parameter for the calendar year that can be interpreted as the inflation rate, and ε is the normal disturbance term.

In the simulation of the first data set, the parameters in the regression equation are pre-specified:

- The horizon of the analysis is 15 years;

- $X_1 = 0, 1, \cdots, 14$, representing the 15 development years;
- $X_2 = 0, 1, \cdots, 14$, representing the 15 calendar years;
- $\alpha = 12$, representing an average value of 12 of loss logarithms;
- β_1 = -0.25, suggesting a constantly 25% exponential decaying rate;
- $\beta_2 = 0.1$ implying a stable inflation rate.

The standard deviation σ of the disturbance term is varied in three scenarios: zero in Scenario 1 (Deterministic), 0.5 in Scenario 2 (Low stochastic), and 2 in Scenario 3 (High stochastic).

The second data set is simulated for the analysis on the impact of unstable inflation rate. The only difference from the first data set is that there is a jump in the inflation rate. This is expressed in the following equation:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + (\beta_2^{'} - \beta_2) \max(X_2 - 9,\ 0) + \varepsilon, \tag{3.2}$$

where $\beta_2^{'}$ is the new inflation rate starting from year 10, and other notations are the same as in Equation (3.1). $\beta_2^{'}$ is set to be 0.2 in the data simulation.

4 Scenario Analysis of Error Volatility

4.1 Scenario 1: deterministic relation

Suppose the loss is perfectly determined by the development year and the calendar year, where $\sigma = 0$. The regression is performed and results are shown in Table 1. As expected, the estimated parameters are exactly the input values used in the simulation. Measures of goodness of fit all suggest that the model fits well to the data, which are of course what should be expected. Residuals are of course zero: the variations in the dependent variable are all explained by the linear regression.

Table 1: Regression results of Scenario 1

Parameter	**Estimate**	**s.e.** (10^{-14})	t**-statistic**(10^{15})	**Lower**	**Upper**	**Width**
α	12.0000	0.1212	9.8972	12.0000	12.0000	0.0000
β_1	-0.2500	0.0140	-1.7886	-0.2500	-0.2500	0.0000
β_2	0.1000	0.0140	0.7155	0.1000	0.1000	0.0000
R^2	F	p-value	$\hat{\sigma}^2$	N		
1.0000	2.1075×10^{30}	0.0000	0.0000	120		

4.2 Scenario 2: low volatility ($\sigma = 0.5$)

A normal random variable with standard deviation of 0.5 is included in the regression equation as shown in Equation (3.1), where $\varepsilon \sim \text{Norm}(0, 0.5^2)$. The regression is performed and results are shown in Table 2. It can be seen from the table that estimated parameters are not equal to the true parameters pre-specified in the data simulation, but the 95% confidence intervals do include the true parameters.

Table 2: Regression results of Scenario 2

Parameter	**Estimate**	**s.e.**	t**-statistic**	**Lower**	**Upper**	**Width**
α	12.2150	0.1233	99.0935	11.9709	12.4592	0.4883
β_1	-0.2554	0.0142	-17.9740	-0.2836	-0.2273	0.0563
β_2	0.0807	0.0142	5.6800	0.0526	0.1089	0.0563
R^2	F	p-value	$\hat{\sigma}^2$	N		
0.7427	168.8227	0.0000	0.2403	120		

4.3 Scenario 3: high volatility ($\sigma = 2$)

A normal random variable with standard deviation of 2 is included in the regression equation as shown in Equation (3.1), where $\varepsilon \sim \text{Norm}(0, 4)$. The regression is performed and results are shown in Table 3. As expected, estimated parameters are not necessarily equal to the true parameters pre-specified in the data simulation and the 95% confidence intervals do include the true parameters. The interval widths are considerably larger than those of Scenario 2. It can be seen from the table that R^2 is less than 0.2, suggesting that no more than 20% of the variations in the dependent variable are explained by the regression equation.

Parameter	Estimate	s.e.	t-statistic	Lower	Upper	Width
α	12.8602	0.4931	26.0817	11.8837	13.8367	1.9530
β_1	-0.2716	0.0568	-4.7791	-0.3842	-0.1591	0.2251
β_2	0.0229	0.0568	0.4021	-0.0897	0.1354	0.2251
R^2	F	p-value	$\hat{\sigma}^2$	N		
0.1937	14.0534	0.0000	3.8447	120		

Table 3: Regression results of Scenario 3

Residuals are separately plotted against the two explanatory variables. The plots are shown in Figure 1. It can be seen that the residuals do no show certain patterns.

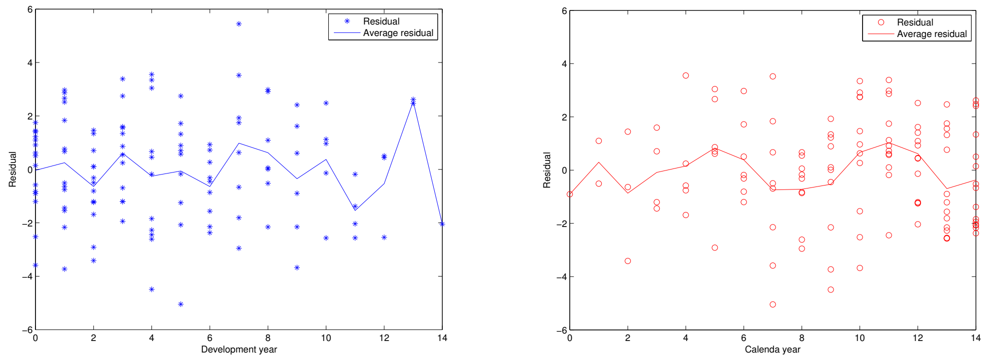

Figure 1: Residual plot of Scenario 3

5 Unstable Inflation Rate

Discrete changes are assumed to take place in the inflation rate. In the first subsection I show the impact of unstable inflation rate on the regression for three different scenarios by using the original regression equation. Then the remedy is shown in the second subsection to address the unstable inflation rate.

5.1 Impact of unstable inflation rate

Without taking into account the problem, the regression equation expressed in Equation (3.1) is again fit to the data simulated based on the unstable inflation. The scenarios of different magnitudes of volatilities are all used in the regression. Results are shown in Table 4. It can be seen from Table 4 that the variations caused by

Table 4: Regression results of data with a discrete change in inflation rate

Scenario 1: Deterministic

Parameter	**Estimate**	**s.e.**	**t-statistic**	**Lower**	**Upper**	**Width**
α	11.7551	0.0244	481.9821	11.7068	11.8034	0.0966
β_1	-0.2500	0.0028	-88.9193	-0.2556	-0.2444	0.0112
β_2	0.1445	0.0028	51.4088	0.1390	0.1501	0.0111
R^2	F	p-value	$\hat{\sigma}^2$	N		
0.9855	3,986	0.0000	0.0094	120		

Scenario 2: Low volatility

Parameter	**Estimate**	**s.e.**	**t-statistic**	**Lower**	**Upper**	**Width**
α	11.9308	0.1378	86.5749	11.6578	12.2037	0.5459
β_1	-0.2424	0.0159	-15.2585	-0.2739	-0.2109	0.0630
β_2	0.1255	0.0159	7.8983	0.0940	0.1569	0.0629
R^2	F	p-value	$\hat{\sigma}^2$	N		
0.6656	116	0.0000	0.3003	120		

Scenario 3: High volatility

Parameter	**Estimate**	**s.e.**	**t-statistic**	**Lower**	**Upper**	**Width**
α	12.4576	0.5544	22.4708	11.3597	13.5556	2.1959
β_1	-0.2196	0.0639	-3.4362	-0.3462	-0.0930	0.2532
β_2	0.0683	0.0639	1.0685	-0.0583	0.1949	0.2532
R^2	F	p-value	$\hat{\sigma}^2$	N		
0.0956	6.1852	0.0028	4.8604	120		

the change in inflation rate cannot be explained by the regression and are captured in the residual part. This also provides the reason for higher estimated variances of errors ($\hat{\sigma}^2$) compared to the estimations from the previous section. For example, the estimated variance for the low volatility scenario is 0.3003, higher than that estimated variance of 0.2403 from Section 4.2. The estimate variances are also higher than the true values that are used as input in the data simulation, which also justifies that failing to capture the change in inflation rate increases the error.

Residual plots for the three scenarios are shown in Figure 2. The discrete change in the inflation rate from year 9 can be easily identified from the two figures in the top panel where results are for the deterministic scenario. Residuals show a sharp turning point at year 9. As the volatility of the error increases, it is more difficult to identify the change. In addition the pattern of the residuals of Scenario 2 and Scenario 3 are very similar except that they are of different scales. This is due to the relatively large volatility compared to the change in the inflation rate.

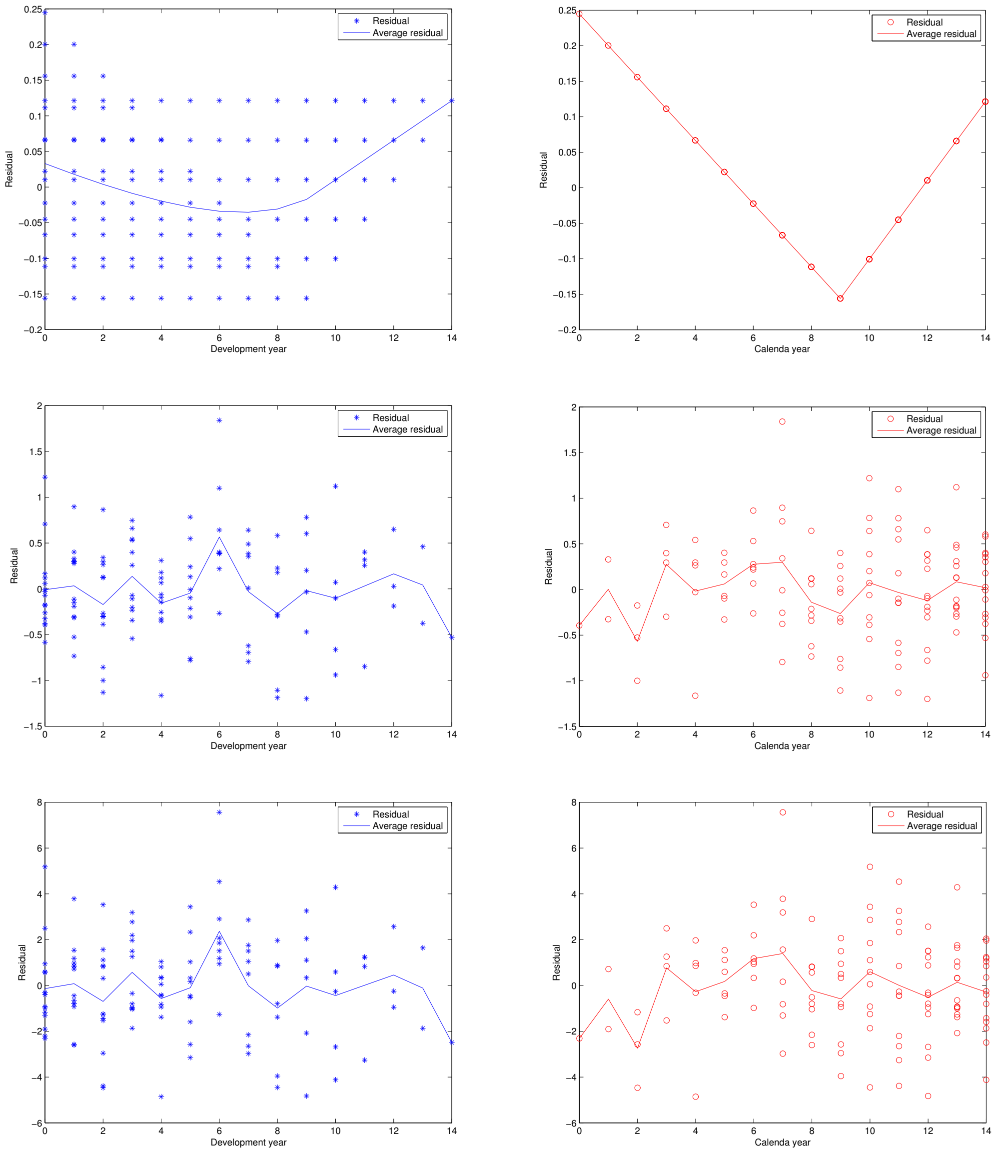

Figure 2: Residual plots of data with a discrete change in the inflation rate from year 9. The top two figures are for Scenario 1, the middle two figures are for Scenario 2, and the bottom figures are for Scenario 3.

5.2 Remedy

Equation (3.2) shows the underlying generation method of data. To address the problem, the maximum part needs more care rather than just ignoring it. In fact, the maximum part is equivalent to the product of a dummy variable and the shifted calendar year. In mathematical notation, it can be expressed in the following equation:

$$\max(X_2 - 9,\ 0) = D(X_2 - 9), \tag{5.1}$$

where D is a dummy variable that equals 0 if $X_2 \leq 9$ and equals 1 if $X_2 > 9$. Therefore, the regression equation can be expressed in Equation (5.2):

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \gamma D(X_2 - 9) + \varepsilon, \tag{5.2}$$

where γ can be interpreted as the difference of the changed inflation rate and the original inflation rate (i.e. $\gamma = \beta_2' - \beta_2$).

Perform regression Equation (5.2) on the second data set, results are shown in Table 5. From the first panel where the deterministic scenario is analysed, it is obvious that the additional variable captures the change in the inflation rate. For the other two scenarios, the improvement in goodness of fit is marginal and sometimes becomes worse off. This is due to the disturbance of the errors with large volatilities. As the volatility gets larger, the improvement is smaller. For example, in the low volatility scenario the estimated incremental inflation rate is somewhat close to the true value, but the estimated increment is negative in the high volatility scenario.

Residual plots for the three scenarios are again shown in Figure 3. It is obvious that the residuals of Scenario 1 become zero once we include an interactional term of dummy variable and the shifted calendar year as the remedy. As analysed above, the improvement in the residuals is minimal in the two scenarios with considerable volatility.

6 Conclusion

This project first assesses the impact of the disturbance term on the regression goodness of fit by varying the volatility of the error term. Results indicate that the error sum of squares gets larger and R^2 gets smaller as the volatility increases.

The impact of unstable parameters on the regression goodness of fit and parameter estimation accuracy is also investigated in this project. Results indicate that failing to capture the changes in parameters can result in larger unexplained variations and residuals with undesirable patterns. The accuracy of the estimation is then questionable. But as the volatility of the true disturbance term increases, the improvement of capturing the changes in parameters becomes marginal.

Table 5: Remedy regression results

Scenario 1: Deterministic

Parameter	**Estimate**	**s.e.** (10^{-14})	***t*-statistic**(10^{16})	**Lower**	**Upper**	**Width**
α	12.0000	0.1066	1.1255	12.0000	12.0000	0.0000
β_1	-0.2500	0.0090	-0.2775	-0.2500	-0.2500	0.0000
β_2	0.1000	0.0160	0.0626	0.1000	0.1000	0.0000
γ	0.1000	0.0296	0.0338	0.1000	0.1000	0.0000
R^2	F	p-value	$\hat{\sigma}^2$	N		
1.0000	2.15×10^{30}	0.0000	0.0000	120		

Scenario 2: Low volatility

Parameter	**Estimate**	**s.e.**	***t*-statistic**	**Lower**	**Upper**	**Width**
α	11.9627	0.1888	63.3695	11.5888	12.3366	0.7478
β_1	-0.2424	0.0160	-15.1972	-0.2740	-0.2108	0.0632
β_2	0.1197	0.0283	4.2305	0.0636	0.1757	0.1121
γ	0.0131	0.0524	0.2488	-0.0908	0.1169	0.2077
R^2	F	p-value	$\hat{\sigma}^2$	N		
0.6658	77.0378	0.0000	0.2803	120		

Scenario 3: High volatility

Parameter	**Estimate**	**s.e.**	***t*-statistic**	**Lower**	**Upper**	**Width**
α	11.8509	0.7551	15.6943	10.3553	13.3465	2.9912
β_1	-0.2196	0.0638	-3.4420	-0.3460	-0.0932	0.2528
β_2	0.1786	0.1131	1.5790	-0.0454	0.4027	0.4481
γ	-0.2478	0.2098	-1.1811	-0.6633	0.1677	0.8310
R^2	F	p-value	$\hat{\sigma}^2$	N		
0.1064	4.6024	0.0044	4.8441	120		

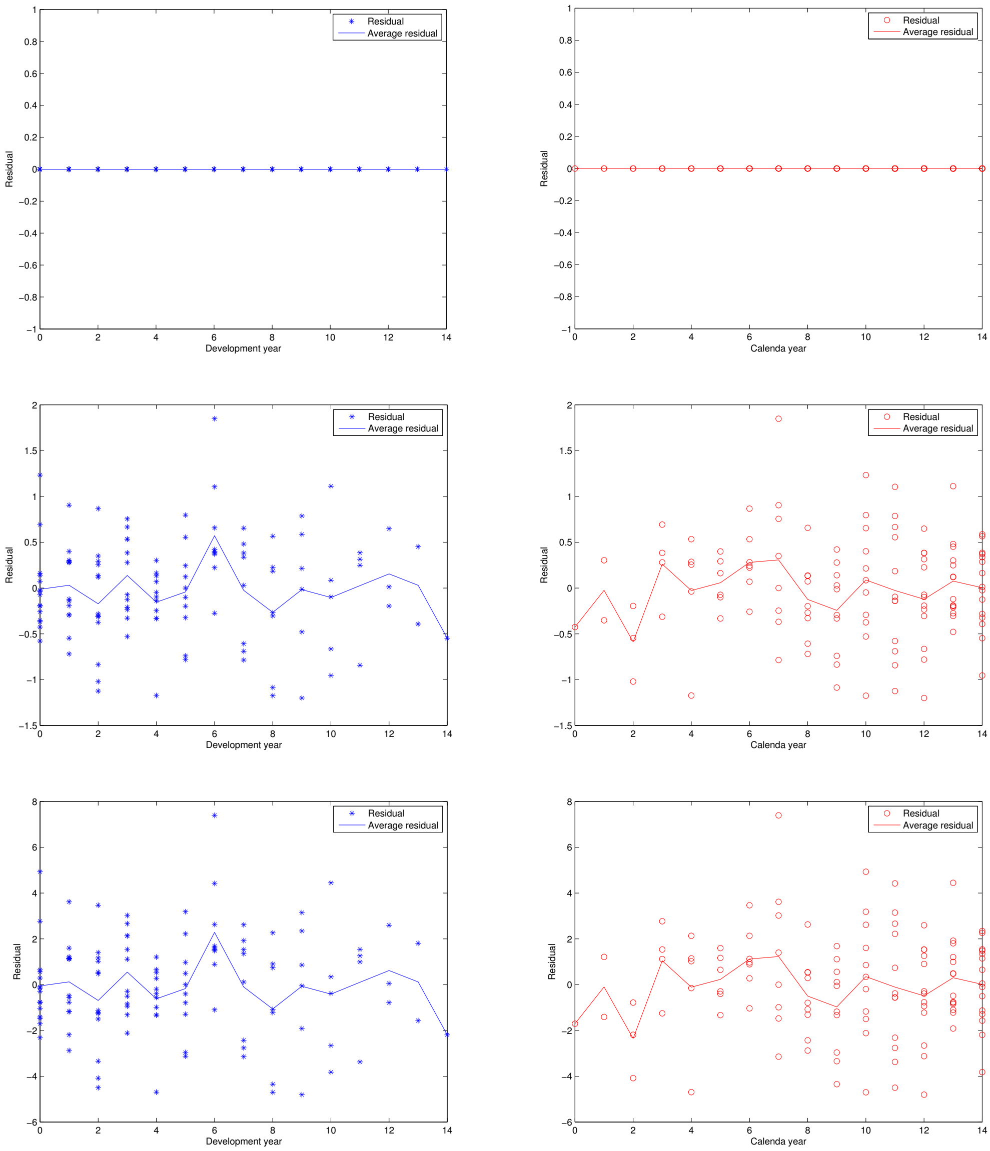

Figure 3: Residual plots of remedy regressions. The top two figures are for Scenario 1, the middle two figures are for Scenario 2, and the bottom figures are for Scenario 3.