Loss Simulation Model Testing and Enhancement

Kailan Shang FSA, CFA, PRM, SCJP

Abstract. This paper is a response to the Casualty Actuarial Society's call for papers on the topic of "Testing Loss Reserving Methods, Models and Data Using the Loss Simulation Model." Its goal is to test and improve the Loss Simulation Model (LSM). The testing methods used are good sources for analyzing real claim data. A two-state regime-switching feature is also built into the model to add an extra layer of flexibility to describe claim data.

Motivation. The testing and enhancement of the Loss Simulation Model helps improve and refine the model. The test method may also be a good reference for performing tests on real claim data.

Method. Statistical tests are applied to the data simulated by the Loss Simulation Model. Standard distribution fitting methods such as maximum likelihood estimation are used to analyze real claim data. The open-source software LSM is enhanced via programming in Visual Basic.

Results. The LSM is enhanced with two-state regime-switching capability. Testing of the Loss Simulation Model according to the list suggested by the Loss Simulation Model Working Party is conducted. It shows the consistency between model input and model output for the addressed issues except case reserve adequacy.

Conclusions. Categorical variable and two-state regime-switching capability are added to the LSM. Testing of the LSM increases the confidence in the accuracy of this advanced and useful tool.

Keywords. simulation model; loss reserving; regime-switching; copula.

Table of Contents

LOSS SIMULATION MODEL TESTING AND ENHANCEMENT 1
1. INTRODUCTION 3
1.1 Research Context 3
1.2 Objective 3
1.3 Outline 4
2. MODEL TESTING 4
2.1 Negative Binomial Frequency Distribution 5
2.2 Correlation 7
2.2.1 Clayton Copula 8
2.2.2 Frank Copula 10
2.2.3 Gumbel Copula 12
2.2.4 *t* Copula 14
2.2.5 Correlation between claim size and report lag 16
2.3 Severity trend 18
2.4 Alpha in Severity Trend 20
2.5 Case Reserve Adequacy Distribution 22
3. REAL DATA AND SIMULATED DATA 25
3.1 Property Line 26
3.2 Liability Line 29
3.3 Correlation 33
4. MODEL ENHANCEMENT 34
4.1 Two-State Regime-Switching Distribution 34
4.2 Testing 37
4.2.1 Frequency 39
4.2.2 Severity 44
4.2.3 Correlation 48
5. CONCLUSION AND FURTHER DEVELOPMENT 53
Acknowledgment 54
APPENDIX A. R CODE 55
A.1 Negative Binomial Frequency Distribution Testing 55
A.2 Correlation Test 56
A.3 Severity Trend 61
A.4 Alpha in Severity Trend 62
A.5 Case Reserve Adequacy 63
A.6 Real Claim Data Fitting 64
A.7 Two-State Regime-Switching Feature Testing 67
APPENDIX B. QUICK GUIDE FOR TWO-STATE REGIME-SWITCHING 71
5. REFERENCES 75
Biography of the Author 75

1. INTRODUCTION

This paper is a response to a call for papers by the Casualty Actuarial Society (CAS) on "Testing Loss Reserving Methods, Models and Data Using the Loss Simulation Model."

1.1 Research Context

The loss simulation model (LSM) is a tool created by the CAS Loss Simulation Model Working Party (LSMWP)[1] to generate claims that can then be used to test loss reserving methods and models.[2] The LSMWP paper suggests some model enhancement and additional tests of the LSM.[3] Based on the suggested list, additional tests are performed on the simulated results to test the correlation, severity trend, negative binomial distribution for frequency, and case reserve adequacy distribution. Real claim data are used to fit into distributions to determine parameters in LSM. The model is also enhanced by allowing a two-state regime-switching distribution model for both frequency and severity.

1.2 Objective

A. Model Testing

1. Frequency distribution testing

   Test the Negative Binomial frequency distribution using various goodness-of-fit testing methods.

2. Test correlation

   Test the frequency correlation between different lines for other copula types in addition to the normal copula: Frank, Gumbel, Clayton, and T copula. Those types of copulas are very important to capture the tail risk while the normal copula that has been tested by LSMWP assumes a linear correlation behavior.

   Test the correlation between report lag and size of loss under a normal copula.

3. Severity trend and Alpha testing

   Apply time series analysis techniques to find the trend and alpha parameters from simulated

[1] For more information about LSM and LSMWP, please visit http://www.casact.org/research/lsmwp.
[2] CAS Loss Simulation Model Working Party Summary Report,. pages 4-5,
[3] CAS Loss Simulation Model Working Party Summary Report, page 33, The paper addresses the first suggestion about model enhancement and tests 1, 2, 3, 5 of the LSM in the suggestion list.

data and compared with parameter inputs to check the statistical credibility. Ordinary least square (OLS) method and hypothesis testing are applied to the deterministic time trend model.

4. Case Reserve Adequacy

A 40% time point case reserve adequacy distribution is tested against simulation model input.

B. Real Data and Simulated Data

Marine claim data are used to fit the distribution for frequency and severity using Maximum Likelihood Estimation (MLE) and OLS for trend and seasonality analysis. The correlation between different lines is also estimated. The estimated distribution type and parameters can then be input into Loss Simulation Model (LSM) for simulation and further testing of different reserve methods. This illustrates how to use real data to determine inputs for the LSM. Unfortunately, only final claim data are available and there is no detailed paid loss history. Therefore, the Meyers' Approach[4] is not applied to test rectangles generated by the simulation model against those from the real data due to the missing details.

C. Model Enhancement

A categorical variable is included to enable setting parameters/distribution type for different states. A two-state regime-switching flexibility is then built in to enable moving from one state to the other state. The transition matrix of states from one period to another is an input table in the user interface. Hopefully, this can add the flexibility to mimic the underlying cycle we normally see in P&C business. The enhancement is intended for frequency and severity distribution. The simulated results based on this enhancement are also tested.

1.3 Outline

The remainder of the paper proceeds as follows. Section 2 will discuss the methodology and results of testing LSM. Section 3 will fit real claim data to distribution and determine trend parameters which are inputs for LSM. Section 4 will present the enhancement being made for the LSM. Section 5 will discuss the conclusion and potential further improvement of the LSM.

2. MODEL TESTING

The LSM is used to simulate claim and transaction data for testing. Once the simulator is run

[4] CAS Loss Simulation Model Working Party Summary Report, pages. 7-8.

with specified parameters, the relevant R code in Appendix A is applied to the claim file and transaction file output from LSM. Running R code, process output data and apply statistical tests. A conclusion based on the statistical test results is then drawn for the addressed issues.

2.1 Negative Binomial Frequency Distribution

This test is to check if the simulated frequency result is consistent with the LSM input parameters for negative binomial distribution.[5]

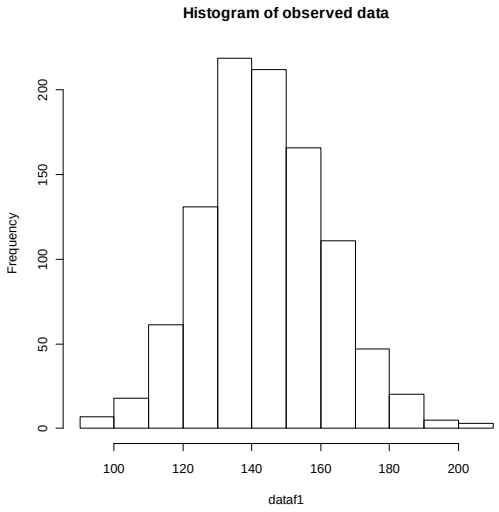
Test Parameters:

- ✓ One Line with annual frequency Negative Binomial (size = 100, probability = 0.4)
- ✓ Monthly exposure: 1
- ✓ Frequency Trend: 1
- ✓ Seasonality: 1
- ✓ Accident Year: 2000
- ✓ Random Seed: 16807
- ✓ # of Simulations: 1000

Firstly, we draw a histogram of the simulated frequency data to give an indication of the distribution type.

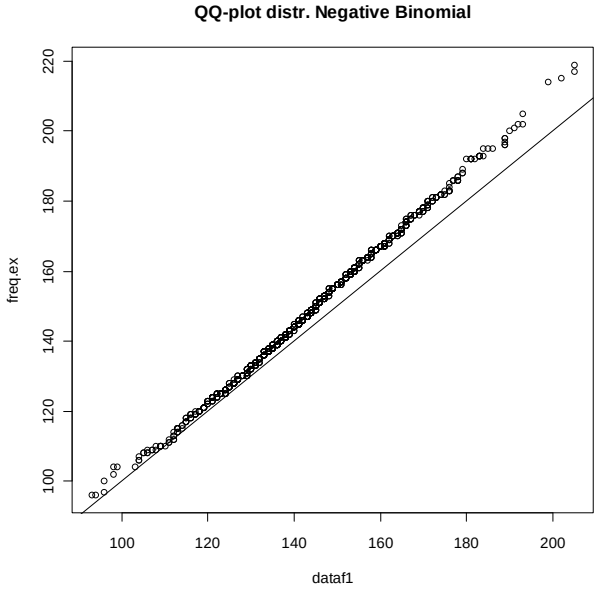
[5] Negative Binomial Distribution: "A discrete probability distribution of the number of successes in a sequence of Bernoulli trials before a specified (non-random) number *r* of failures occurs."
probability mass function as $\binom{k+r-1}{k} \cdot (1-p)^r \cdot p^k$ p: probability of success, *k*: number of successes.
More details can be found at http://en.wikipedia.org/wiki/Negative_binomial_distribution.

Figure 1. Histogram of simulated frequency data (Negative Binomial)



A QQ plot would also be a straightforward way to compare the simulated results with the intended distribution – Negative Binomial (Size = 100, probability = 0.4). From Figure 2, we can see that it is a good fit although the expected frequency distribution in the LSM has a slightly longer tail than the simulated results.

Figure 2. QQ Plot – Simulated results vs. Negative Binomial (size = 100, prob. = 0.4)



Goodness-of-fit test using Pearson's Chi-squared statistic is performed. The results disallow rejecting the null hypothesis that the simulated frequency follows negative binomial distribution.

Goodness-of-fit test for nbinomial distribution

	X^2	df	$P(> X$^2)
Pearson	197.3816	205	6.360712*e*-01

In addition, using maximum likelihood (ML) method to fit the negative binomial distribution and calculate the likelihood ratio statistics implies the same conclusion.

Goodness-of-fit test for nbinomial distribution

	X^2	df	$P(> X$^2)
Likelihood Ratio	113.3462	94	0.08499854

Using ML method gives us an estimation of the parameters as follows:

	size	**mu**
Estimation	117.2378284	144.1840000
Standard deviation	9.5150285	0.5670163

Our LSM inputs (size = 100 and prob = 0.4) imply mu = 150 and variance = 375. The estimated value gives us size = 117 and prob = 0.448. The variance is 321.5. Here prob = size/(size+mu) and variance = mu + mu^2/size.[6]

We can see that at the significance level of 5%, the confidence interval for size is (98.59, 135.89) which includes the model input size = 100. The mean and variance of the model input and simulated results are also not too far away. Those results together with the goodness-of-fit tests indicate that simulated frequencies are consistent with the negative binomial distribution.

2.2 Correlation

In LSM, there are two places where correlation can be built between variables. One is the correlation between frequencies of different product lines. The other is the correlation between claim size and report lag. The method of modeling correlation in LSM is a copula, which can capture tail risk better than standard linear correlation assumption. Available copula types in LSM include Clayton, Frank, Gumbel, *t*, and normal copula. A normal copula among different lines' frequencies was tested and summarized in LSMWP paper.[7]

Sections 2.2.1 to 2.2.4 discuss the correlations among frequencies of different lines. Section 2.2.5

[6] Package *stats* version 2.12.0, R Documentation, The Negative Binomial Distribution.
[7] CAS Loss Simulation Model Working Party Summary Report, Section 6.2.3, pages 29-33.

discusses the correlation between claim size and report lag. In each section, once the simulator is run with these parameters, the R code in Appendix A.2 is applied to the output claim file and/or transaction file. Running the code produces joint frequencies for two lines of correlated loss size and report lag. Statistical methods are then applied to test the consistency between model inputs and model outputs. Each section contains the model parameters used and a discussion of how well the copula fits the output of the simulation.

2.2.1 Clayton Copula

This test is to check if the Clayton Copula[8] modeling in LSM is appropriate for correlation between frequencies of different lines.

Test Parameters:

- ✓ Two Lines with annual frequency Poisson ($\lambda = 96$)
- ✓ Monthly exposure: 1
- ✓ Frequency Trend: 1
- ✓ Seasonality: 1
- ✓ Accident Year: 2000
- ✓ Random Seed: 16807
- ✓ Frequency correlation: Θ 5, $n = 2$ (see footnote 8)
- ✓ # of Simulations: 1000

A simple way to compare is to draw a scatter plot for the intended copula and simulated frequency pairs. Figures 3 and 4 below show that they are of similar patterns.

[8] Clayton Copula: $C_\theta^n(u) = (u_1^{-\theta} + u_1^{-\theta} + \cdots + u_n^{-\theta} - n + 1)^{-1/\theta}$ $\theta > 0$. Details can be found on page 153 of Nelsen 2006.

Figure 3. Clayton Copula (5)

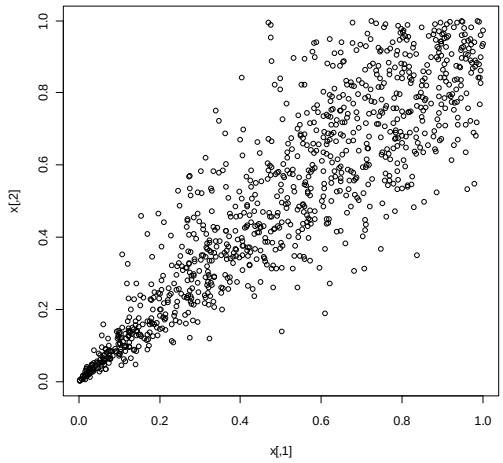
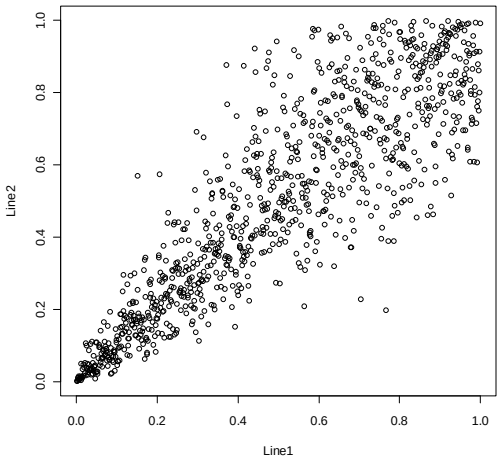


Figure 4. Simulated Frequencies



Clayton copula parameter is then estimated based on simulated frequency data using two methods.

(1) The estimation is based on the maximum likelihood and a sample of size 998.

	Estimate	Std. Error	z value	Pr(>|z|)
parameter	4.112557	0.1441209	28.53546	0

The maximized loglikelihood is 822.3826.

(2) The estimation is based on the inversion of Kendall's tau and a sample of size 998.

	Estimate	Std. Error	z value	Pr(>|z|)
parameter	4.623835	0.2434634	18.99191	0

We can see that the model parameter Θ as 5 is within 95% confidence interval based on inversion of Kendall's tau but not that for maximum likelihood estimation. This is also consistent with goodness-of-fit test results as below. We use two methods to test whether the correlation between simulated frequencies is consistent with assumed copula.

(1) Using Maximum Likelihood method for parameter estimation:

Parameter estimate(s): 4.112557

Cramer-von Mises statistic:[9] 0.03709138 with *p*-value 0.004950495

(2) Using Inversion of Kendall's tau method for parameter estimation:

Parameter estimate(s): 4.623835

Cramer-von Mises statistic: 0.01276128 with *p*-value 0.2623762

Based on Inversion of Kendall's tau method, we cannot reject the null hypothesis that the simulated frequencies have a relationship as the Clayton copula with $\Theta = 5$. But using Maximum Likelihood method, it is the opposite conclusion. It would be conservative for us not to reject the null hypothesis given the mixture of statistical test results.

2.2.2 Frank Copula

This test is to check if the Frank Copula[10] modeling in LSM is appropriate for correlation between frequencies of different lines.

Test Parameters:

- ✓ Two Lines with annual frequency Poisson ($\lambda = 96$)
- ✓ Monthly exposure: 1
- ✓ Frequency Trend: 1
- ✓ Seasonality: 1
- ✓ Accident Year: 2000
- ✓ Random Seed: 16807
- ✓ Frequency correlation: $\Theta = 8$, $n = 2$ (see footnote 10)
- ✓ # of Simulations: 1000

A simple way to compare is to draw the scatter plot for the intended copula and simulated frequency pairs. Figures 5 and 6 below show that they are of the similar patterns.

---

[9] $S_n^{(k)} = \sum_{i=1}^{n} \{C_n^{(k)}(\widehat{U}_i^{(k)}) - C_{\theta_n}^{(k)}(\widehat{U}_i^{(k)})\}^2$. Details can be found on page 6 of Kojadinovic and Yan 2010.

[10] Frank Copula: $C_\theta^n(u) = -\frac{1}{\theta}\ln(1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)\cdots(e^{-\theta u_n} - 1)}{(e^{-\theta} - 1)^{n-1}}$ $\theta > 0$ Details can be found on page 152 in Nelsen 2006,.

Figure 5. Frank Copula (8)

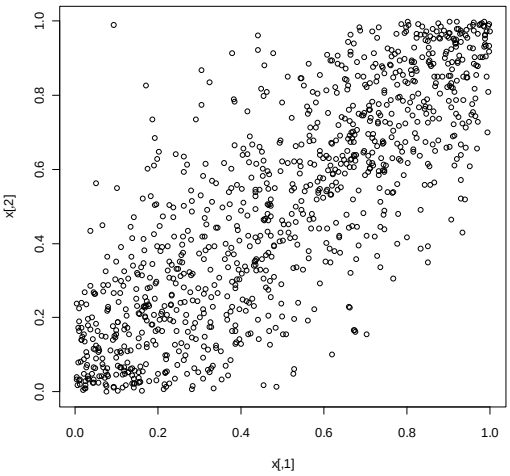
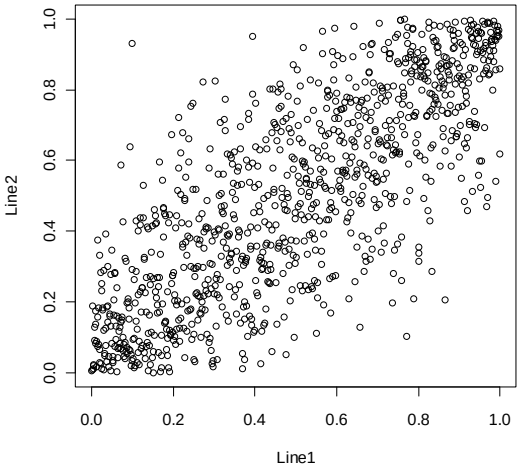


Figure 6. Simulated Frequencies



Frank copula parameter is then estimated based on simulated frequency data using two methods.

(1) The estimation is based on the maximum likelihood and a sample of size 1000.

	Estimate	Std. Error	z value	Pr(>|z|)
parameter	7.508134	0.2770857	27.09679	0

The maximized loglikelihood is 455.8911.

(2) The estimation is based on the inversion of Kendall's tau and a sample of size 1000.

	Estimate	Std. Error	z value	Pr(>|z|)
parameter	7.544506	0.3076033	24.52674	0

We can see that the model parameter **Θ** as 8 is within the 95% confidence interval based on either maximum likelihood or inversion of Kendall's tau.

Goodness-of-fit Test

(1) Using Maximum Likelihood method for parameter estimation:

Parameter estimate(s): 7.508134

Cramer-von Mises statistic: 0.01648723 with *p*-value 0.3118812

(2) Using Inversion of Kendall's tau method for parameter estimation:

Parameter estimate(s): 7.544506

Cramer-von Mises statistic: 0.01664421 with *p*-value 0.2029703

Based on those testing results, we cannot reject the null hypothesis that the simulated results are consistent with Frank Copula with Θ equal to 8.

2.2.3 Gumbel Copula

This test is to check if the Gumbel Copula[11] modeling in LSM is appropriate for correlation among frequencies of different lines.

Test Parameters:

- ✓ Two Lines with annual frequency Poisson (λ = 96)
- ✓ Monthly exposure: 1
- ✓ Frequency Trend: 1
- ✓ Seasonality: 1
- ✓ Accident Year: 2000
- ✓ Random Seed: 16807
- ✓ Frequency correlation: Θ = 6, n = 2 (see footnote 11)
- ✓ # of Simulations: 1000

A simple way to compare is to draw the scatter plot for the intended copula and simulated frequency pairs. Figures 7 and 8 below show that they are of similar patterns.

[11] Gumbel Copula: $C_\theta^n(u) = \exp(-[(-\ln u_1)^\theta + (-\ln u_2)^\theta + \cdots + (-\ln u_n)^\theta]^{1/\theta})$ $\theta \geq 1$. Details can be found on page 153 in Nelsen 2006.

Figure 7. Gumbel Copula (6)

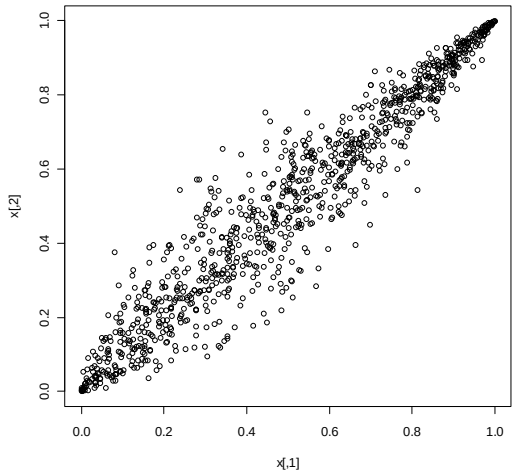
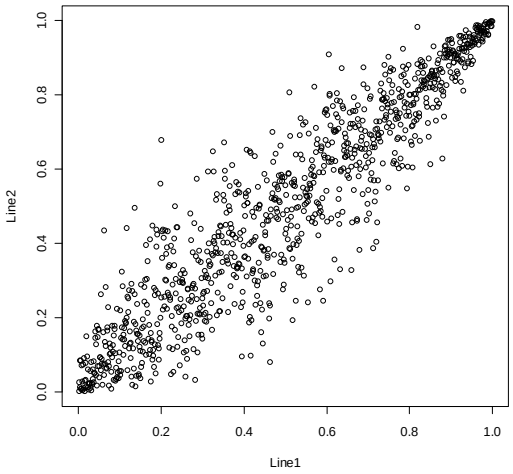


Figure 8. Simulated Frequencies



Gumbel copula parameter is then estimated based on simulated frequency data using two methods.

(1) The estimation is based on the maximum likelihood and a sample of size 1000.

	Estimate	Std. Error	z value	Pr(>|z|)
parameter	4.223043	0.1111714	37.98677	0

The maximized loglikelihood is 1038.727.

(2) The estimation is based on the inversion of Kendall's tau and a sample of size 1000.

	Estimate	Std. Error	z value	Pr(>|z|)
parameter	4.419024	0.1603205	27.56369	0

We can see that the model parameter Θ as 6 is out of the 95% confidence interval based on either maximum likelihood or inversion of Kendall's tau.

Goodness-of-fit Test

(1) Using Maximum Likelihood method for parameter estimation:

Parameter estimate(s): 4.223043

Cramer-von Mises statistic: 0.01498423 with p-value 0.1237624

(2) Using Inversion of Kendall's tau method for parameter estimation:

Parameter estimate(s): 4.419024

Cramer-von Mises statistic: 0.01063169 with *p*-value 0.2623762

Based on those testing results, we would reject the null hypothesis that the simulated results are consistent with Gumbel Copula with Θ equal to 6.

2.2.4 *t* Copula

This test is to check if the *t* Copula[12] modeling in LSM is appropriate for correlation between frequencies of different lines

Test Parameters:

- ✓ Two Lines with annual frequency Poisson (λ = 96)
- ✓ Monthly exposure: 1
- ✓ Frequency Trend: 1
- ✓ Seasonality: 1
- ✓ Accident Year: 2000
- ✓ Random Seed: 16807
- ✓ Frequency correlation: v (degree of freedom) = 5, correlation = 0.8, *n* = 2 (see footnote 12)
- ✓ # of Simulations: 1000

A simple way to compare is to draw the scatter plot for the intended copula and simulated frequency pairs. Figures 9 and 10 below show that they are of the similar patterns.

[12] *t* Copula, or Student *t* copula, $C^n_{\nu,\Sigma}(u) = T_{\nu,\Sigma}(T_\nu^{-1}(u_1),\cdots,T_\nu^{-1}(u_n))$ *ν*: degree of freedom, Σ: correlation matrix, *T*: *t* cumulative distribution function.

Figure 9. *t* Copula (dof = 5, 0.8)

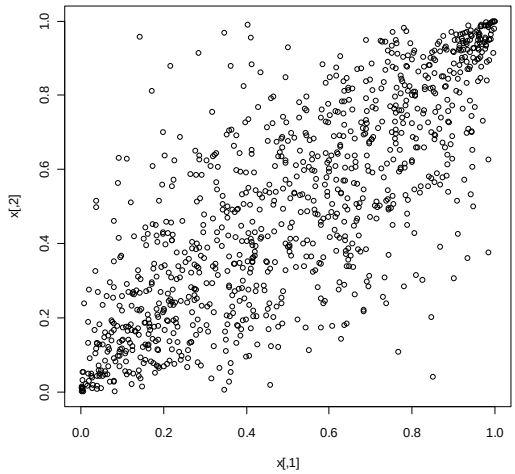
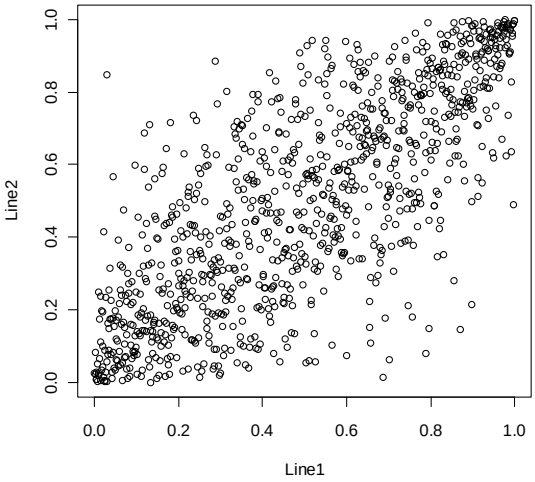


Figure 10. Simulated Frequencies



The *t* copula parameter is then estimated based on simulated frequency data using two methods.

(1) The estimation is based on the maximum likelihood and a sample of size 1000.

	Estimate	Std. Error	z value	Pr(>|z|)
parameter	0.7614685	0.01254461	60.70086	0

The maximized loglikelihood is 444.3589.

(2) The estimation is based on the inversion of Kendall's tau and a sample of size 1000.

	Estimate	Std. Error	z value	Pr(>|z|)
parameter	0.7840726	0.01343576	58.35713	0

We can see that the correlation assumption of 0.8 is within the 95% confidence interval based on inversion of Kendall's tau and within the 99% confidence interval based on maximum likelihood.

Goodness-of-fit Test

(1) Using Maximum Likelihood method for parameter estimation:

Parameter estimate(s): 0.7614685

Cramer-von Mises statistic: 0.04547016 with *p*-value 0.01485149

(2) Using Inversion of Kendall's tau method for parameter estimation:

Parameter estimate(s): 0.7840726

Cramer-von Mises statistic: 0.0259301 with *p*-value 0.04455446

Based on those testing results, it is conservative for us not to reject the null hypothesis that the simulated results are consistent with the *t* Copula that has correlation = 0.8 and degree of freedom = 5.

2.2.5 Correlation between claim size and report lag

This test is to check if the correlation between claim size and report lag in LSM is appropriately modeled.

Test Parameters:

- ✓ One Line with annual frequency Poisson ($\lambda = 120$)
- ✓ Monthly exposure: 1
- ✓ Frequency Trend: 1.05
- ✓ Seasonality: 1
- ✓ Accident Year: 2000
- ✓ Random Seed: 16807
- ✓ Payment Lag: Exponential with rate = 0.002739726, which implies a mean of 365 days.
- ✓ Size of entire loss: Lognormal with mu = 11.16636357 and sigma = 0.832549779
- ✓ Correlation between payment lag and size of loss: normal copula[13] with correlation = 0.85, dimension 2 (See footnote 13)
- ✓ # of Simulations: 10[14]

A simple way to compare is to draw the scatter plot for the intended copula and simulated frequency pairs. Figures 11 and 12 below show that they are of similar patterns.

---

[13] Normal Copula, or Gaussian Copula, $C_{\Sigma}^{n}(u) = \Phi_{\Sigma}(\Phi^{-1}(u_1),\cdots,\Phi^{-1}(u_n))$ Σ: correlation matrix

Φ: normal cumulative distribution function. Details can be found on pages 43-54 of Li 2000.

[14] The reason to use 10 simulations instead of 1000 simulations is that 120 claims are expected (Frequency distribution with l = 120) in each simulation. The total expected number of pairs of data is 1200 with 10 simulations for correlation analysis.

Figure 11. Normal Copula (0.85)

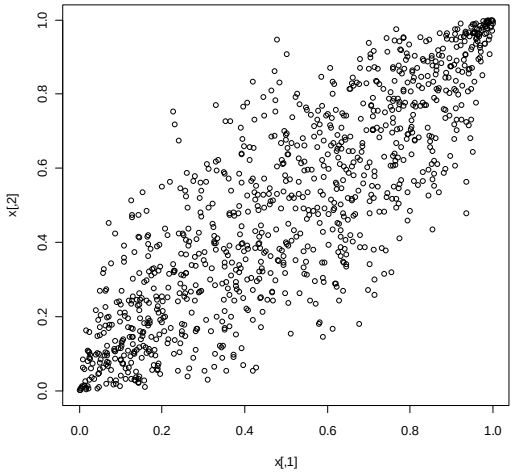
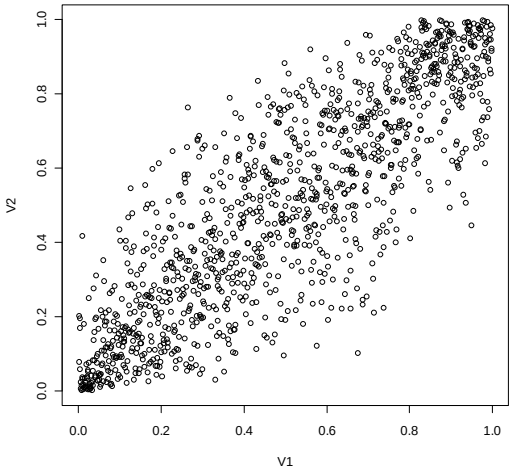


Figure 12. Simulated claim size vs. report lag



The normal copula parameter is then estimated based on simulated frequency data using two methods.

(1) The estimation is based on the maximum likelihood and a sample of size 1000.

	Estimate	Std. Error	z value	Pr(>|z|)
rho.1	0.8317376	0.006878922	120.9110	0

The maximized loglikelihood is 694.6756.

(2) The estimation is based on the inversion of Kendall's tau and a sample of size 1000.

	Estimate	Std. Error	z value	Pr(>|z|)
parameter	0.8538963	0.007917961	107.8430	0

We can see that the correlation assumption (0.85) is within the 95% confidence interval based on inversion of Kendall's tau.

Goodness-of-fit Test

(1) Using Maximum Likelihood method for parameter estimation:

Parameter estimate(s): 0.8317376

Cramer-von Mises statistic: 0.06218935 with p-value 0.004950495

(2) Using Inversion of Kendall's tau method for parameter estimation:

Parameter estimate(s): 0.8538963

Cramer-von Mises statistic: 0.02898052 with *p*-value 0.01485149

Based on those testing results, we would reject the null hypothesis at the significance level larger than 1.5% that the simulated results are consistent with Normal Copula that has correlation = 0.85. The difference in the value of correlation coefficients between model input and model output is not small. However, the simulated data still have a strong correlation as intended.

2.3 Severity trend

This test is to check if the severity trend in LSM is modeled as intended.

Test Parameters:

- ✓ One Line with annual frequency Poisson ($\lambda = 96$)
- ✓ Monthly exposure: 1
- ✓ Frequency Trend: 1
- ✓ Seasonality: 1
- ✓ Accident Years: 2000 to 2005
- ✓ Random Seed: 16807
- ✓ Size of entire loss: Lognormal with mu = 11.16636357 and sigma = 0.832549779
- ✓ Severity Trend: 1.5
- ✓ # of Simulations: 300

Figure 13 shows the mean value of loss size over time. There is a clear consistent trend. Figure 14 shows that Seasonal Decomposition of Time Series by Loess (STL),[15] which decomposes a time series into seasonal, trend, and irregular components using loess.[16] It is very obvious there is no seasonality and there exists an upward sloping trend. The residual errors behave like white noise.

[15] Package *stats* version 2.12.0, R Documentation, Seasonal Decomposition of Time Series by Loess. A description of STL is available in Cleveland et al., 1990.
[16] Loess stands for Locally Weighted Regression Fitting.

Figure 13. Mean Loss Size

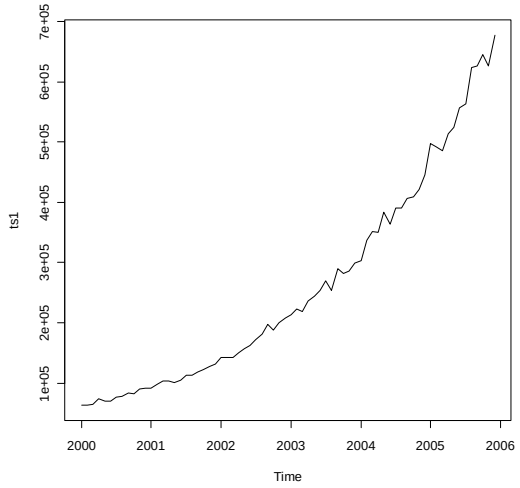
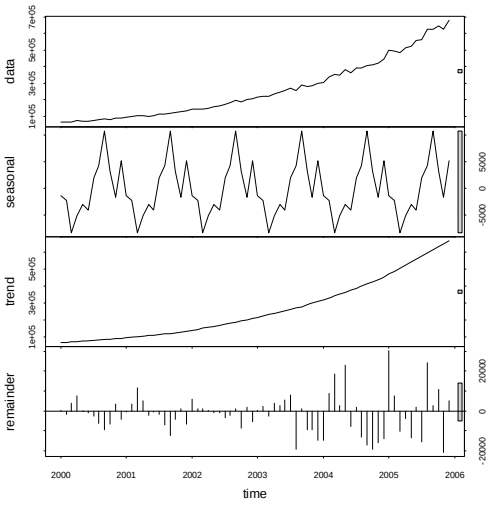


Figure 14. STL



Based on the log of mean loss size, a linear regression that estimates the linear trend factor supports our assumptions.

Log(Mean Loss Size) = Intercept + trend * (time – 2000) + error term

We get the following results using R.

Residuals:

Min	1Q	Median	3Q	Max
-0.051579	-0.023194	-0.007886	0.023918	0.078750

Coefficients:

	Estimate	Std. Error	*t* value	Pr(>|*t*|)
(Intercept)	11.034162	0.007526	1466.1	<2*e*-16
trend	0.405552	0.002196	184.7	<2*e*-16

Residual standard error: 0.03226 on 70 degrees of freedom

Multiple R-squared: 0.998, Adjusted R-squared: 0.9979

F-statistic: 3.412e+04 on 1 and 70 DF, *p*-value: < 2.2*e*-16

We can see that the *t* test shows that the trend is not equal to 0 at a significant level less than 0.1%. The high adjusted R2 and the *F* test also show that the trend is obvious. The trend factor

0.405552 is based on the log of the mean loss size and is equivalent to the trend factor of 1.5 for loss size (exp(0.405552) = 1.50013). This is also our model input. Figure 15 shows a good fitting of the regression. Residual graph (Figure 16) shows a white noise pattern. Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) also support the existence of linear trend.

Figure 15. Trend fitting

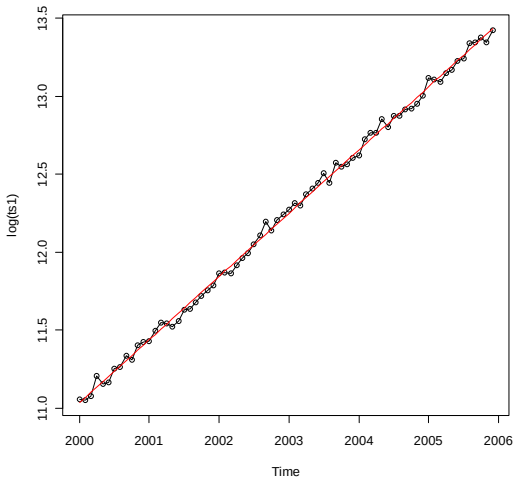
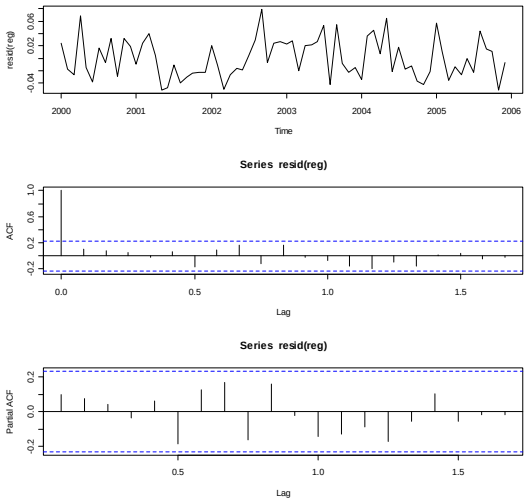


Figure 16. Residual, ACF, PACF



2.4 Alpha in Severity Trend

This test checks if the alpha that determines the persistency of the force of trend for severity in LSM is modeled as intended. As described in LSM,[17] the cumulative trend amounts (cum) are calculated first and then the trend multiplier is calculated as

$$trend = (cum_{acc\_date})\left(\frac{cum_{pmt\_date}}{cum_{acc\_date}}\right)^{\alpha} = (cum_{acc\_date})^{1-\alpha}(cum_{pmt\_date})^{\alpha}.$$

Test Parameters:

- ✓ One Line with annual frequency Poisson (λ = 96)
- ✓ Monthly exposure: 1
- ✓ Frequency Trend: 1
- ✓ Seasonality: 1

[17] CAS Loss Simulation Model Working Party Summary Report, pages 66-67.

- ✓ Accident Years: 2000 to 2001
- ✓ Random Seed: 16807
- ✓ Size of entire loss: Lognormal with mu = 11.16636357 and sigma = 0.832549779
- ✓ Severity Trend: 1.5
- ✓ Alpha: 0.4
- ✓ # of Simulations: 1000

We choose the sample loss payments with report date during the 1st month and payment date during the 7th month.

Therefore, the severity trend multiple is $(1.5^{1/12})^{(1-0.4)} \cdot (1.5^{7/12})^{0.4} \approx 1.122$ for those chosen claims.

The expected loss size is $1.122 \cdot e^{11.166+0.83255^2/2} \approx 112{,}175$.

The histogram and QQ plot show that the fit is not perfect, but not too far away.

Figure 17. Histogram of severity

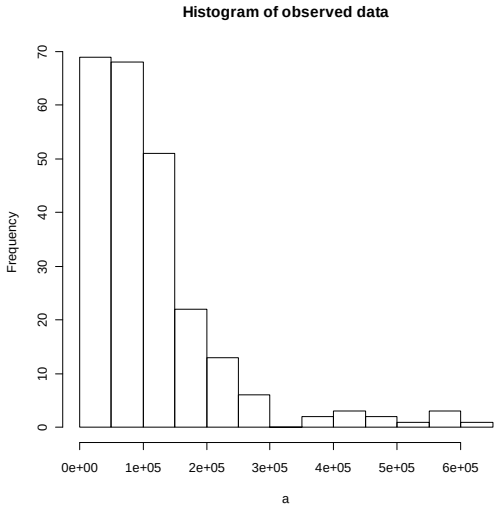


Figure 18. QQ plot of severity

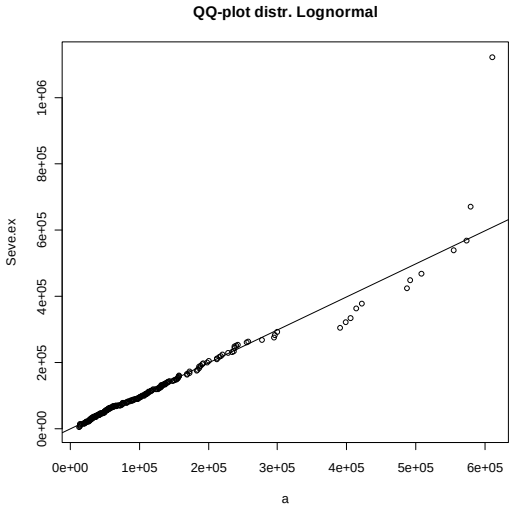
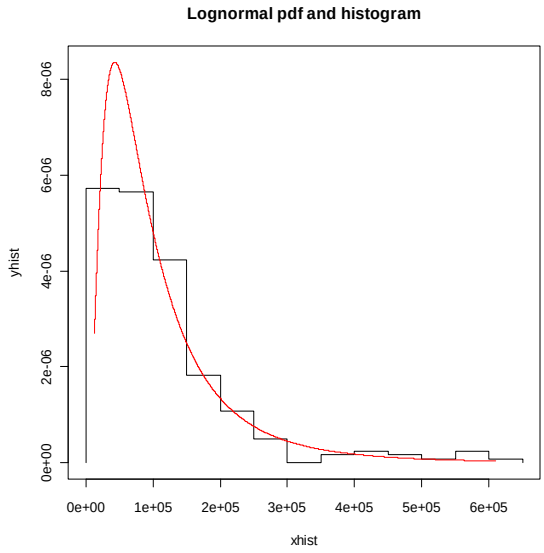


Figure 19. Histogram and fitted probability density function



Maximum likelihood estimation gives us the following fitted parameters and standard deviation. The mean value of severity is 113,346. When only volatility of meanlog estimation is considered, the mean loss derived by model input is within 95% confidence interval.

	meanlog	sdlog
Estimation	11.31595927	0.80279226
Standard Deviation	0.05171240	0.03656619

Results of Kolmogorov-Smirnov test and Anderson-Darling normality test support the lognormal distribution of the sampled payments.

One-sample Kolmogorov-Smirnov test

$D = 0.0405$, p-value = 0.8249

alternative hypothesis: two-sided

Anderson-Darling normality test

$A = 0.4114$, p-value = 0.3384

2.5 Case Reserve Adequacy Distribution

In the LSM, the case reserve adequacy distribution parameters are intended to model

characteristics of an insurer's case loss reserving process. For example, some insurers set a nominal reserve until a claim is investigated while others may set up a formula or "average" reserve initially. The ultimate claim value may be the same in both cases, but the timing and amount of the reserve changes may be quite different. The case reserve adequacy distribution attempts to model this process by generating case reserve adequacy ratio at each valuation date. Case reserve is determined by multiplying the generated final claim amount by case reserve adequacy ratio.

Notice that, for simulated data, the case reserve adequacy parameters do not affect the ultimate claim value. However, in determining LSM parameters from real data where some of the accident years are not fully developed, the case reserve adequacy assumption may be crucial.

This test is to check if the *X*% time point case reserve adequacy distribution in LSM is modeled as intended. We choose the 40% time point[18] in this paper.

Test Parameters

- ✓ One Line with annual frequency Poisson ($\lambda = 96$)
- ✓ Monthly exposure: 1
- ✓ Frequency Trend: 1
- ✓ Seasonality: 1
- ✓ Accident Years: 2000 to 2001
- ✓ Random Seed: 16807
- ✓ Size of entire loss: Lognormal with mu = 11.16636357 and sigma = 0.832549779
- ✓ 40% Case Reserve: Lognormal with mu = 0.25 and sigma = 0.05
- ✓ Severity Trend: 1
- ✓ $P(0) = 0.4$
- ✓ Est $P(0) = 0.4$
- ✓ # of Simulations: 8[19]

From the test assumption, we know that the mean 40% case reserve adequacy ratio is

[18] The 40% time point is the date that is equal to the 60% Report Date + 40% Final Payment Date.
[19] Similar reason as indicated in footnote 14 as the number of simulated claims is large enough for statistical testing with 8 simulations.

$e^{0.25+0.05^2/2} \approx 1.2856$. The transaction output is used to calculate the case reserve at 40% of payment lag using linear interpolation method. Those values are then used for testing purposes.

The histogram, QQ plot, and probability density function show that the fit is not good.

Figure 20. Histogram of severity

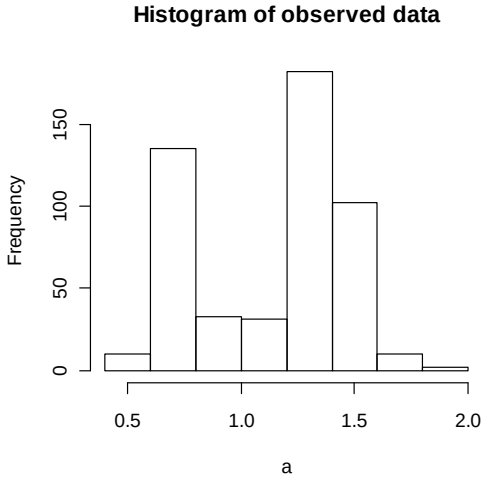


Figure 21. QQ Plot of severity

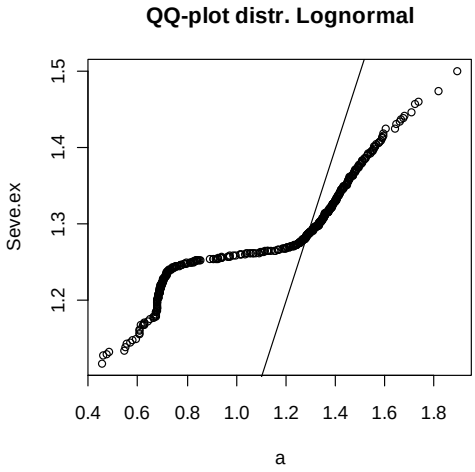
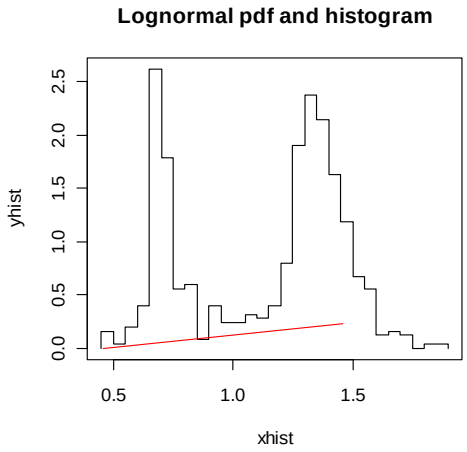


Figure 22. Histogram and fitted probability density function



Maximum likelihood estimation gives us the following fitted parameters and standard deviation. The mean value of severity is 1.141. When only volatility of meanlog estimation is considered, the mean loss derived by model input is within 95% confidence interval.

	meanlog	sdlog
Estimation	0.07973950	0.32269631

Standard Deviation	0.01435980	0.01015391

Results of Kolmogorov-Smirnov test and Anderson-Darling normality test do not support the lognormal distribution of the sampled case reserve adequacy.

One-sample Kolmogorov-Smirnov test

$D = 0.3869$, p-value < 2.2e-16

Anderson-Darling normality test

$A = 33.2183$, p-value < 2.2e-16

Model input and output are not consistent for both the distribution type and the fitted parameters. In the simulation, valuation dates of each claim are generated based on an assumption of waiting period (inter-valuation lag assumption). Before the final payment, case reserve is generated on the simulated valuation dates. Since valuation dates are randomly generated, it often does not coincide with the 40% time point. In those cases, linear interpolation method is used to get case reserve ratio at 40% time point for testing. On the first valuation date, i.e., the report date, a case reserve of 2,000 will be allocated for each claim without any adjustment related to the claim size. If the second valuation date happens after 40% time point, it is clear that linear interpolation method can give us false estimation of what is assumed in the model inputs. Therefore, there is no confident conclusion about whether the model is correct or not.

A way to overcome this is to change the way in which transaction date is determined. In current coding, report date and final settlement date are generated before transaction date and case reserves are generated. We can set a few transaction dates as report date + X%(payment date – report date) instead of generating them based on waiting period distribution assumption. X% could be 40%, 70%, and 90% to be consistent with current case reserve adequacy model input setting. In this way, linear interpolation is not needed anymore and the output data we got are also easier for testing the model and reserve methods.

3. REAL DATA AND SIMULATED DATA

Marine claim data are used for distribution fitting, trend analysis, and correlation analysis. Those estimated distributions and parameters could be input for LSM to generate stochastic claim data. Based on those claim data, reserve methods can be tested and evaluated. Unfortunately, paid loss

history data are not available in this example and Meyers' Approach[20] cannot be applied due to the lack of details. Below is a snapshot of the claim data used in this section. It has two product lines: Property and Liability. The data period is from 2006 to 2010. The number of accidents is 317 for Property Insurance and 428 for Liability Insurance. All the claims are closed with a final payment.

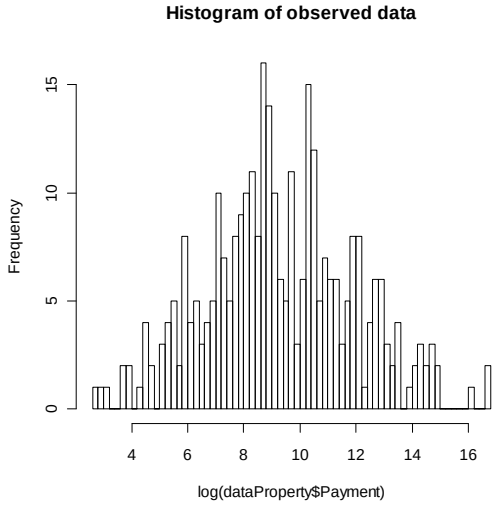
Accident Date	Payment data	Line	Final Payment
12/31/2006	3/30/2008	Property	249
5/1/2006	11/27/2006	Property	16,293
1/22/2010	4/22/2010	Property	65,130
1/22/2006	8/20/2006	Liability	38,544
7/27/2010	2/22/2011	Liability	13,206

3.1 Property Line

Fit the severity

1. Draw a histogram of logarithm of payment to find out the most appropriate claim-size distribution type. Lognormal distribution seems to be a good candidate for describing claim size.

Figure 23. Histogram of Log (Claim Size)



2. Use lognormal distribution fitting for claim size.

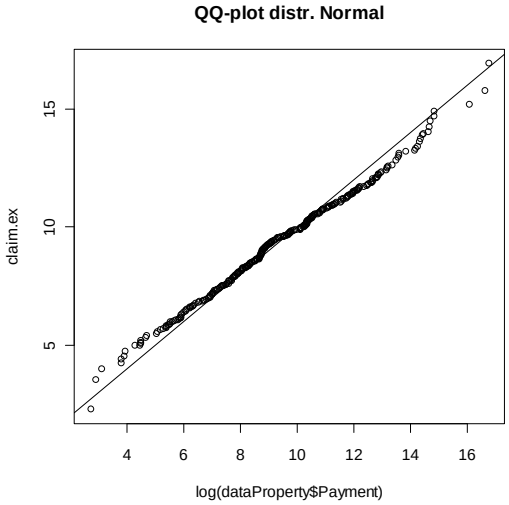
	meanlog	sdlog
Estimation	9.2848522	2.6269670

[20] CAS Loss Simulation Model Working Party Summary Report, pages 7-8.

Standard Deviation 0.1484850 0.1049947

3. Use a QQ plot to check the fitting. It is not a perfect fitting but this is probably the best we can achieve.

Figure 24. QQ Plot of Log(Claim Size)



Fit the frequency

4. Draw a time series of frequency data and conduct a Seasonal Decomposition of Time Series. There is no strong evidence of linear trend and seasonality during this period.

Figure 24. Frequency

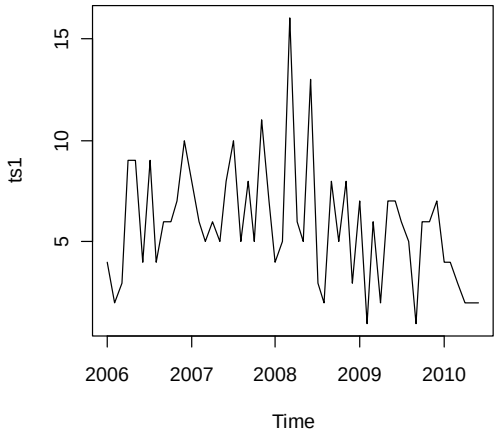
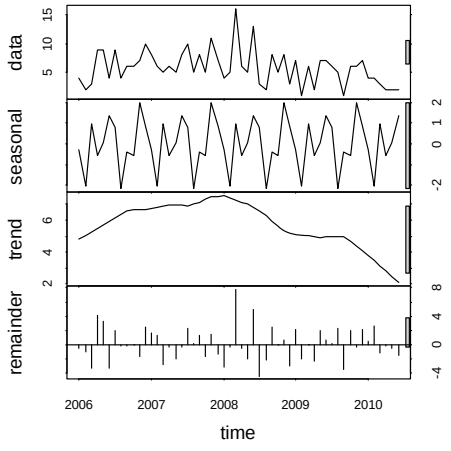


Figure 25. STL



5. Perform a linear regression for trend analysis.

Log(Monthly Frequency) = Intercept + trend * (time – 2006) + error term.

Residuals:

Min	1Q	Median	3Q	Max
-1.48135	-0.36849	0.04697	0.38654	1.15768

Coefficients:

	Estimate	Std. Error	*t* value	Pr(>|*t*|)
(Intercept)	1.93060	0.15164	12.732	<2*e*-16
trend	-0.14570	0.05919	-2.462	0.0172

Residual standard error: 0.5649 on 52 degrees of freedom.

Multiple R-squared: 0.1044, Adjusted R-squared: 0.08715.

F-statistic: 6.06 on 1 and 52 DF, *p*-value: 0.01718.

Figure 26. Trend fitting

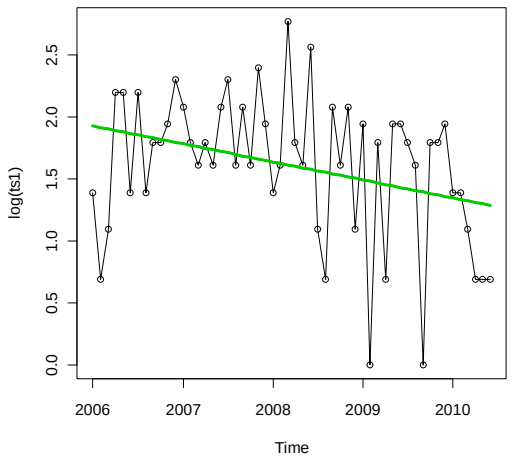
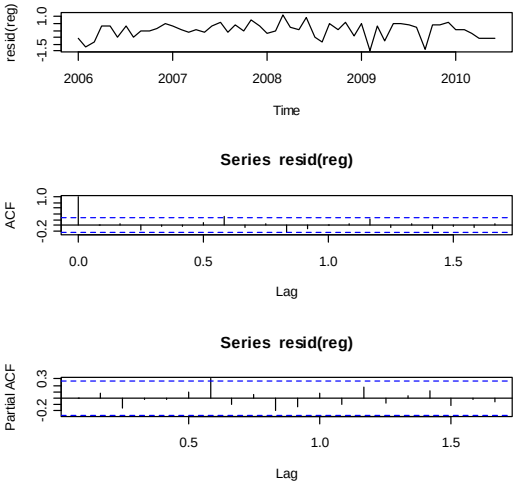


Figure 27. Residual, ACF, PACF



6. Detrend the frequency and fit to the frequency distribution.

It looks like that lognormal distribution fits the detrended data better.

	meanlog	sdlog
Estimation	9.5539259	3.1311762

Standard Deviation	0.4260991	0.3012976

Figure 28. Histogram of detrend freq.

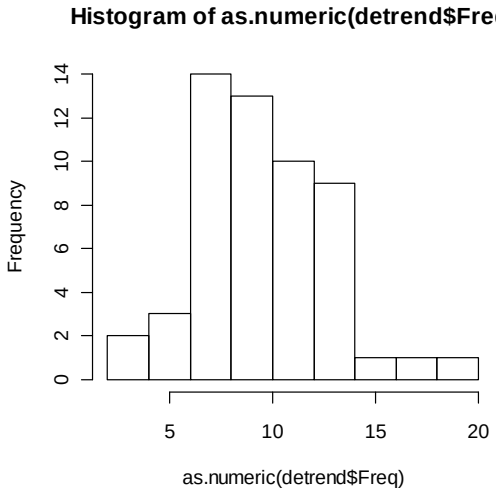
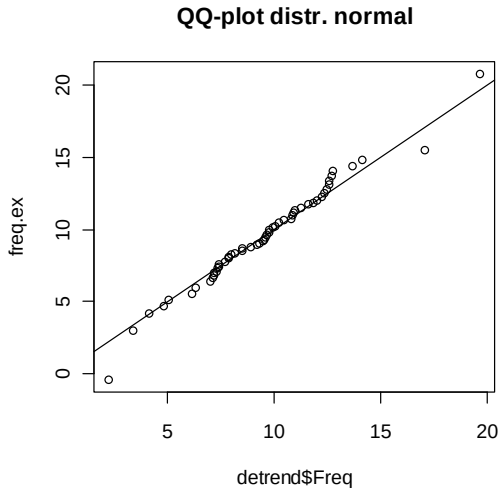


Figure 29. QQ Plot of detrend freq.



The Kolmogorov-Smirnov test result also supports lognormal distribution assumption.

One-sample Kolmogorov-Smirnov test is as follows:

$D = 0.0814$, p-value = 0.8384.

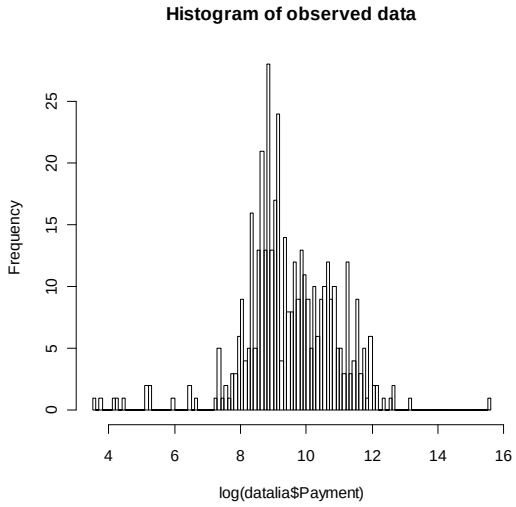
Therefore, we have all the parameters for frequency and severity distribution and trend of frequency for property line.

3.2 Liability Line

Fit the severity

1. Draw a histogram of the logarithm of payments to find out candidates for the distribution type of the claim size. Lognormal distribution seems to be a good candidate for describing claim size.

Figure 30. Histogram of Log (Claim Size)

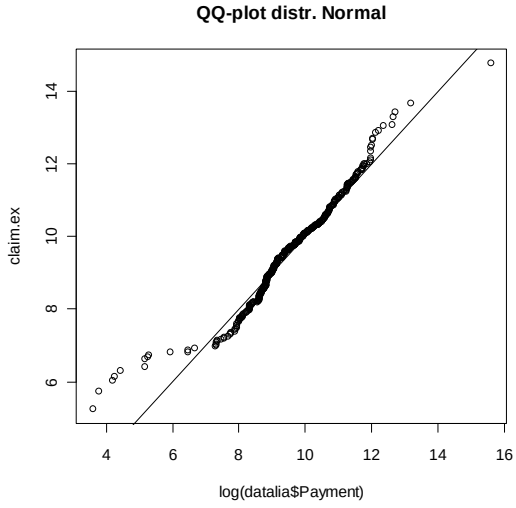


2. Use lognormal distribution fitting for claim size:

	meanlog	sdlog
Estimation	9.50314718	1.42545383
Standard Deviation	0.06890191	0.04872101

3. Use a QQ plot to check the fitting. The fit is not good at the low end.

Figure 31. QQ Plot of Log (Claim Size)



Fit the frequency

4. Draw a time series of frequency data and conduct a Seasonal Decomposition of Time Series. There are no strong evidence of linear trend and seasonality during this period.

Figure 32. Frequency

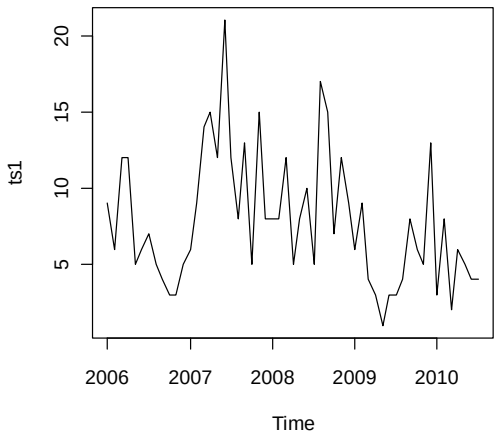
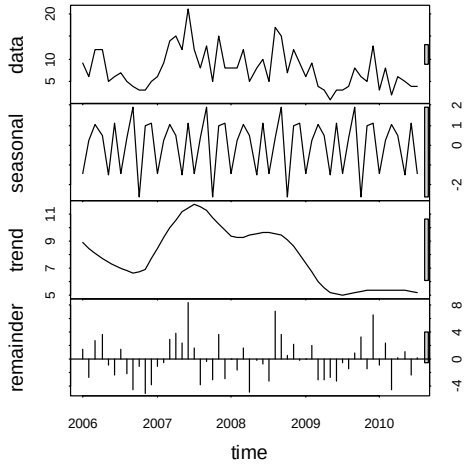


Figure 33. STL



5. Use linear regression for trend analysis.

Log(Monthly Frequency) = Intercept + trend * (time – 2006) + error term.

Residuals:

Min	1Q	Median	3Q	Max
-1.74504	-0.36590	0.09695	0.42571	1.03941

Coefficients:

	Estimate	Std. Error	*t* value	Pr(>|*t*|)
(Intercept)	2.3330	0.2060	11.327	9.03*e*-16
trend	-0.1357	0.0587	-2.311	0.0247

Residual standard error: 0.5759 on 53 degrees of freedom.

Multiple R-squared: 0.09158, Adjusted R-squared: 0.07444.

F-statistic: 5.343 on 1 and 53 DF, *p*-value: 0.02472.

Figure 34. Trend fitting

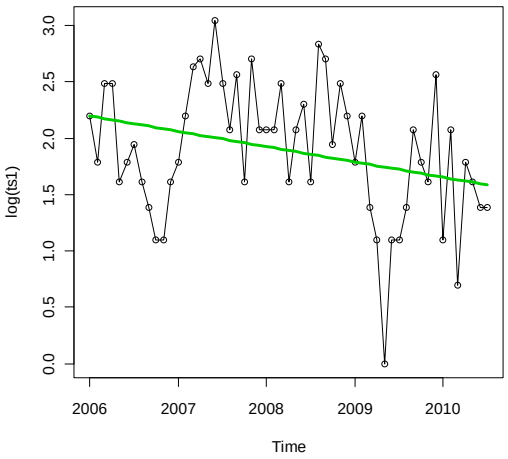
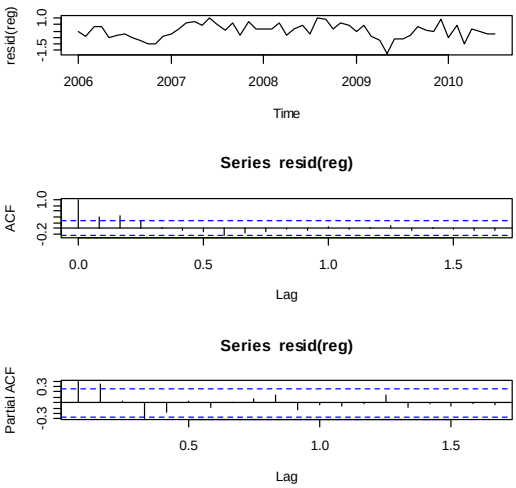


Figure 35. Residual, ACF, PACF



6. Detrend the frequency and fit it to the frequency distribution. It looks like lognormal distribution fits the detrended data better.

	meanlog	sdlog
Estimation	2.35724617	0.38449461
Standard Deviation	0.05184524	0.03666012

Figure 36. Histogram of detrend freq.

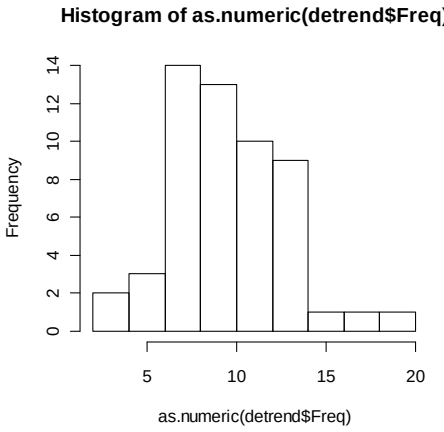
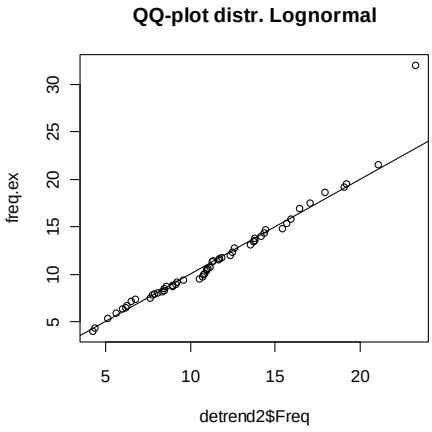


Figure 37. QQ Plot of detrend freq.



The Kolmogorov-Smirnov test also supports the assumption of lognormal distribution.

One-sample Kolmogorov-Smirnov test

$D = 0.0981$, p-value = 0.6293,

therefore, we have all the parameters for frequency and severity distribution and trend of frequency for liability line.

3.3 Correlation

First, we calculate the correlation coefficient between the two lines' frequencies.

	Line1	Line2
Line1	1.0000000	0.2800634
Line2	0.2800634	1.0000000

The Frank copula parameter is then estimated based on simulated frequency data using two methods. Other types of copula can and should also be used to determine the best fit.

(1) The estimation is based on the maximum likelihood and a sample of size 55.

	Estimate	Std. Error	z value	$Pr(>\|z\|)$
rho.1	1.512390	0.854729	1.769438	0.07682074

The maximized loglikelihood is 1.533443.

(2) The estimation is based on the inversion of Kendall's tau and a sample of size 55.

	Estimate	Std. Error	z value	$Pr(>\|z\|)$
parameter	1.325654	0.918666	1.443020	0.1490148

A simple way to compare is to draw the scatter plot for the intended copula and simulated frequency pairs. The figures below show that they are of the similar patterns.

Figure 38. Frank Copula (1.325654)

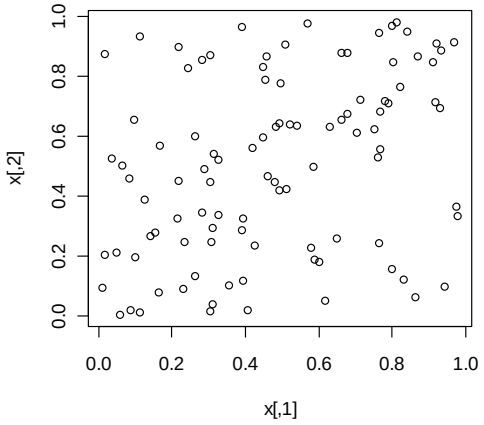
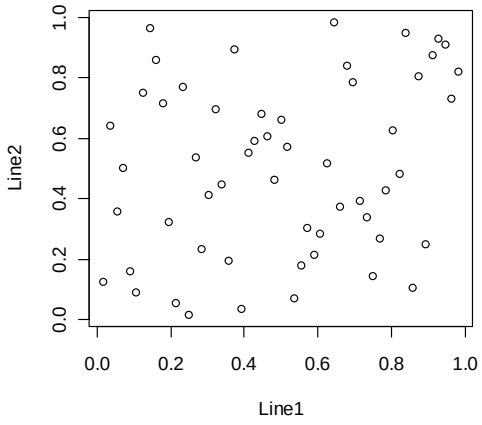


Figure 39. Simulated Frequencies



Goodness-of-fit Test

(1) Using Maximum Likelihood method for parameter estimation:

Parameter estimate(s): 1.512390

Cramer-von Mises statistic: 0.02652859 with *p*-value 0.3514851

(2) Using Inversion of Kendall's tau method for parameter estimation:

Parameter estimate(s): 1.325654

Cramer-von Mises statistic: 0.02780636 with *p*-value 0.4009901

Based on those testing results, we would not reject the null hypothesis that the real data are consistent with the Frank copula with parameter 1.325654.

4. MODEL ENHANCEMENT

4.1 Two-State, Regime-Switching Distribution

Sometimes in the real world, one single distribution may not be able to represent the past frequency and severity experience data well. There are normally three reasons behind this:

(1) Structural change: some exogenous impact causes distribution (distribution type and/or parameters) to change drastically during a time period and last thereafter.

(2) Cyclical pattern: The business may have some cyclical characteristics. A normal case is the

underwriting cycle where for a certain period of time, the claim frequencies and/or severities will increase a lot and after that, it will return to a lower level.

(3) Idiosyncratic risk: The claim data cannot be described by available distribution types. The randomness due to idiosyncratic characteristics makes it hard to fit a certain distribution along the time.

In the LSM, if the structural change is predicted, it can be incorporated by setting frequency/severity trend and even using different severity distributions for different months when the distribution type is expected to change.

However, the current model does not have a direct solution for incorporating the cyclical pattern and idiosyncratic characteristics. In order to add the flexibility of LSM to handle the modeling of them, a categorical variable is included to enable setting parameters/distribution type for different states. For all the variables that are modeled as distribution, two-state regime-switching capability is built in to enable moving from one state to the other state. A two-state, regime-switching model is commonly used in time series analysis. Here state means the status of the object such as frequency and/or severity that is described as a certain distribution.

The user can set two distributions with different parameters and determine the transition probability from one state to another. At the beginning of each month, the model will determine which distribution/state it will be for this month based on the transition matrix.

Let's take frequency distribution as an example to illustrate the process in the model.

Input

- ✓ State 1: Poisson Distribution ($\lambda = 120$)
- ✓ State 2: Negative Binomial Distribution (size = 36, prob = 0.5)
- ✓ Assume the trend, monthly exposure, and seasonality are all 1
- ✓ State 1 persistency: 0.5
- ✓ State 2 persistency: 0.7
- ✓ Seed: 16807

Markov Chain Transition Matrix

State persistency represents the probability that the variable will remain in the same state next

month. Here we assume the transition follows discrete Markov Chain.[21]. It means that the state of next month only depends on the state of the current month but does not depend on the state before the current month. In other words, it is not path-dependent.

Another thing that needs to be determined is the state of the first month. In the current model setting, steady-state probabilities are used. Let's define some variables first:

- ✓ P_{11}: state 1 persistency, the probability that the state will be 1 next month given that it is 1 this month.
- ✓ P_{12}: the probability that the state will be 2 next month given that it is 1 this month.
- ✓ P_{21}: the probability that the state will be 1 next month given that it is 2 this month.
- ✓ P_{22}: state 2 persistency, the probability that the state will be 2 next month given that it is 2 this month.
- ✓ Π_1: steady probability of state 1.
- ✓ Π_2: steady probability of state 2.

We have the following relationship held.

$$\begin{pmatrix}\Pi_1 & \Pi_2\end{pmatrix}\begin{pmatrix}P_{11} & P_{12}\\ P_{21} & P_{22}\end{pmatrix}=\begin{pmatrix}\Pi_1 & \Pi_2\end{pmatrix}$$

$$P_{11}=1-P_{12}$$

$$P_{21}=1-P_{22}$$

$$\Pi_1+\Pi_2=1$$

We can then derive the steady-state probabilities Π_1 and Π_2 based on state persistencies P_{11} and P_{22}.

$$\Pi_1=\frac{1-P_{22}}{2-P_{11}-P_{22}}=\frac{1-0.7}{2-0.5-0.7}=0.375$$

$$\Pi_2=\frac{1-P_{11}}{2-P_{11}-P_{22}}=\frac{1-0.7}{2-0.5-0.7}=0.625$$

Calculation Steps

(1) Generate uniform random number $randf_0$ on range [0,1].

[21] http://en.wikipedia.org/wiki/Markov_chain

(2) If $randf_0 < \Pi_1$, state of first month state is 1, else, it is 2.

(3) Generate uniform random number $randf_i$ on range [0,1].

(4) For previous month state I, if $randf_i < P_{i1}$, then state is 1, else it is 2.

(5) Repeat step 3 and 4 until the end of the simulation is reached.

Table 1 shows the two-state, regime-switching result for the first simulation.

Table 1. Two-State, Regime-Switching Example

Random Number (RN)	State	Criteria
0.634633548790589	2	RN>0.375
0.801362191326916	1	RN>0.7
0.529508789768443	2	RN>0.5
0.0441845036111772	2	RN<0.7
0.994539848994464	1	RN>0.7
0.21886122901924	1	RN<0.5
0.0928565948270261	1	RN<0.5
0.797880138037726	2	RN>0.5
0.129500501556322	2	RN<0.7
0.24027365935035	2	RN<0.7
0.797712686471641	1	RN>0.7
0.0569291599094868	1	RN<0.5

Based on those generated frequency states, the claim and transaction are populated. This enhancement is intended for frequency and severity distribution although the flexibility is given to all the variables that are modeled as distribution in the LSM.

4.2 Testing

The following model setting is used for testing two-state, regime-switching feature.

Test Parameters:

- ✓ Accident Year: 2000
- ✓ Random Seed: 16807
- ✓ # of Simulations: 300
- ✓ Frequency correlation: Normal Copula with correlation as 95%

Line 1

Annual frequency:

- ✓ State 1: Poisson (λ = 120), State 2: Negative Binomial (Size = 36, prob = 0.5)
- ✓ State 1 persistency: 0.15
- ✓ State 2 persistency: 0.9. It is equivalent to Π_1 = 10.53% and Π_2 = 89.47%. We can consider state 2 as the long-term normal case while state 1 is the short period where the cases of claim increase a lot compared to state 1.
- ✓ Monthly exposure: 1
- ✓ Frequency Trend: 1
- ✓ Seasonality: 1

Size of entire loss

- ✓ State 1: Lognormal with mu = 10 and sigma = 0.832549779
- ✓ State 2: Lognormal with mu = 2 and sigma = 0.832549779
- ✓ State 1 persistency: 0.3
- ✓ State 2 persistency: 0.8. It is equivalent to Π_1 = 22.22% and Π_2 = 77.78%.
- ✓ Severity Trend: 1
- ✓ $P(0) = 0$
- ✓ Est $P(0) = 0$

Line 2

Annual frequency:

- ✓ State 1: Poisson (λ = 120), State 2: Negative Binomial (Size = 36, prob = 0.5)
- ✓ State 1 persistency: 0.2
- ✓ State 2 persistency: 0.9. It is equivalent to Π_1 = 11.11% and Π_2 = 88.89%
- ✓ Monthly exposure: 1
- ✓ Frequency Trend: 1
- ✓ Seasonality: 1

Size of entire loss:

- ✓ Lognormal with mu = 10 and sigma = 0.832549779

- ✓ Severity Trend: 1
- ✓ $P(0) = 0$
- ✓ Est $P(0) = 0$

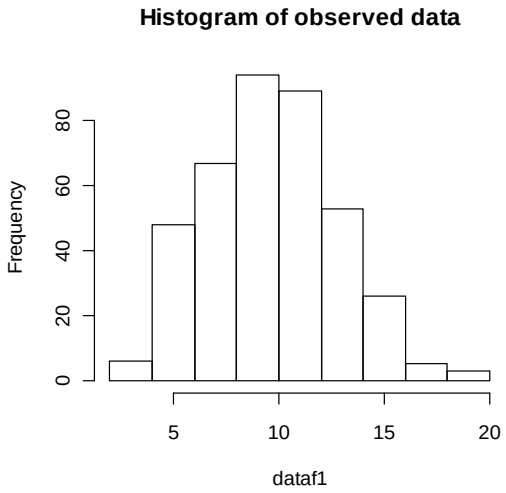
4.2.1 Frequency

We split the claim data according to the state of the monthly frequency and test whether the distribution for each state follows our model assumption.

State 1

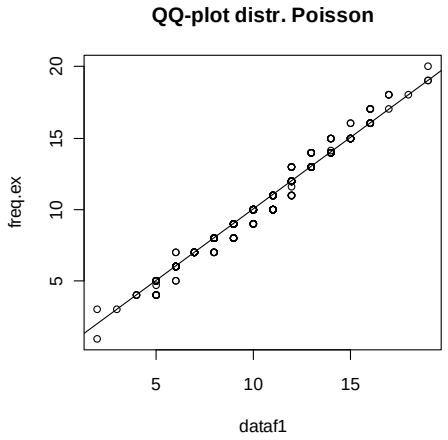
First, we draw a histogram of the simulated frequency data to give intuition of the distribution type.

Figure 40. Histogram of simulated frequency data (State 1)



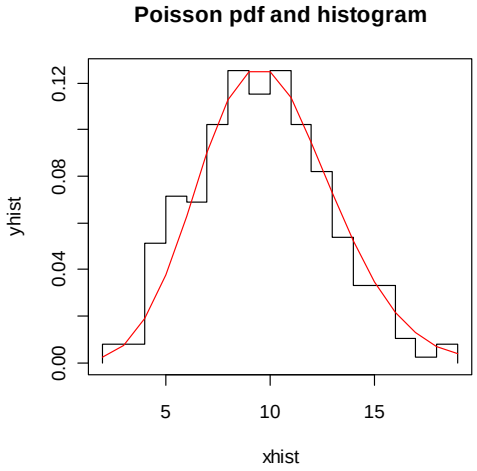
A QQ plot would also be a straightforward way to compare the simulated results with the intended distribution – Poisson (lambda = 10). In Figure 41, we can see that it is a good fit.

Figure 41. QQ Plot – Simulated results vs. Poisson (lambda = 10)



Comparing the probability distribution functions also gives us a vivid illustration of the fit.

Figure 42. PDF – simulated vs. assumption



Goodness-of-fit test using Pearson's Chi-squared statistic is performed. The results disallow rejecting the null hypothesis that the simulated frequencies follow a Poisson distribution.

Goodness-of-fit test for Poisson distribution

	X^2	df	*P*(> *X*^2)
Pearson	15.30052	19	0.703315

In addition, using maximum likelihood (ML) method to fit the Poisson distribution and calculate the likelihood Ratio statistics implies the same conclusion.

Goodness-of-fit test for Poisson distribution

	X^2	df	*P*(> *X*^2)
Likelihood Ratio	20.27080	17	0.260613

Using ML method gives us an estimation of the parameters as follows:

	lambda
Estimation	10.1329923
Standard deviation	0.1609832

Comparing with our LSM input: lambda = 120, which implies a monthly frequency as Poisson distribution with lambda = 10. We can see that at significance level of 5%, the confidence interval for size is (9.82, 10.45), which includes the model input (lambda = 10).

Two-sample Kolmogorov-Smirnov test

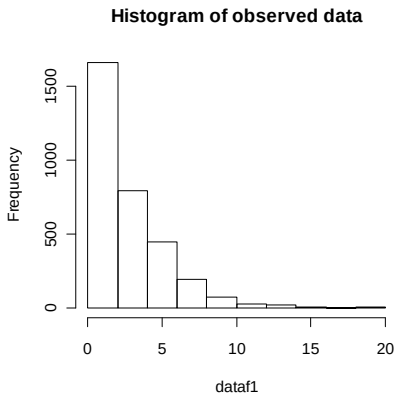
$D = 0.0411$, p-value = 0.7286

The Kolmogorov-Smirnov test also shows a reliable fit. Those results together with the goodness-of-fit tests indicate that simulated frequencies are Poisson distribution.

State 2

Firstly, we draw a histogram of the simulated frequency data to have an indication of the distribution type.

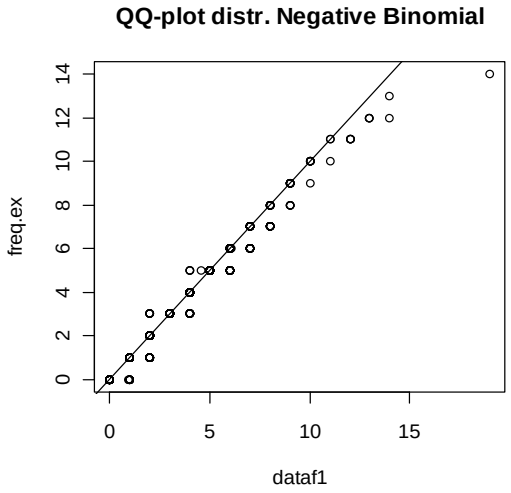
Figure 43. Histogram of simulated frequency data (State 2)



A QQ plot would also be a straightforward way to compare the simulated results with the intended distribution – Negative Binomial (size = 3, prob = 0.5). From the figure below, we can see

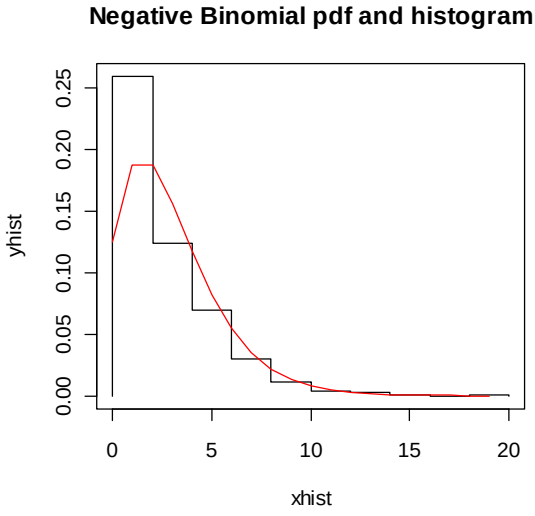
that it is a good fit. The expected frequency distribution in the LSM has a slightly shorter tail than the simulated results.

Figure 44. QQ Plot – Simulated results vs. Negative Binomial (size = 3, prob = 0.5)



Comparing the probability distribution function also shows fit below.

Figure 45. PDF – simulated vs. assumption



Goodness-of-fit test using Pearson's Chi-squared statistic is performed. The results allow us to reject the null hypothesis that the simulated frequencies follow negative binomial distribution.

Goodness-of-fit test for nbinomial distribution

	X^2	df	$P(> X$^2)
Pearson	30.75979	19	0.042890443

In addition, using maximum likelihood (ML) method to fit the Poisson distribution and calculate the likelihood ratio statistics also implies the same conclusion.

Goodness-of-fit test for Poisson distribution

	X^2	df	$P(> X$^2)
Likelihood Ratio	32.36216	16	0.008968028

Using ML method gives us an estimation of the parameters as follows:

	size	mu
Estimation	2.78375646	3.00250312
Standard deviation	0.14274338	0.04418285

The estimated value gives us size = 2.78 and prob = 0.48. The derived variance is 6.24

Where prob = size/(size+mu) and variance = mu + $mu^2/size$[22]

Our LSM inputs of size = 36 and prob =0.5 implies a monthly frequency as negative binomial distribution with size = 3, prob = 0.5. In comparison to estimated parameters based on simulated frequencies, they are not too far away.

Those results are somewhat consistent with the negative binomial frequency distribution testing results in section 2.1 as the *p* values are not very high but disallow us rejecting the null hypotheses at low significance level.

Transition Matrix

The implied steady-state probability of the transition matrix is tested against the simulation result. The results and calculation step are shown below. The simulation results show the similar steady-state probability.

[22] Package *stats* version 2.12.0, R Documentation, The Negative Binomial Distribution.

Line 1 Frequency	Line 2 Frequency
$\begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} = \begin{pmatrix} 0.15 & 0.85 \\ 0.1 & 0.9 \end{pmatrix}$	$\begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} = \begin{pmatrix} 0.2 & 0.8 \\ 0.1 & 0.9 \end{pmatrix}$
$(\Pi_1 \quad \Pi_2) = (10.53\% \quad 89.47\%)$	$(\Pi_1 \quad \Pi_2) = (11.11\% \quad 88.89\%)$
Non Zero Cases:	
State 1: 391	State 1: 410
State 2: 2797	State 2: 2733
Probability of Zero Cases:	
State 1: 0.005% (e^{-10})	State 1: 0.005% (e^{-10})
State 2: 0.125 ($prob^3$)	State 2: 0.135 (e^{-2})
Estimated all Cases: Non Zero Cases/ (1 – Probability of Zero Cases)	
State 1: 391	State 1: 410
State 2: 3188 (2797/(1-0.125))	State 2: 3161 (2733/(1-0.135))
Total Cases: # of simulations * 12 months = 3600	
Steady-state probability (compared with Π_1 & Π_2)	
State 1: 391/3600 = 10.86%	State 1: 410/3600 = 11.4%
State 2: 1-10.86% = 89.14%	State 2: 1-11.4% = 88.6%

4.2.2 Severity

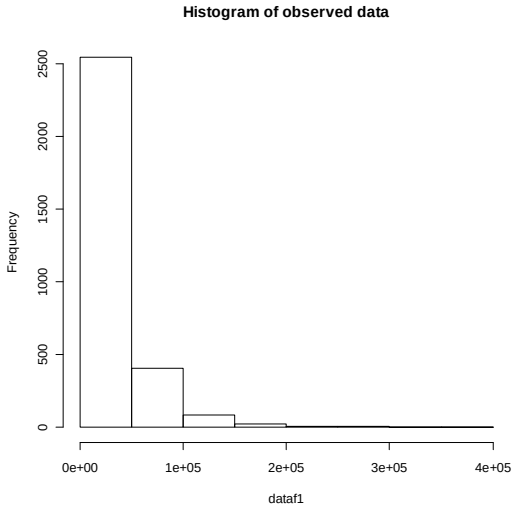
In testing Line 1 severity data, one thing worth noticing is that the size of loss assumption in the LSM is based on report date. Accident date might be a better choice to link size of loss with date of occurrence. For example, the size of loss might be more relevant to the time of catastrophic event like the 2011 Japanese earthquake instead of the time that the loss caused by the event is reported. From the modeling perspective, it also creates difficulties to realize the two-state, regime-switching function as the simulation is looped around each accident date instead of reporting date. In this testing, size of loss assumption is changed to be linked with accident date.

We split the claim data according to the state of the severity and test whether the distribution for each state follows our model assumption.

State 1

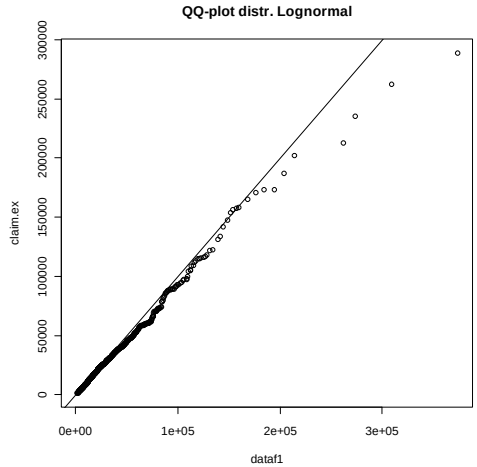
First, we draw a histogram of the simulated severity data to have an indication of the distribution type.

Figure 46. Histogram of simulated severity data (State 1)



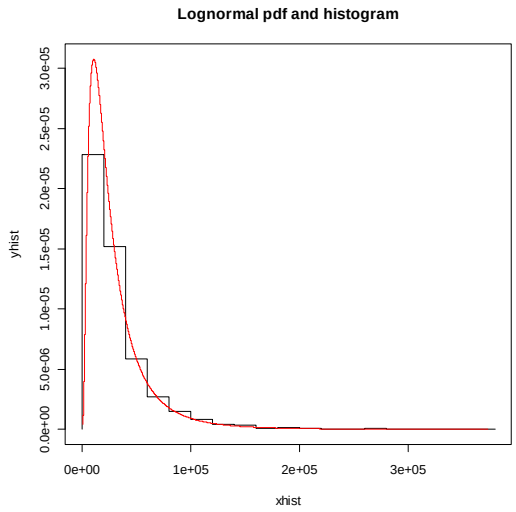
A QQ plot compares the simulated results with the intended distribution – Lognormal (mu = 10 and sigma = 0.832549779). From Figure 47, we can see that it is a good fit although the expected severity distribution as in the LSM has a slightly shorter tail than the simulated results.

Figure 47. QQ Plot – Simulated results vs. Lognormal (μ = 10 and σ = 0.832549779)



Comparing the probability distribution functions also gives us a vivid illustration of the fit.

Figure 48. PDF – simulated vs. assumption



Using ML method gives us an estimation of the parameters as follows:

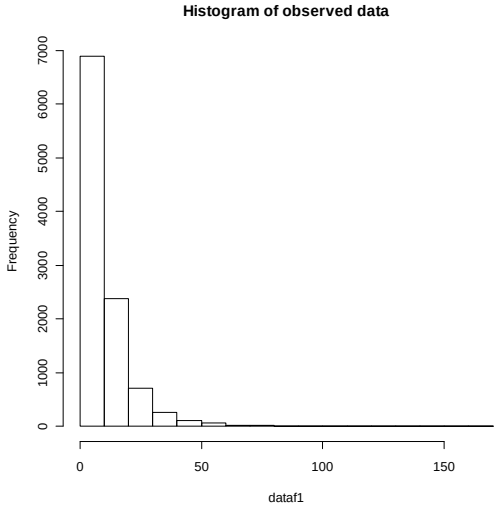
	meanlog	sdlog
Estimation	10.00677788	0.85323121
Standard deviation	0.01536917	0.01086764

Compare with our LSM input: mu = 10 and sigma = 0.832549779. We can see that at significance level of 5%, the confidence intervals for both parameters include the model input.

State 2

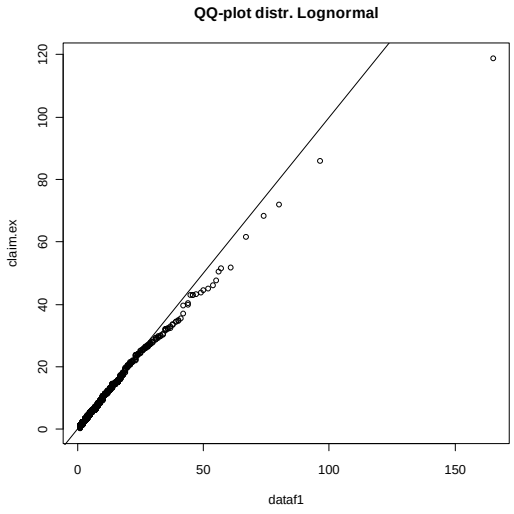
First, we draw a histogram of the simulated severity data to have an indication of the distribution type.

Figure 49. Histogram of simulated severity data (State .2)



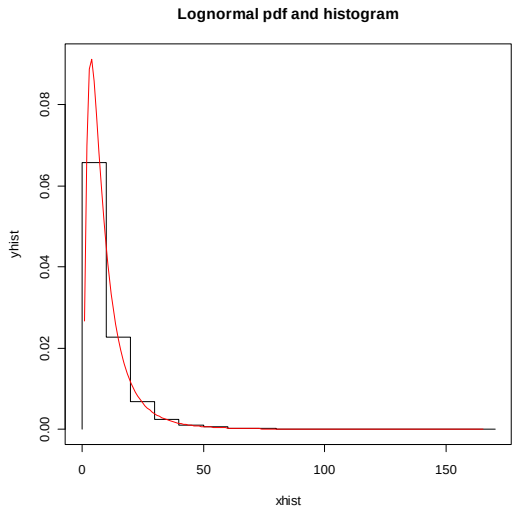
A QQ plot compares the simulated results with the intended distribution –Lognormal (mu = 2, sigma = 0.832549779). From Figure 50, we can see that it is a good fit. The expected severity distribution in the LSM also has a slightly shorter tail than the simulated results as in state 1.

Figure 50. QQ Plot – Simulated results vs. Lognormal ($\mu = 2$, $\sigma = 0.832549779$)



Comparing the probability distribution functions also shows the fit below.

Figure 51. PDF – simulated vs. assumption



Using the ML method gives us an estimation of the parameters as follows:

	meanlog	sdlog
Estimation	2.00714752	0.83957055
Standard deviation	0.00820275	0.00580022

In comparison to our LSM input of mu = 2 and sigma = 0.832549779, we can see at a significant level of 5% that the confidence intervals for both parameters include the model input.

4.2.3 Correlation

Correlation is tested to make sure that the correlation modeling using Copula is not affected by a two-state, regime-switching model. Correlation between frequencies of two lines is chosen for testing. We have four sets of data to test:

Set 1: Line 1: State 1 and Line 2: State 1

Set 2: Line 1: State 1 and Line 2: State 2

Set 3: Line 1: State 2 and Line 2: State 1

Set 4: Line 1: State 2 and Line 2: State 2

Scatter plots for the intended copula and simulated frequency pairs are shown below. Figures from 52 to 56 below show that they are of similar patterns.

Figure 52. Normal Copula (0.95)

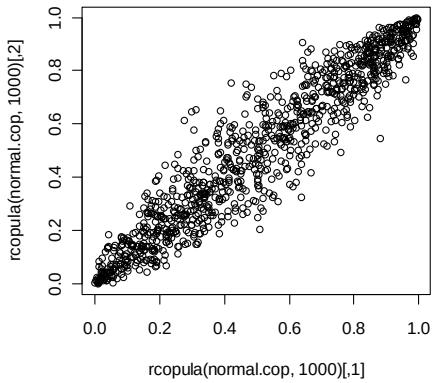


Figure 53. Set 1

Figure 54. Set 2

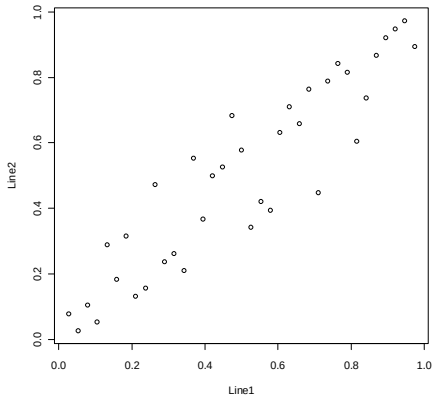


Figure 55. Set 3

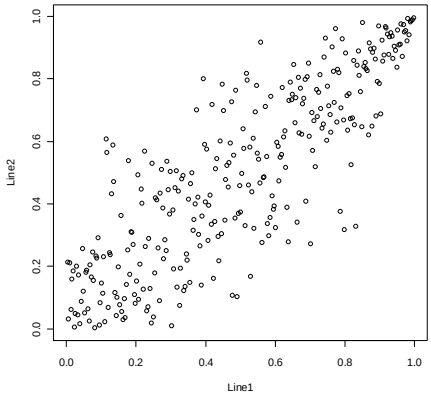
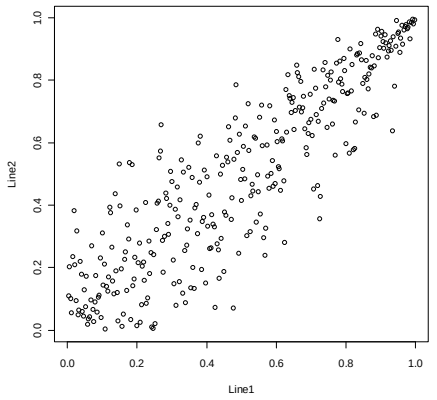
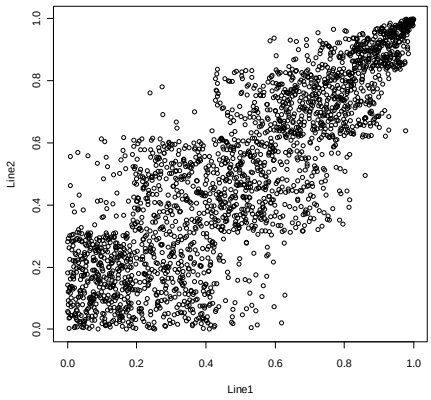


Figure 56. Set 4





For each set, we use maximum likelihood and inversion of Kendall's tau for parameter estimation

and goodness-of-fit test. Below are the results.

Set 1: State 1 for Line 1 and State 1 for Line 2

Normal copula parameter is estimated based on simulated frequency data using two methods.

(1) The estimation is based on the maximum likelihood and a sample of size 37.

	Estimate	Std. Error	z value	Pr(>|z|)
rho.1	0.9344341	0.01531399	61.01832	0

The maximized loglikelihood is 35.42264.

(2) The estimation is based on the inversion of Kendall's tau and a sample of size 1000.

	Estimate	Std. Error	z value	Pr(>|z|)
parameter	0.9380688	0.02458959	38.14903	0

We can see that the model parameter 0.95 is within the 95% confidence interval based on either of the two methods.

Goodness-of-fit Test

(1) Using Maximum Likelihood method for parameter estimation:

Parameter estimate(s): 0.9344341

Cramer-von Mises statistic: 0.01936648 with *p*-value 0.6980198

(2) Using Inversion of Kendall's tau method for parameter estimation:

Parameter estimate(s): 0.9380688

Cramer-von Mises statistic: 0.01821279 with *p*-value 0.7079208

Kolmogorov-Smirnov test is also done for testing the copula.

Two-sample Kolmogorov-Smirnov test

$D = 0.0423$, p-value = 0.9995

Based on those testing results, we can conclude that the simulated results show the same correlation as defined in model input.

Set 2: State 1 for Line 1 and State 2 for Line 2

Normal copula parameter is estimated based on simulated frequency data using two methods.

(1) The estimation is based on the maximum likelihood and a sample of size 307.

	Estimate	Std. Error	z value	Pr(>|z|)
rho.1	0.8400551	0.01290163	65.1123	0

The maximized loglikelihood is 183.7114.

(2) The estimation is based on the inversion of Kendall's tau and a sample of size 307.

	Estimate	Std. Error	z value	Pr(>|z|)
parameter	0.852917	0.01677851	50.83388	0

We can see that the model parameter 0.95 is out of the 95% confidence interval based on either of the two methods.

Goodness-of-fit Test

(1) Using Maximum Likelihood method for parameter estimation:

Parameter estimate(s): 0.8400551

Cramer-von Mises statistic: 0.03961167 with *p*-value 0.01485149

(2) Using Inversion of Kendall's tau method for parameter estimation

Parameter estimate(s): 0.852917

Cramer-von Mises statistic: 0.03370755 with *p*-value 0.01485149

Two-sample Kolmogorov-Smirnov test

$D = 0.0213$, p-value = 0.9837

The testing results show mixed information. One of the possible reasons for this is that we are not using all the simulated data for Set 2 but truncated data. If the number of claim is zero for a particular month, this data is not included in the claim output file from the LSM. Therefore, we are testing against non-zero monthly data only. As state 2 has a 12.5% and 13.5% probability of zero monthly claims for the two lines, respectively, we can see that except for Set 1, all other sets have the similar problem. It is still safe to conclude that high correlation exists as desired by model input.

Set 3: State 2 for Line 1 and State 1 for Line 2

Normal copula parameter is estimated based on simulated frequency data using two methods.

(1) The estimation is based on the maximum likelihood and a sample of size 329.

	Estimate	Std. Error	z value	Pr(>|z|)
rho.1	0.8644334	0.01056627	81.81065	0

The maximized loglikelihood is 222.0031.

(2) The estimation is based on the inversion of Kendall's tau and a sample of size 329.

	Estimate	Std. Error	z value	Pr(>|z|)
parameter	0.893593	0.01178312	75.8367	0

We can see that the model parameter 0.95 is out of the 95% confidence interval based on either of the two methods.

Goodness-of-fit Test

(1) Using Maximum Likelihood method for parameter estimation:

Parameter estimate(s): 0.8644334

Cramer-von Mises statistic: 0.07412085 with *p*-value 0.004950495

(2) Using Inversion of Kendall's tau method for parameter estimation:

Parameter estimate(s): 0.893593

Cramer-von Mises statistic: 0.04756158 with *p*-value 0.004950495

Two-sample Kolmogorov-Smirnov test

$D = 0.016$, p-value = 0.9996

Similar with Set 2, high correlation exists in the simulated data.

Set 4: State 2 for Line 1 and State 2 for Line 2

Normal copula parameter is estimated based on simulated frequency data using two methods.

(1) The estimation is based on the maximum likelihood and a sample of size 2376.

	Estimate	Std. Error	z value	Pr(>|z|)

rho.1	0.8114362	0.005444864	149.0278	0

The maximized loglikelihood is 1270.765.

(2) The estimation is based on the inversion of Kendall's tau and a sample of size 2376.

	Estimate	Std. Error	z value	Pr(>|z|)
parameter	0.845676	0.006024305	140.3773	0

We can see that the model parameter 0.95 is out of the 95% confidence interval based on either of the two methods.

Goodness-of-fit Test

(1) Using Maximum Likelihood method for parameter estimation:

Parameter estimate(s): 0.8114362

Cramer-von Mises statistic: 0.5949188 with *p*-value 0.004950495

(2) Using Inversion of Kendall's tau method for parameter estimation:

Parameter estimate(s): 0.845676

Cramer-von Mises statistic: 0.4380294 with *p*-value 0.004950495

Two-sample Kolmogorov-Smirnov test

$D = 0.0289$, *p*-value = 0.1900

Similar with Set 2, high correlation exists in the simulated data.

5. CONCLUSION AND FURTHER DEVELOPMENT

Based on the tests that have been conducted on the LSM, we cannot reject the assumption that model input and output are consistent regarding the following:

(1) Negative binomial frequency distribution.

(2) All copula types for frequencies among different lines except Gumbel Copula.

(3) the correlation modeling between report lag and loss size based on Normal Copula.

(4) Severity trend.

(5) Alpha in severity trend.

Though the statistical test results does not support Gumbel Copula applied to frequencies correlation very well, it is safe to not reject the null hypothesis as at a lower significance level such as 1%; it still passes the goodness-of-fit test.

A case reserve adequacy test shows that the assumption is not consistent with simulation data. This may be caused by the linear interpolation method used to derive 40% time point case reserve. It is suggested revising the way in which valuation date is determined in the LSM. In addition to the simulated valuation dates based on the waiting-period distribution assumption as in the LSM, some deterministic time points can be added as valuation dates. The deterministic valuation dates are interpolated between the report date and the payment date. In the LSM, 0%, 40%, 70%, and 90% time-points, case reserve, adequacy distribution can be input into the model. Therefore, 0%, 40%, 70% and 90% time points may be added as deterministic valuation dates.

Marine claim data are used to fit the distribution for frequency and severity. Trend, seasonality, and correlation analyses are also conducted to determine model parameters. These could be examples of how we use real data to determine appropriate LSM input which can be used for simulation and further testing of different reserve methods. If there are some data about paid loss history of the claims, the LSM can be better utilized to test different reserving methods. This could be an area for further research on the LSM.

Some enhancements have been made to the LSM. In the LSM, size of loss is linked to report date. The accident date might be a better choice for linking the size of loss with date of occurrence as the report lag would only have slight impact in loss size. From the modeling perspective, it also creates difficulties to realize the two-state, regime-switching function as the simulation is looped around each accident date instead of reporting date.

A categorical variable is included to enable setting parameters/distribution type for different states. Two-state, regime-switching flexibility is built in to enable moving from one state to the other state with a specified transition matrix. This, hopefully, can add the flexibility to mimic the underlying cycle we normally see in P&C business. Relevant testing is performed on the simulation data, which shows the consistency between model input and model output.

Acknowledgment

The author acknowledges Robert Bear, Dana F. Joseph, Joe Marker, Bryan Ware, and Kun Zhang for guidance, review, comments, and full support of this research. They helped indentify a lot of errors in the earlier versions and made suggestions on the revision. The author is also grateful for the opportunity provided by Casualty Actuarial Society.

APPENDIX A. R[23] CODE

Statistical software R is used for loss simulation testing purpose. Based on the claim and transaction files output from the loss simulation model, R is used to process the data, conduct the statistical test for copula and distribution, and draw graphics for viewing goodness of fit. The R codes are listed below for each test. The input/output directory shall be revised if the codes are to be reused. Lines start with "#" is the description of the codes below it.

A.1 Negative Binomial Frequency Distribution Testing

```
# Read raw data (Claim output file)
rawdata<-read.csv("F:/Research/copula/copula test/Negative Binomial Frequency 100
0.4/co.csv",skip=1,header=TRUE)

# Manipulate claim output file to retrieve annual frequency data for each simulation/line
fcn<-function(dataset){
  x<-floor((dataset[4]-20000000)/100)
  return(x)}

# apply fcn which returns the month of accident date
dataindex<-apply(rawdata,1,fcn)
rawdata2<-cbind(rawdata,dataindex)
rawdata3<-aggregate(rawdata2, list(rawdata2$Simulation.No), length)
rawdata4<-rawdata3[,1:2]
dataf1<-rawdata4$Simulation.No
write.csv(datar,"F:/Research/copula/copula test/Negative Binomial Frequency 100 0.4/freq.csv")

#draw histogram
hist(dataf1,main="Histogram of observed data")

#QQPlot
freq.ex<-(rnbinom(n=1000,size=100,prob=0.4))
qqplot(dataf1,freq.ex,main="QQ-plot distr. Negative Binomial")
abline(0,1) ### a 45-degree reference line is plotted

#Histogram and PDF
h<-hist(dataf1,breaks=10)
xhist<-c(min(h$breaks),h$breaks)
yhist<-c(0,h$density,0)
xfit<-seq(min(dataf1),max(dataf1),by=1)
yfit<-dnbinom(xfit,size=100,prob=0.4)
plot(xhist,yhist,type="s",ylim=c(0,max(yhist,yfit)), main="Negative Binomial pdf and histogram")
lines(xfit,yfit, col="red")

#Goodness of fit test
library(vcd)
gf<-goodfit(dataf1,type= "nbinom",par=list(size=100,prob=0.4))
```

[23] R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

```
summary(gf)
plot(gf)
gf<-goodfit(dataf1,type= "nbinom",method= "ML")
fitdistr(dataf1, "Negative Binomial")
```

A.2 Correlation Test

Correlation among the frequencies of different lines

1. Clayton Copula

```
# Read raw data (Claim output file)
rawdata<-read.csv("F:/Research/copula/copula test/clayton 5/co.csv",skip=1,header=TRUE)

# Manipulate claim output file to retrieve annual frequency data for each simulation/line
fcn<-function(dataset){
  x<-floor((dataset[4]-20000000)/100)
  return(x)}
# apply fcn which returns the month of accident date
dataindex<-apply(rawdata,1,fcn)
rawdata2<-cbind(rawdata,dataindex)
#1st month instead of one year occurrences
rawdata2m<-rawdata2[rawdata2$dataindex==1,]
rawdata3<-aggregate(rawdata2m, list(rawdata2m$Simulation.No,rawdata2m$Line), length)
rawdata4<-rawdata3[,1:3]
data1<-rawdata4[rawdata4$Group.2==1,]
data2<-rawdata4[rawdata4$Group.2==2,]
rawdata5<-merge(data1,data2,by="Group.1")
datar<-cbind(rawdata5$Simulation.No.x,rawdata5$Simulation.No.y)
colnames(datar)<-c("Line1","Line2")
write.csv(datar,"F:/Research/copula/copula test/clayton 5/x.csv")

#copula test
n<-length(datar[,1])
set.seed(123)
x<- sapply(as.data.frame(datar), rank, ties.method = "random") / (n + 1)
plot(x)

#Load R packages
library(MASS)
library(methods)
library(mvtnorm)
library(scatterplot3d)
library(mnormt)
library(sn)
library(pspline)
library(copula)

#Set up copula object for copula distribution and goodness-of-fit test later
clayton.cop <- claytonCopula(6, dim=2)

#Copula fit with prespecified type.
```

```
fit.clayton<-fitCopula(clayton.cop,x,method="ml")
fit.clayton
fit.clayton<-fitCopula(clayton.cop,x,method="itau")
fit.clayton

#Copula Goodness-of-fit test
gofCopula(clayton.cop, x, N=100, method = "mpl")
gofCopula(clayton.cop, x, N=100, method = "itau")
```

2. Frank Copula

```
# Read raw data (Claim output file)
rawdata<-read.csv("F:/Research/copula/copula test/frank 8/co.csv",skip=1,header=TRUE)
# Manipulate claim output file to retrieve annual frequency data for each simulation/line
fcn<-function(dataset){
  x<-floor((dataset[4]-20000000)/100)
  return(x)}
# apply fcn which returns the month of accident date
dataindex<-apply(rawdata,1,fcn)
rawdata2<-cbind(rawdata,dataindex)
#1st month instead of one year occurences
rawdata2m<-rawdata2[rawdata2$dataindex==1,]
#rawdata3<-aggregate(rawdata2, list(rawdata2$Simulation.No,rawdata2$Line), length)
#1st month instead of one year occurences
rawdata3<-aggregate(rawdata2m, list(rawdata2m$Simulation.No,rawdata2m$Line), length)
rawdata4<-rawdata3[,1:3]
data1<-rawdata4[rawdata4$Group.2==1,]
data2<-rawdata4[rawdata4$Group.2==2,]
rawdata5<-merge(data1,data2,by="Group.1")
datar<-cbind(rawdata5$Simulation.No.x,rawdata5$Simulation.No.y)
colnames(datar)<-c("Line1","Line2")
write.csv(datar,"F:/Research/copula/copula test/frank 8/x.csv")

#copula test
n<-length(datar[,1])
set.seed(123)
x<- sapply(as.data.frame(datar), rank, ties.method = "random") / (n + 1)
plot(x)

#Load R packages
library(MASS)
library(methods)
library(mvtnorm)
library(scatterplot3d)
library(mnormt)
library(sn)
library(pspline)
library(copula)

#Set up copula object for copula distribution and goodness-of-fit test later
frank.cop <- frankCopula(8, dim=2)
```

```
#Copula fit with prespecified type.
fit.frank<-fitCopula(frank.cop,x,method="ml")
fit.frank
fit.frank<-fitCopula(frank.cop,x,method="itau")
fit.frank

#Copula Goodness-of-fit test
gofCopula(frank.cop, x, N=100, method = "mpl")
gofCopula(frank.cop, x, N=100, method = "itau")
```

3. Gumbel Copula

```
# Read raw data (Claim output file)
rawdata<-read.csv("F:/Research/copula/copula test/Gumbel 6/co.csv",skip=1,header=TRUE)
# Manipulate claim output file to retrieve annual frequency data for each simulation/line
fcn<-function(dataset){
  x<-floor((dataset[4]-20000000)/100)
  return(x)}
# apply fcn which returns the month of accident date
dataindex<-apply(rawdata,1,fcn)
#rawdata2<-cbind(rawdata,dataindex)
#1st month instead of one year occurences
rawdata2m<-rawdata2[rawdata2$dataindex==1,]
#rawdata3<-aggregate(rawdata2, list(rawdata2$Simulation.No,rawdata2$Line), length)
#1st month instead of one year occurences
rawdata3<-aggregate(rawdata2m,          list(rawdata2m$Simulation.No,rawdata2m$Line),          length)rawdata4<-
rawdata3[,1:3]
data1<-rawdata4[rawdata4$Group.2==1,]
data2<-rawdata4[rawdata4$Group.2==2,]
rawdata5<-merge(data1,data2,by="Group.1")
datar<-cbind(rawdata5$Simulation.No.x,rawdata5$Simulation.No.y)
colnames(datar)<-c("Line1","Line2")
write.csv(datar,"F:/Research/copula/copula test/Gumbel 6/x.csv")

#copula test
n<-length(datar[,1])
set.seed(123)
x<- sapply(as.data.frame(datar), rank, ties.method = "random") / (n + 1)
plot(x)

#Load R packages
library(MASS)
library(methods)
library(mvtnorm)
library(scatterplot3d)
library(mnormt)
library(sn)
library(pspline)
library(copula)

#Set up copula object for copula distribution and goodness-of-fit test later
gumbel.cop <- gumbelCopula(3, dim=2)
```

```
#Copula fit with prespecified type.
fit.gumbel<-fitCopula(gumbel.cop,x,method="ml")
fit.gumbel
fit.gumbel<-fitCopula(gumbel.cop,x,method="itau")
fit.gumbel

#Copula Goodness-of-fit test
gofCopula(gumbel.cop, x, N=100, method = "mpl")
gofCopula(gumbel.cop, x, N=100, method = "itau")
```

4. T Copula

```
# Read raw data (Claim output file)
rawdata<-read.csv("F:/Research/copula/copula test/t50.8/co.csv",skip=1,header=TRUE)
# Manipulate claim output file to retrieve annual frequency data for each simulation/line
fcn<-function(dataset){
  x<-floor((dataset[4]-20000000)/100)
  return(x)}
# apply fcn which returns the month of accident date
dataindex<-apply(rawdata,1,fcn)
rawdata2<-cbind(rawdata,dataindex)
#1st month instead of one year occurences
rawdata2m<-rawdata2[rawdata2$dataindex==1,]
#1st month instead of one year occurences
rawdata3<-aggregate(rawdata2m, list(rawdata2m$Simulation.No,rawdata2m$Line), length)
rawdata4<-rawdata3[,1:3]
data1<-rawdata4[rawdata4$Group.2==1,]
data2<-rawdata4[rawdata4$Group.2==2,]
rawdata5<-merge(data1,data2,by="Group.1")
datar<-cbind(rawdata5$Simulation.No.x,rawdata5$Simulation.No.y)
colnames(datar)<-c("Line1","Line2")
write.csv(datar,"F:/Research/copula/copula test/t50.8/x.csv")

#copula test
n<-length(datar[,1])
set.seed(123)
x<- sapply(as.data.frame(datar), rank, ties.method = "random") / (n + 1)
plot(x)

#Load R packages
library(MASS)
library(methods)
library(mvtnorm)
library(scatterplot3d)
library(mnormt)
library(sn)
library(pspline)
library(copula)

#Set up copula object for copula distribution and goodness-of-fit test later
t.cop <- tCopula(c(0.8), dim=2, dispstr="un", df=5, df.fixed=TRUE)
```

```
#Copula fit with prespecified type.
fit.t<-fitCopula(t.cop,x,method="ml")
fit.t
fit.t<-fitCopula(t.cop,x,method="itau")
fit.t

#Copula Goodness-of-fit test
gofCopula(t.cop, x, N=100, method = "mpl")
gofCopula(t.cop, x, N=100, method = "itau")
```

Correlation between claim size and report lag

```
# Read raw data (Claim and transaction output file)
rawdatap<-read.csv("F:/Research/copula/copula test/copula2/co.csv",skip=1,header=TRUE)
rawdataa<-read.csv("F:/Research/copula/copula test/copula2/to.csv",skip=1,header=TRUE)

# Manipulate transaction output file to retrieve final payment amount
rawdataa2<-rawdataa[rawdataa$Transaction=="CLS",]
data1<-rawdatap[,c(1,2,3,5)]
data2<-rawdataa2[,c(1,2,3,4,7)]
datan<-merge(data1,data2,by=c("Simulation.No","Occurrence.No","Claim.No"))

# Translate payment date in terms of years
fcn<-function(dataset){
x<-floor(dataset[5]/10000)-floor(dataset[4]/10000)
y<-floor(dataset[5]/100)-floor(dataset[5]/10000)*100-(floor(dataset[4]/100)-floor(dataset[4]/10000)*100)
z<-dataset[5]-floor(dataset[5]/100)*100-(dataset[4]-floor(dataset[4]/100)*100)
r<-x+y/12+z/365
return(r)}
paymentlag<-apply(datan,1,fcn)
rawdatap2<-cbind(datan,paymentlag)
datar<-cbind(rawdatap2$paymentlag,rawdatap2$Payment)
write.csv(datar,"F:/Research/copula/copula test/copula2/100/x.csv")

#copula test
n<-length(datar[,1])
set.seed(123)
x<- sapply(as.data.frame(datar), rank, ties.method = "random") / (n + 1)
plot(x)

#Load R packages
library(MASS)
library(methods)
library(mvtnorm)
library(scatterplot3d)
library(mnormt)
library(sn)
library(pspline)
library(copula)

#Set up copula object for copula distribution and goodness-of-fit test later
normal.cop <- normalCopula(c(0),dim=2,dispstr="un")
```

```
#Copula fit with pre specified type.
fit.normal<-fitCopula(normal.cop,x,method="ml")
fit.normal
fit.normal<-fitCopula(normal.cop,x,method="itau")
fit.normal

#Copula Goodness-of-fit test
gofCopula(normal.cop, x, N=100, method = "mpl")
gofCopula(normal.cop, x, N=100, method = "itau")
```

A.3 Severity Trend

```
# Read raw data (Claim and transaction output file)
rawdatap<-read.csv("F:/Research/copula/copula test/strend/co.csv",skip=1,header=TRUE)
rawdataa<-read.csv("F:/Research/copula/copula test/strend/to.csv",skip=1,header=TRUE)

# Manipulate transaction output file to retrieve final payment amount
rawdataa2<-rawdataa[rawdataa$Transaction=="CLS",]
fcn<-function(dataset){
  x<-floor((dataset[4]-20000000)/100)
  return(x)}
# apply fcn which returns the month of accident date
dataindex<-apply(rawdatap,1,fcn)
rawdatap2<-cbind(rawdatap,dataindex)
data1<-rawdatap2[,c(1,2,8)]
data2<-rawdataa2[,c(1,2,7)]
datan<-merge(data1,data2,by=c("Simulation.No","Occurrence.No"))
rawdata3<-aggregate(datan, list(datan$dataindex), mean)
#rawdata4<-rawdata3[,c(3,5,6)]
rawdata4<-rawdata3[,c(4,5)]
colnames(rawdata4)<-c("Month","MeanPayment")
write.csv(rawdata4,"F:/Research/copula/copula test/strend/x.csv")
datar<-rawdata4$MeanPayment

#set up time series
ts1<-ts(datar,start=2000,frequency=12)
plot(ts1)
plot(stl(ts1,s.window="periodic"))

#linear trend fitting
trend = time(ts1)-2000
reg = lm(log(ts1)~trend, na.action=NULL)
summary(reg)
plot(log(ts1), type="o")
lines(fitted(reg), col=2)
par(mfrow=c(3,1))
plot(resid(reg))
acf(resid(reg),20)
pacf(resid(reg),20)
```

A.4 Alpha in Severity Trend

```
# Read raw data (Claim and transaction output file)
rawdatap<-read.csv("F:/Research/copula/copula test/Alpha/co.csv",skip=1,header=TRUE)
rawdataa<-read.csv("F:/Research/copula/copula test/Alpha/to.csv",skip=1,header=TRUE)

# Manipulate transaction output file to retrieve final payment amount
rawdataa2<-rawdataa[rawdataa$Transaction=="CLS",]
fcn<-function(dataset){
  x<-floor((dataset[4]-20000000)/100)
  return(x)}
# apply fcn which returns the month of accident date
dataindex<-apply(rawdatap,1,fcn)
rawdatap2<-cbind(rawdatap,dataindex)
dataindex2<-apply(rawdataa2[,c(1:4)],1,fcn)
rawdataa3<-cbind(rawdataa2,dataindex2)
data1<-rawdatap2[,c(1,2,8)]
data2<-rawdataa3[,c(1,2,7,8)]
datan<-merge(data1,data2,by=c("Simulation.No","Occurrence.No"))
datam<-datan[datan$dataindex==1,]
b<-datam[datam$dataindex2==7,]
c<-b[b$Payment!=0,]
a<-c$Payment
length(a)

#draw histogram
hist(a,main="Histogram of observed data")
library(MASS)
fitdistr(a, "Lognormal")

#QQPlot
Seve.ex<-(rlnorm(n=1000,meanlog=-0.8726,sdlog=0.9567))
qqplot(a,Seve.ex,main="QQ-plot distr. Lognormal")
abline(0,1) ### a 45-degree reference line is plotted

#Histogram and PDF
h<-hist(a,breaks=10)
xhist<-c(min(h$breaks),h$breaks)
yhist<-c(0,h$density,0)
xfit<-seq(min(a),max(a),by=1)
yfit<-dlnorm(xfit,meanlog=-0.8726,sdlog=0.9567)
plot(xhist,yhist,type="s",ylim=c(0,max(yhist,yfit)), main="Lognormal pdf and histogram")
lines(xfit,yfit, col="red")

#Kolmogorov-Smirnov Tests
ks.test(a,"plnorm", meanlog=-0.8726,sdlog=0.9567)

#Anderson-Darling Test
datas1.norm<-log(a)
library(nortest) ### package loading
ad.test(datas1.norm)
```

A.5 Case Reserve Adequacy

```
# Read raw data (Claim output file)
rawdatap<-read.csv("D:/LS/RS/case reserve/025/to.csv",skip=1,header=TRUE)

# Manipulate transaction output file to retrieve final payment amount
rawdataa<-rawdatap[rawdatap$Simulation.No<101,]

# Calculate the number of days that have passed since Jan 1,2000 until the accident date
x<-(floor(rawdataa[4]/10000)-2000)*365+(floor(rawdataa[4]/100)-floor(rawdataa[4]/10000)*100)*30+rawdataa[4]-
floor(rawdataa[4]/100)*100
rawdatap2<-cbind(rawdataa,x)
# Linear Interpolation of generated case reserves to get 40% time point case reserve
fcn<-function(dataset){
  aa<-dataset$Date
  b<-dataset$Case.Reserve
  bb<-dataset$Case.Reserve
  cc<-dataset$Payment
  count<-length(dataset$Date)
  temp<-0
  for(k in 1:(count-1)){
  bb[k]<-b[k]+temp
  temp<-temp+b[k]
  }
  bb[count]<-cc[count]
  f<-approxfun(aa,bb)
  xmin<-min(dataset[5])
  xmax<-max(dataset[5])
  x<-0.6*xmin+0.4*xmax
  if(cc[count]==0){
  return(0)
  }else{
  return(f(x)/cc[count]/0.6)}
  }
rawdata0<-rawdatap2[,c(1,2,6,7,8)]
m<-max(rawdata0$Simulation.No)
a<-matrix(rep(0,m*134),nrow=134,ncol=m)
# Get 40% case reserve for all claims
for(i in 1:m) {
 rawdata00<-rawdata0[rawdata0$Simulation.No==i,]
 rawdata<-as.data.frame(apply(rawdata00,2,abs))
 n<-max(rawdata$Occurrence.No)
 for(j in 1:n) {
 dataset<-as.data.frame(rawdata[rawdata$Occurrence.No==j,])
 a[j,i]=fcn(dataset)
 }
}
a<-as.vector(a)
a<-a[a!=0]

#draw histogram
hist(a,main="Histogram of observed data")
library(MASS)
```

```
fitdistr(a, "Lognormal")

#QQPlot
Seve.ex<-(rlnorm(n=1000,meanlog=0.25,sdlog=0.05))
qqplot(a,Seve.ex,main="QQ-plot distr. Lognormal")
abline(0,1) ### a 45-degree reference line is plotted

#Histogram and PDF
h<-hist(a,breaks=30)
xhist<-c(min(h$breaks),h$breaks)
yhist<-c(0,h$density,0)
xfit<-seq(min(a),max(a),by=1)
yfit<-dlnorm(xfit,meanlog=0.25,sdlog=0.05)
plot(xhist,yhist,type="s",ylim=c(0,max(yhist,yfit)), main="Lognormal pdf and histogram")
lines(xfit,yfit, col="red")

#Kolmogorov-Smirnov Tests
ks.test(a,"plnorm", meanlog=0.25,sdlog=0.05)

#Anderson-Darling Test
datas1.norm<-log(a)
library(nortest) ### package loading
ad.test(datas1.norm)
```

A.6 Real Claim Data Fitting

```
# Read raw data
rawdata<-read.csv("D:/LS/RS/PL/pl.csv",header=TRUE)
rawdata1<-rawdata[rawdata$Payment>0,]
dataProperty0<-rawdata1[rawdata1$Line=="Property",]
dataProperty<-dataProperty0[,-3]
datalia0<-rawdata1[rawdata1$Line=="Liability",]
datalia<-datalia0[,-3]

#Property
#draw histogram of claim
hist(log(dataProperty$Payment),breaks=100,main="Histogram of observed data")
library(MASS)
fitdistr(log(dataProperty$Payment), "normal")

#QQPlot of claim
claim.ex<-(rlnorm(n=1000,mean=9.285,sd=2.267))
qqplot(log(dataProperty$Payment),claim.ex,main="QQ-plot distr. Normal")
abline(0,1) ### a 45-degree reference line is plotted

rawdata3<-aggregate(dataProperty, list(dataProperty$dataindex), length)
rawdata4<-rawdata3[,1:2]
colnames(rawdata4)<-c("tMonth","Freq")
summary(rawdata4)

#set up time series for frequency
ts1<-ts(rawdata4$Freq,start=2006,frequency=12)
```

```
plot(ts1)
plot(stl(ts1,s.window="periodic"))

#trend analysis
trend = time(ts1)-2006
reg = lm(log(ts1)~trend, na.action=NULL)
summary(reg)
plot(log(ts1), type="o")
lines(fitted(reg), col=3, lwd=3)

par(mfrow=c(1,1))
plot(resid(reg))
acf(resid(reg),20)
pacf(resid(reg),20)

trendreg<--0.136*rawdata4[1]
detrend<-rawdata4[2]-trendreg

hist(as.numeric(detrend$Freq))
fitdistr(detrend$Freq,"normal")

#QQPlot of detrended frequency
freq.ex<-(rnorm(n=1000,mean=9.554,sd=3.131))
qqplot(detrend$Freq,freq.ex,main="QQ-plot distr. normal")
abline(0,1) ## a 45-degree reference line is plotted

ks.test(detrend$Freq,"pnorm", mean=9.554,sd=3.131)

#Histogram and PDF
h<-hist(detrend$Freq,breaks=15)
xhist<-c(min(h$breaks),h$breaks)
yhist<-c(0,h$density,0)
xfit<-seq(min(detrend$Freq),max(detrend$Freq),length=40)
yfit<-dnorm(xfit,mean=9.554,sd=3.131)
plot(xhist,yhist,type="s",ylim=c(0,max(yhist,yfit)), main="Normal pdf and histogram")
lines(xfit,yfit, col="red")

#Liability
#draw histogram of claim
hist(log(datalia$Payment),breaks=100,main="Histogram of observed data")

fitdistr(log(datalia$Payment), "normal")

#QQPlot of claim
claim.ex<-(rlnorm(n=1000,mean=9.5,sd=1.425))
qqplot(log(datalia$Payment),claim.ex,main="QQ-plot distr. Lognormal")
abline(0,1) ## a 45-degree reference line is plotted

rawdata3<-aggregate(datalia, list(datalia$dataindex), length)
rawdata4<-rawdata3[,1:2]
colnames(rawdata4)<-c("tMonth","Freq")
summary(rawdata4)
```

```
#set up time series
ts1<-ts(rawdata4$Freq,start=2006,frequency=12)
plot(ts1)
plot(stl(ts1,s.window="periodic"))

#trend analysis
trend = time(ts1)-2005
reg = lm(log(ts1)~trend, na.action=NULL)
summary(reg)
plot(log(ts1), type="o")
lines(fitted(reg), col=3,lwd=3)

par(mfrow=c(1,1))
plot(resid(reg))
acf(resid(reg),20)
pacf(resid(reg),20)

trendreg<--0.127*rawdata4[1]
detrend2<-rawdata4[2]-trendreg

#histogram of detrended data
hist(as.numeric(detrend$Freq))
fitdistr(detrend2$Freq,"lognormal")
fitdistr(detrend2$Freq,"normal")

#QQPlot of detrended frequency
freq.ex<-(rlnorm(n=100,meanlog=2.357,sdlog=0.3845))
qqplot(detrend2$Freq,freq.ex,main="QQ-plot distr. Lognormal")
abline(0,1) ### a 45-degree reference line is plotted

ks.test(detrend2$Freq,"plnorm", meanlog=2.357,sdlog=0.3845)

#Histogram and PDF
h<-hist(detrend2$Freq,breaks=15)
xhist<-c(min(h$breaks),h$breaks)
yhist<-c(0,h$density,0)
xfit<-seq(min(detrend2$Freq),max(detrend2$Freq),length=40)
yfit<-dlnorm(xfit,meanlog=2.357,sdlog=0.3845)
plot(xhist,yhist,type="s",ylim=c(0,max(yhist,yfit)), main="Normal pdf and histogram")
lines(xfit,yfit, col="red")

datar<-cbind(detrend$Freq,detrend2$Freq)
colnames(datar)<-c("Line1","Line2")

#copula test
n<-length(datar[,1])
set.seed(123)
x<- sapply(as.data.frame(datar), rank, ties.method = "random") / (n + 1)
plot(x)

cor(datar)
```

```
#Load R packages
library(MASS)
library(methods)
library(mvtnorm)
library(scatterplot3d)
library(mnormt)
library(sn)
library(pspline)
library(copula)

#Set up copula object for copula distribution and goodness-of-fit test later. Only Frank copula
#is shown here while in real testing different types of copula should all be tested against the data

frank.cop <- frankCopula(6, dim=2)

#Copula fit with pre specified type.
fit.frank<-fitCopula(frank.cop,x,method="ml")
fit.frank
fit.frank<-fitCopula(frank.cop,x,method="itau")
fit.frank

#Copula Goodness-of-fit test
gofCopula(frank.cop, x, N=100, method = "mpl")
gofCopula(frank.cop, x, N=100, method = "itau")
```

A.7 Two-State, Regime-Switching Feature Testing

Frequency

```
# Read raw data (Claim output file)
rawdata<-read.csv("D:/LS/RS/tsw/cc.csv",skip=1,header=TRUE)
# Manipulate claim output file to retrieve annual frequency data for each simulation/line
fcn<-function(dataset){
  x<-floor((dataset[4]-20000000)/100)
  return(x)}
# apply fcn which returns the month of accident date
dataindex<-apply(rawdata,1,fcn)
rawdata1<-cbind(rawdata,dataindex)
rawdata2<-rawdata1[rawdata1$Line==1,]

## State 1 Frequency Testing
rawdatas1<-rawdata2[rawdata2$State==1,]
rawdata3<-aggregate(rawdatas1, list(rawdatas1$Simulation.No,rawdatas1$dataindex), length)
dim(rawdata3)
rawdata4<-rawdata3[,1:3]
dataf1<-rawdata4$Simulation.No

#draw histogram
hist(dataf1,main="Histogram of observed data")

#QQPlot
```

```
freq.ex<-(rpois(n=1000,lambda=10))
qqplot(dataf1,freq.ex,main="QQ-plot distr. Poisson")
abline(0,1) ## a 45-degree reference line is plotted

#Histogram and PDF
h<-hist(dataf1,breaks=20)
xhist<-c(min(h$breaks),h$breaks)
yhist<-c(0,h$density,0)
xfit<-seq(min(dataf1),max(dataf1),by=1)
yfit<-dpois(xfit,lambda=10)
plot(xhist,yhist,type="s",ylim=c(0,max(yhist,yfit)), main="Poisson pdf and histogram")
lines(xfit,yfit, col="red")

#Goodness of fit test
library(vcd)
gf<-goodfit(dataf1,type= "pois",par=list(lambda=10),method= "MinChisq")
summary(gf)
plot(gf)
fitdistr(dataf1, "Poisson")
#Kolmogorov-Smirnov Tests
ks.test(dataf1,freq.ex,exact=NULL)

## State 2 Frequency Testing
rawdatas2<-rawdata2[rawdata2$State==2,]
rawdata3<-aggregate(rawdatas2, list(rawdatas2$Simulation.No,rawdatas2$dataindex), length)
dim(rawdata3)
rawdata4<-rawdata3[,1:3]
datafs1<-rawdata4$Simulation.No
dataf1<-c(rep(0,400),datafs1)

#draw histogram
hist(dataf1,main="Histogram of observed data")

#QQPlot
freq.ex<-(rnbinom(n=1000,size=3,prob=0.5))
qqplot(dataf1,freq.ex,main="QQ-plot distr. Negative Binomial")
abline(0,1) ## a 45-degree reference line is plotted

#Histogram and PDF
h<-hist(dataf1,breaks=10)
xhist<-c(min(h$breaks),h$breaks)
yhist<-c(0,h$density,0)
xfit<-seq(min(dataf1),max(dataf1),by=1)
yfit<-dnbinom(xfit,size=3, prob=0.5)
plot(xhist,yhist,type="s",ylim=c(0,max(yhist,yfit)), main="Negative Binomial pdf and histogram")
lines(xfit,yfit, col="red")

#Goodness of fit test
library(vcd)
gf<-goodfit(dataf1,type= "nbinom",par=list(size=3,prob=0.5),method= "MinChisq")
summary(gf)
plot(gf)
```

```
fitdistr(dataf1, "Negative Binomial")
```

Severity

```
# Read raw data (Claim output file)
rawdatap<-read.csv("D:/LS/RS/tsw/cc.csv",skip=1,header=TRUE)
rawdataa<-read.csv("D:/LS/RS/tsw/tt.csv",skip=1,header=TRUE)
# Manipulate transaction output file to retrieve final payment amount
rawdataa2<-rawdataa[rawdataa$Transaction=="CLS",]
fcn<-function(dataset){
  x<-floor((dataset[4]-20000000)/100)
  return(x)}
# apply fcn which returns the month of accident date
dataindex<-apply(rawdatap,1,fcn)
rawdatap2<-cbind(rawdatap,dataindex)

data1<-rawdatap2[,c(1,2,6,9)]
data2<-rawdataa2[,c(1,2,7,8)]
datan<-merge(data1,data2,by=c("Simulation.No","Occurrence.No"))
datal<-datan[datan$Line==1,]
datam<-aggregate(datal, list(datal $Simulation.No, datal $dataindex), mean)
dim(datam[datam$State==1,])
dim(datam[datam$State==2,])

datal1<-datan[datan$Line==1,]
datans1<-datal1[datal1$State==1,]
datans2<-datal1[datal1$State==2,]

## State 1 Severity Testing
dataf1<-datans1$Payment

#draw histogram
hist(dataf1,main="Histogram of observed data")

#QQPlot
claim.ex<-(rlnorm(n=1000,meanlog=10,sdlog=0.83255))
qqplot(dataf1,claim.ex,main="QQ-plot distr. Lognormal")
abline(0,1) ## a 45-degree reference line is plotted

#Histogram and PDF
h<-hist(dataf1,breaks=20)
xhist<-c(min(h$breaks),h$breaks)
yhist<-c(0,h$density,0)
xfit<-seq(min(dataf1),max(dataf1),by=1)
yfit<-dlnorm(xfit,meanlog=10,sdlog=0.83255)
plot(xhist,yhist,type="s",ylim=c(0,max(yhist,yfit)), main="Lognormal pdf and histogram")
lines(xfit,yfit, col="red")

# State 2 Severity Testing
dataf1<-datans2$Payment

#draw histogram
hist(dataf1,main="Histogram of observed data")
```

```
#QQPlot
claim.ex<-(rlnorm(n=1000,meanlog=2,sdlog=0.83255))
qqplot(dataf1,claim.ex,main="QQ-plot distr. Lognormal")
abline(0,1) ## a 45-degree reference line is plotted

#Histogram and PDF
h<-hist(dataf1,breaks=20)
xhist<-c(min(h$breaks),h$breaks)
yhist<-c(0,h$density,0)
xfit<-seq(min(dataf1),max(dataf1),by=1)
yfit<-dlnorm(xfit,meanlog=2,sdlog=0.83255)
plot(xhist,yhist,type="s",ylim=c(0,max(yhist,yfit)), main="Lognormal pdf and histogram")
lines(xfit,yfit, col="red")
```

Correlation

```
# Read raw data (Claim output file)
rawdata<-read.csv("D:/LS/RS/tsw/cc.csv",skip=1,header=TRUE)
# Manipulate claim output file to retrieve monthly frequency data for each simulation/line
fcn<-function(dataset){
  x<-floor((dataset[4]-20000000)/100)
  return(x)}
# apply fcn which returns the month of accident date
dataindex<-apply(rawdata,1,fcn)
rawdata2<-cbind(rawdata,dataindex)
rawdata3<-aggregate(rawdata2,    list(rawdata2$Simulation.No,rawdata2$Line,rawdata2$dataindex,rawdata2$State),
length)
rawdata4<-rawdata3[,1:5]
data1<-rawdata4[rawdata4$Group.2==1,]
data2<-rawdata4[rawdata4$Group.2==2,]
rawdata5<-merge(data1,data2,by=c("Group.1","Group.3"))

# Test for Line 1 State 1 and Line 2 State 1. This can be changed to other combinations of states 1&2, #2&1, and
2&2
rawdata6<-rawdata5[rawdata5$Group.4.x==1,]
rawdata7<-rawdata6[rawdata6$Group.4.y==1,]
datar<-cbind(rawdata7$Simulation.No.x,rawdata7$Simulation.No.y)
colnames(datar)<-c("Line1","Line2")

#copula test
n<-length(datar[,1])
set.seed(123)
x<- sapply(as.data.frame(datar), rank, ties.method = "random") / (n + 1)
plot(x)

#Load R packages
library(MASS)
library(methods)
library(mvtnorm)
library(scatterplot3d)
library(mnormt)
library(sn)
```

```
library(pspline)
library(copula)

#Set up copula object for copula distribution and goodness-of-fit test later
normal.cop <- normalCopula(c(0),dim=2,dispstr="un")

#Copula fit with prespecified type.
fit.normal<-fitCopula(normal.cop,x,method="ml")
fit.normal
fit.normal<-fitCopula(normal.cop,x,method="itau")
fit.normal

#Copula Goodness-of-fit test
gofCopula(normal.cop, x, N=100, method = "mpl")
gofCopula(normal.cop, x, N=100, method = "itau")

#K-S test.
normal.fit<-normalCopula(0.95, dim=2)
y<-rcopula(normal.fit,1000)
ks.test(x,y)
```

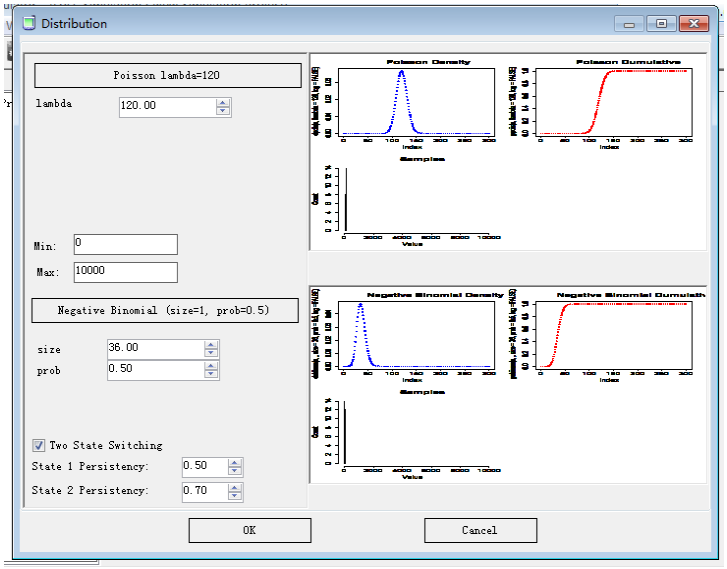
APPENDIX B. QUICK GUIDE FOR TWO-STATE REGIME-SWITCHING

The two-state, regime-switching feature is allowed for all variables that were modeled as distribution in LSM. Below is a short description about the related model input and output.

Model Input

Figure 57 below shows the model input interface of the example in section 4.1. By checking the checkbox "Two-state Switching," two distribution set up panels will be shown. You would also need to input the State 1 persistency and State 2 persistency. If only one distribution is desired, you could either uncheck the checkbox "Two-State Switching" or input the same distribution type and parameters for the 1st and 2nd distributions. By default, frequency and severity has two-state regime switching. For others like report lag, only one distribution is allowed. Those, however, can be changed. XML import/export setting is also revised for this enhancement.

Figure 57. Model input of two-state switching feature



Model Output

Claim and transaction output files: A new column "State" is added to record the state of frequency in claim output file and state of severity in transaction output file.

Claim output example snapshot

Simulation 2011/4/10 0:30:14

Simulation	Occurrence No	Claim No	Accident Date	Report Date	Line	Type	State
1	1	1	20000126	20000701	1	1	2
1	2	1	20000101	20000318	1	1	2
1	3	1	20000106	20010105	1	1	2
1	4	1	20000123	20000823	1	1	2
1	5	1	20000116	20000129	1	1	2
1	6	1	20000223	20000514	1	1	1
1	7	1	20000213	20000327	1	1	1
1	8	1	20000218	20000530	1	1	1
1	9	1	20000223	20010209	1	1	1
1	10	1	20000222	20000823	1	1	1
1	11	1	20000210	20000309	1	1	1
1	12	1	20000326	20000413	1	1	2
1	13	1	20000307	20000614	1	1	2
1	14	1	20000412	20000528	1	1	2
1	15	1	20000422	20000816	1	1	2
1	16	1	20000402	20000626	1	1	2

Transaction output example snapshot

Transactions 2011/4/10 0:30:14

Simulation	Occurrence No	Claim No	Date	Transaction	Case Reserve	Payment	State
1	1	1	20000701	REP	2000	0	2
1	1	1	20000703	RES	-1999	0	2
1	1	1	20000726	RES	0	0	2
1	1	1	20000804	RES	0	0	2
1	1	1	20000907	RES	0	0	2
1	1	1	20001023	RES	0	0	2
1	1	1	20001108	RES	0	0	2
1	1	1	20010115	RES	0	0	2
1	1	1	20010120	RES	0	0	2
1	1	1	20010501	CLS	-1	2	2
1	2	1	20000318	REP	2000	0	2
1	2	1	20000706	RES	-1998	0	2
1	2	1	20000729	CLS	-2	3	2
1	3	1	20010105	REP	2000	0	2
1	3	1	20010115	RES	-1994	0	2
1	3	1	20010202	RES	0	0	2
1	3	1	20010424	RES	0	0	2
1	3	1	20010515	RES	0	0	2
1	3	1	20010812	CLS	-6	0	2
1	4	1	20000823	REP	2000	0	2
1	4	1	20000907	RES	-1998	0	2
1	4	1	20001129	CLS	-2	0	2
1	5	1	20000129	REP	2000	0	2
1	5	1	20000228	RES	-1992	0	2
1	5	1	20000301	CLS	-8	13	2
1	6	1	20000514	REP	2000	0	1
1	6	1	20000524	RES	-1990	0	1
1	6	1	20000907	RES	0	0	1
1	6	1	20000914	RES	0	0	1

In tab "Summary" of simulation results, the state of each month's frequency and severity is output for checking and records.

Figure 58. Frequency state output

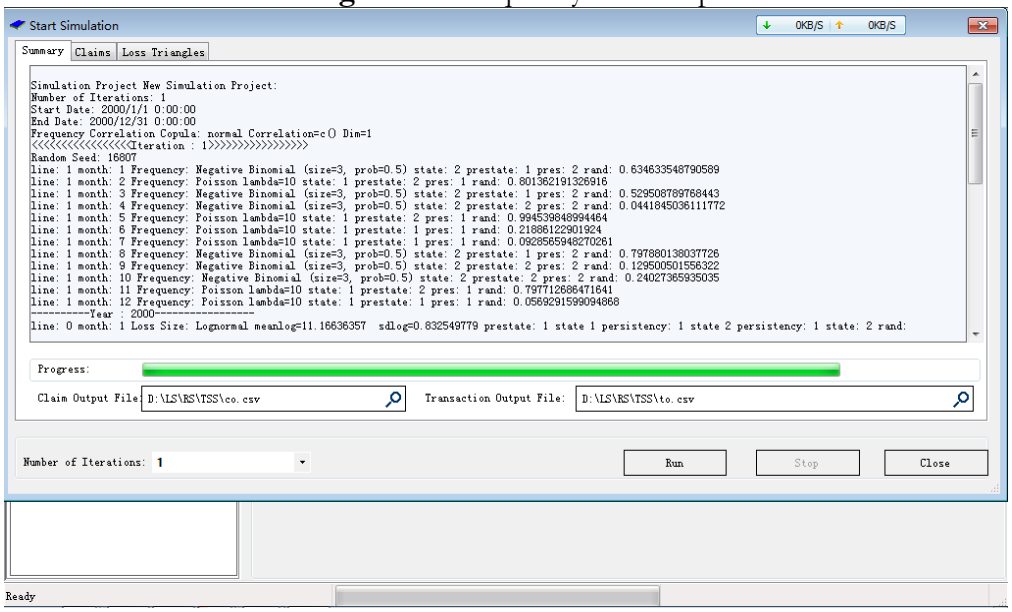
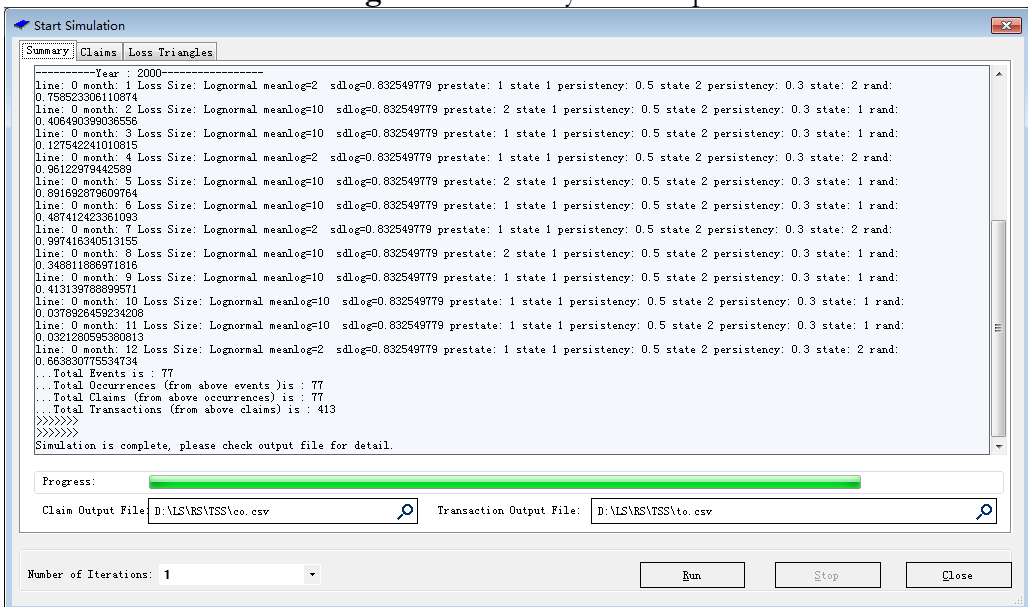


Figure 59. Severity state output



5. REFERENCES

[1] Cleveland, Robert B., William S. Cleveland, Jean E. McRae, and Irma Terpenning, "STL: A Seasonal-Trend Decomposition Procedure Based on Loess," *Journal of Official Statistics* 6, no. 2 (1990): 3-73.

[2] Genest, Christian, Bruno Rémillard, and David Beaudoin, "Goodness-of-fit tests for copulas: A review and a power study," *Insurance: Mathematics and Economics* 44, no. 2 (1999): 199-213.

[3] Kojadinovic, Ivan, and Yan Jun, "Modeling Multivariate Distributions with Continuous Margins Using the Copula R Package," *Journal of Statistical Software* 34, no. 9 (2010): 1-20.

[4] Li, David, "On Default Correlation A Copula Function Approach," *The Journal of Fixed Income* 9, no. 4 (March 2000): 43-54.

[5] LSMWP, "Modeling Loss Emergence and Settlement Processes-CAS Loss Simulation Model Working Party Summary Report," *Casualty Actuarial Society Forum*, 2010, 4-43.

[6] Nelsen, R.B., *An Introduction to Copulas,* 2nd ed. (New York: Springer, 2006), 109-155.

Abbreviations and notations

Collect here in alphabetical order all abbreviations and notations used in the paper

df, degree of freedom
LSM, Loss Simulation Model
LSMWP, Loss Simulation Model Working Party
ML, Maximum Likelihood
MLE, Maximum Likelihood Estimation
OLS, Ordinary Least Square
QQ, Quantile-Quantile plot
RN, Random Number

Biography of the Author

Kailan Shang is a pricing actuary at Manulife Financial in Canada. Before that, he worked in the area of financial risk management in AIA. Years of actuarial and risk management experience has allowed him to get a broad exposure in the fields of economic capital, market-consistent embedded value, financial engineering, dynamic management options and policyholder behavior modeling, product development and management, financial reporting, dynamic solvency testing, and the like.

As an FSA, CFA, PRM, and SCJP, he is also an enthusiast of actuarial research through both volunteer works and funded research program. He participated in the LSMWP and IAA Comprehensive Actuarial Risk Evaluation project and he is now working on the SOA research projects, "Valuation of Embedded Option in Pension Plan" and "Linkage Between Risk Appetite and Strategic Planning."

He can be reached at kailan_shang@manulife.com.