Forecasting 12-Month Operational Losses for Risk Management

Group 3 Project

ABOUT THE PROJECT

This project aims to forecast operational losses for the next 12 months using the Loss Distribution Approach (LDA). The LDA involves modeling the frequency and severity of loss events separately and then combining them to estimate the overall loss distribution.

we have utilize the Internal Loss Database (ILD) to analyze historical loss data. Key considerations include addressing data limitations, such as the absence of losses below a certain threshold, and accounting for potential overdispersion and correlation in the data. By carefully selecting appropriate frequency and severity distributions, estimating their parameters, and simulating future loss scenarios, the project aims to provide a reliable forecast of operational losses and inform risk management decisions.

Team

- Jeff Mathew Sam
- Melda Akan
- Oluwatobi Ige
- Rakesh Jayswal
- Salmah Animashaun

Objective

Objective Forecast 12-month operational losses for a firm's operational risk department, utilizing the internal loss database (ILD) with data on 213 unique loss events over 200 months.

Problem Statement The operational risk department of a firm needs to quantify operational risk and forecast 12-month operational losses. Loss events are recorded only when the loss amount exceeds a threshold of $1 million, leading to gaps in monthly data. Challenges include addressing overdispersion and handling the absence of data for losses below the threshold.

Possible Outcome A 12-month operational loss forecast model that provides monthly loss predictions with quantified uncertainties levels. This model will provide the risk department with a clearer understanding of future operational risk exposure. Key findings will recognize that the firm needs to hold a significant amount of regulatory capital to meet Basel II regulation, potentially increasing risk-weighted asset (RWA) and reducing the amount of capital available for revenue-generating activities.

Operational Risk

Basel Committee on Banking Supervision (BSBC): "the risk of loss resulting from inadequate for failed internal processes, people and systems or from external events. This definition includes legal risk, but excludes strategic and reputational risk."

Basel III (SMA)

- Consistent
- Can lead to higher capital requirement

Basel II Loss Event Type, Level 1 (Annex 9)

- Internal fraud
- External fraud
- Employment practices
- Clients, products and business practices
- Damages to physical assets
- Business disruption and system failures
- Execution, delivery and process management

Methodology

Basel II Advanced Management Approach (AMA)

- Basel II is a set of international banking regulations released by the Basel Committee on Banking Supervision (BSBC) in 2004 that establish the minimum amount of capital banks must hold to cover risks and maintain financial stability.
- Under Basel II, the Advanced Measurement Approach (AMA) was permitted for calculating operational risk capital, but this approach has since been replaced by the Standardized Approach (SA) in Basel III, which was developed in response to the financial crisis of 2007-2009.

Loss Distribution Approach (LDA): This is a separate modeling of the frequency and severity of losses.

Simulation Technique: applying Monte Carlo simulation to combine the frequency and severity distributions into a single compound loss distribution.

Frequency Modeling: using Poisson or Negative Binomial distribution *(Hull 2015, Kuketayev Lecture 2024)*

Severity Modeling: fit a log-normal, exponential or Weibull distribution for loss amount data.*(Hull 2015, Kuketayev Lecture 2024)*

Assumptions & choice of loss severity model

- Addressing overdispersion: **Lognormal Distribution as a Baseline:** The lognormal distribution is a common starting point for modeling operational loss severity due to its ability to handle moderately heavy-tailed data, characteristic of operational risk. *(Peters, Dutta, K. C., & Perry, J. (2012)*
- **Alternative Distributions for Extreme Events:** More flexible models like α-stable distributions, g-and-h distributions, and generalized beta (GB2) distributions are suggested to better capture the heavy tails and extreme losses. *(Peters, G. W., & Sisson, S. A. (2009)*
- **Model Selection is Crucial:** The "best" distribution for a specific case depends the and risk on the profile. Techniques like Bayesian model selection can be used to identify the most appropriate distribution for a particular institution and its operational risk. *(Frachot, A., Georges, P., & Roncalli, T. (2001)*
- **Poisson** as an industry standard **for Loss Frequency** choice even though it may not fit: 'For Loss Frequency, the exact distribution of frequencies is not important. Poisson distribution is the chosen and preferred distribution even though it may not fit.' ()

These assumptions highlight the need for careful consideration when choosing a loss distribution for operational risk modeling. While the lognormal distribution offers a starting point, it's important to acknowledge its limitations and explore alternative models that better capture the full spectrum of potential losses, especially for extreme events.

Models and supporting research:

- **Loss Severity:** Log-normal, Gamma (Frachot, A., Georges, P., & Roncalli, T. (2001), (Dutta K. and J. Perry (2006), (El Adlouni et.al. 2011); Exponential (Kuketayev Lecture 2024)
- **Loss Frequency**: Poisson. 'For loss distribution the exact distribution of frequencies is not so important.' (Kuketayev Lecture 2024)

Assumptions

- Dutta and Perry found that while the lognormal distribution often fits operational loss severity data well, especially in capturing heavy tails, its application within the Loss Distributional Approach (LDA) sometimes produces unrealistic capital estimates for extreme losses, highlighting the variability of model performance across institutions and business lines.

- Peters, Shevchenko, Young, and Yip discuss that while the lognormal distribution is frequently used in Loss Distributional Approach (LDA) models as a traditional severity model, it may not effectively capture the extreme tails of operational risk losses. They suggest that more flexible models, like α-stable distributions, could better represent the heavy tails and rare, severe losses characteristic of operational risk
-
- In "Bayesian Inference, Monte Carlo Sampling, and Operational Risk", Peters and Sisson discuss that while the log-normal distribution is commonly used in the Loss Distributional Approach (LDA) for its heavy-tailed nature, it often fails to accurately model extreme operational loss events; they suggest more flexible alternatives like the g-and-h and generalized beta (GB2) distributions, which can better capture the skewness and heavy tails typical of operational risk data

- In "Analytic Loss Distributional Approach Models for Operational Risk from the α-Stable Doubly Stochastic Compound Processes and Implications for Capital Allocation", the authors note that while the log-normal distribution is commonly used in the Loss Distributional Approach (LDA) for modeling moderate heavy tails, it may not effectively capture extreme heavy tails. They advocate for α-stable distributions as a more flexible alternative to accommodate rare, severe losses in operational risk data

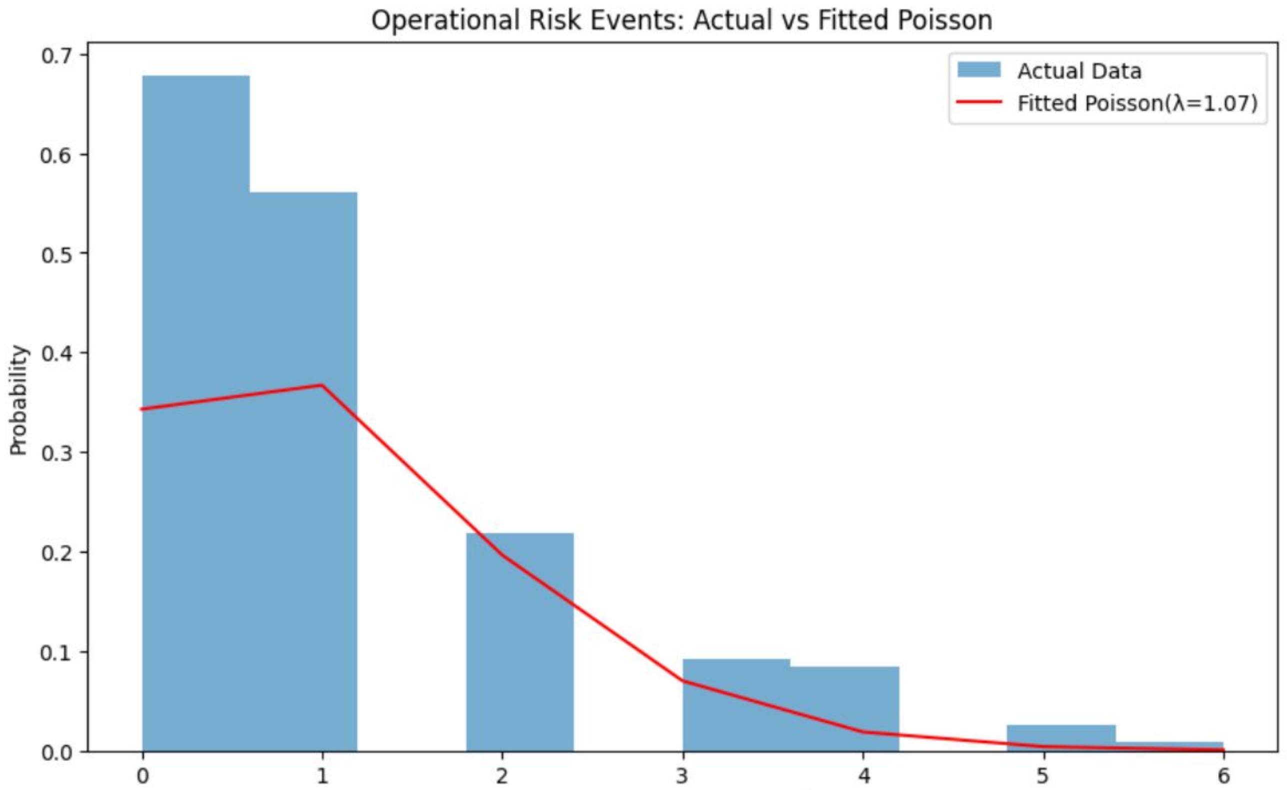
- In "Loss Distribution Approach for Operational Risk", Frachot, Georges, and Roncalli discuss the log-normal distribution as a viable model for loss severity within the Loss Distribution Approach (LDA) due to its ability to handle moderately heavy tails typical in operational risk. However, they note its limitations in capturing extreme tail risk, suggesting that alternative heavy-tailed distributions may be necessary for modeling rare, severe loss events, depending on an institution's specific risk profile.

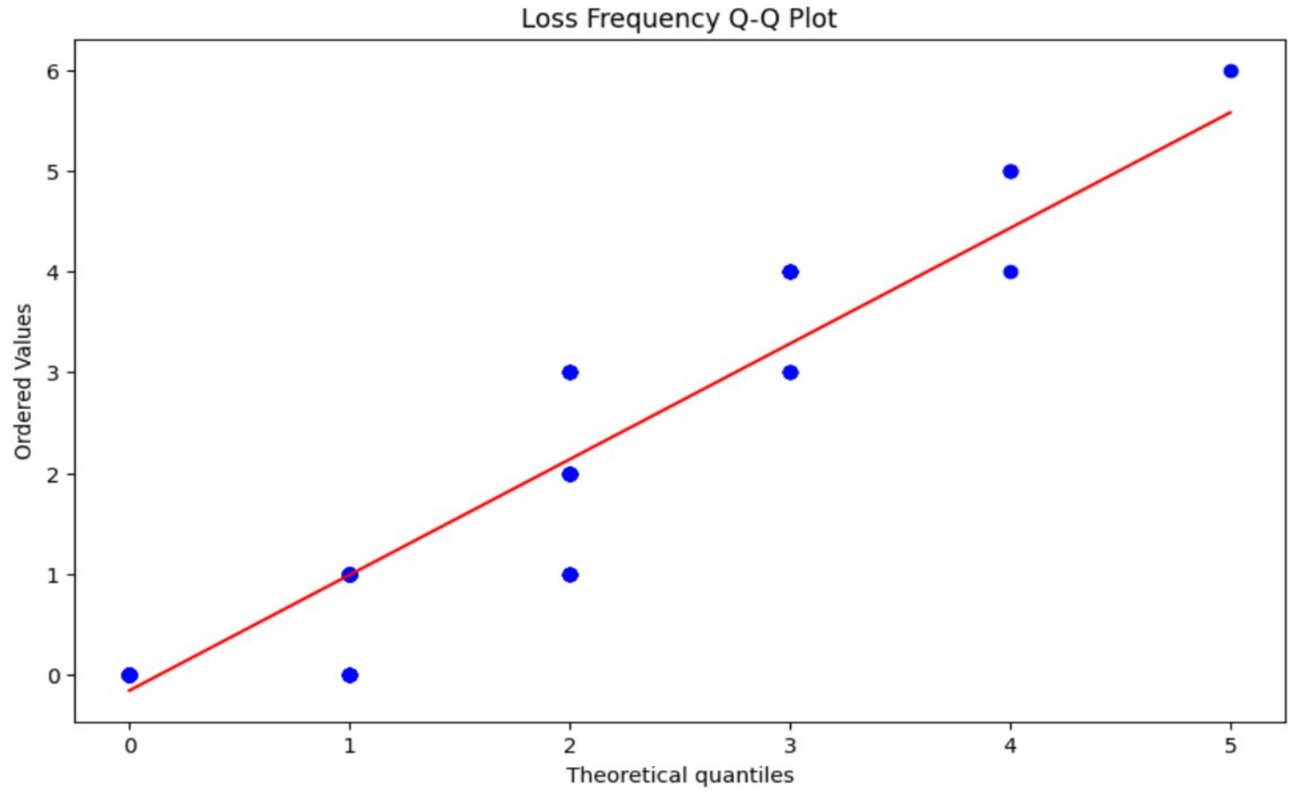
- In "On the Selection of Loss Severity Distributions to Model Operational Risk", the authors discuss the log-normal distribution as a commonly chosen model for loss severity in the Loss Distributional Approach (LDA), noting that while its right-skewed nature effectively captures moderate loss severity, it may be inadequate for modeling the extreme heavy tails often observed in operational risk data.

- Hugen indicates that even though the lognormal distribution may not perfectly fit the data, it is still frequently used for modeling loss severity in insurance contexts. Actuaries in the Property and Casualty fields have long used the lognormal distribution for calculating the risk of the probability of extreme, high-loss events

Model Description and Summary Statistics

Loss Frequency:





Summary Statistics with Skewness and Kurtosis:

	Amount	log_Amount
count	213.000000	213.000000
mean	4.879663	1.159349
std	6.309824	0.839217
min	1.000678	0.000678
25%	1.611940	0.477439
50%	2.593938	0.953177
75%	5.932096	1.780378
max	48.306953	3.877576
skewness	3.951344	0.822397
kurtosis	19.729102	0.139431

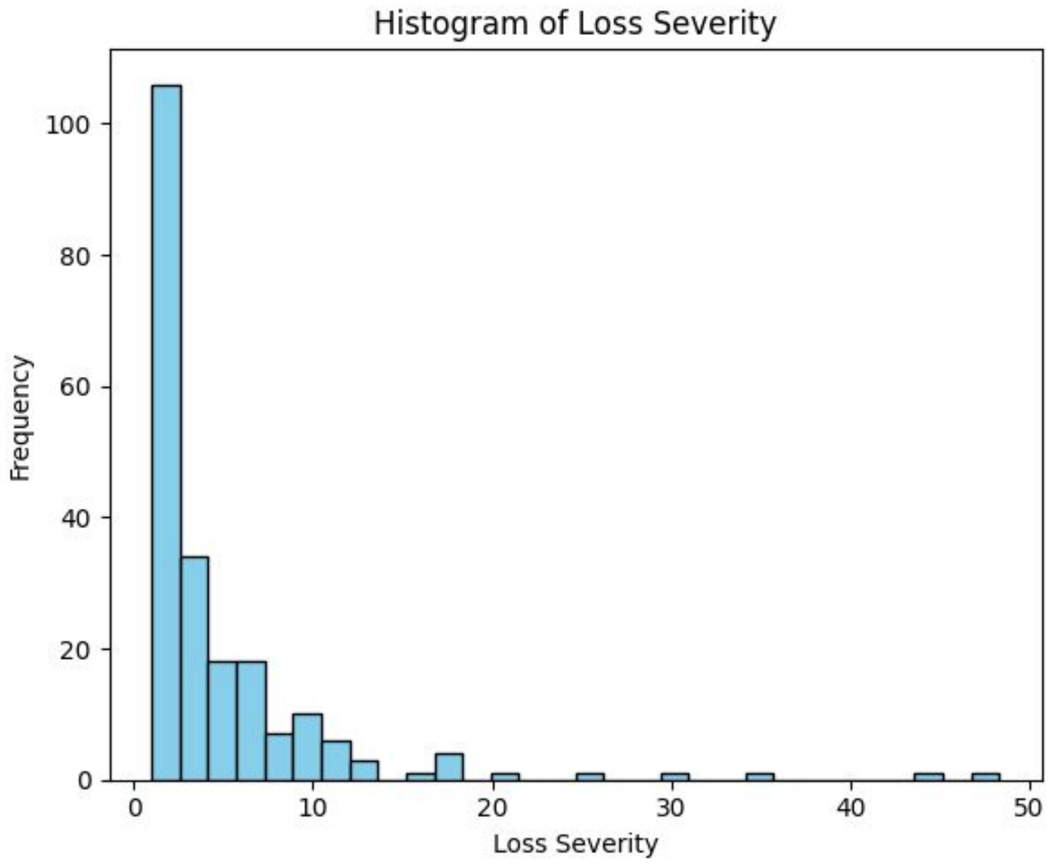
Goodness of fit tests:

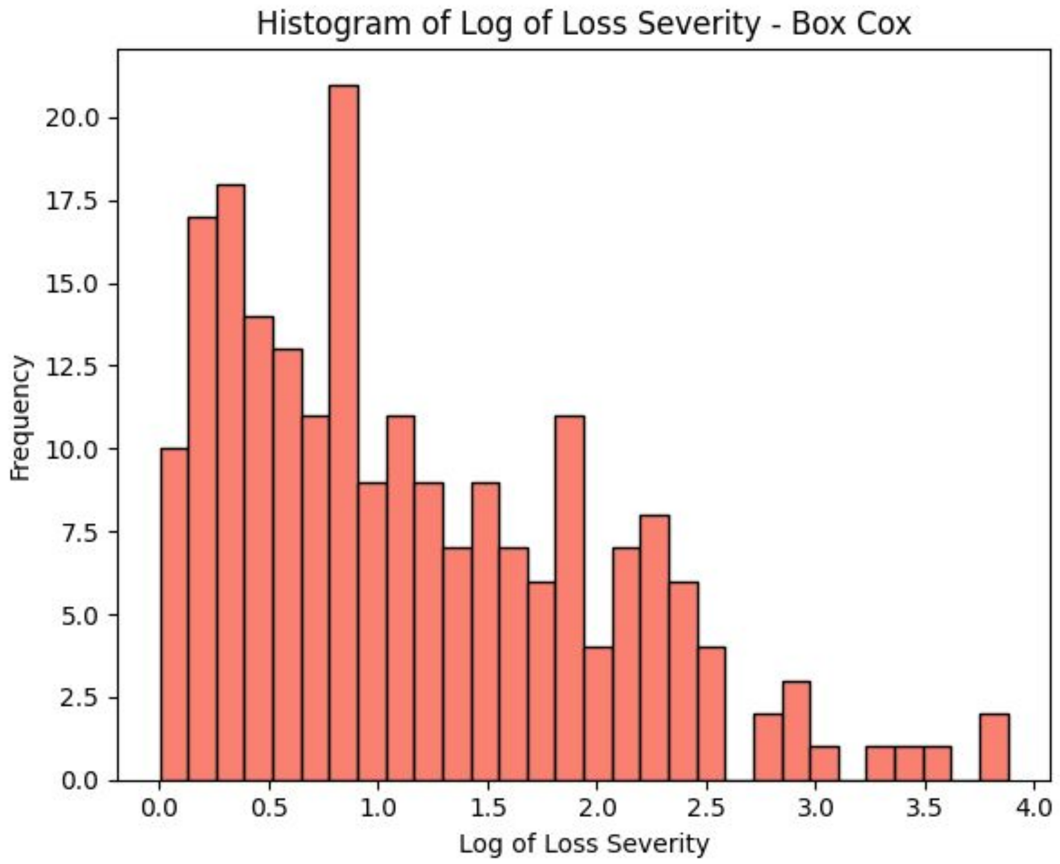
KS test p-value: 0.0000

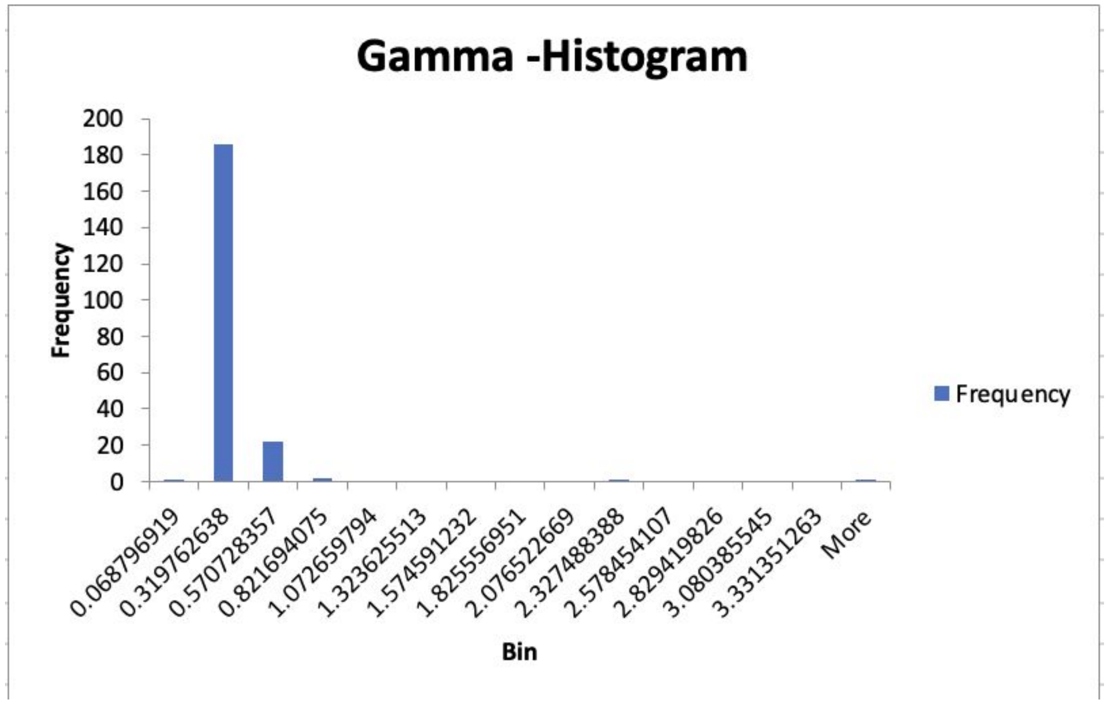
{'ks': (0.3428878823747608, 2.152466627831393e-21)}

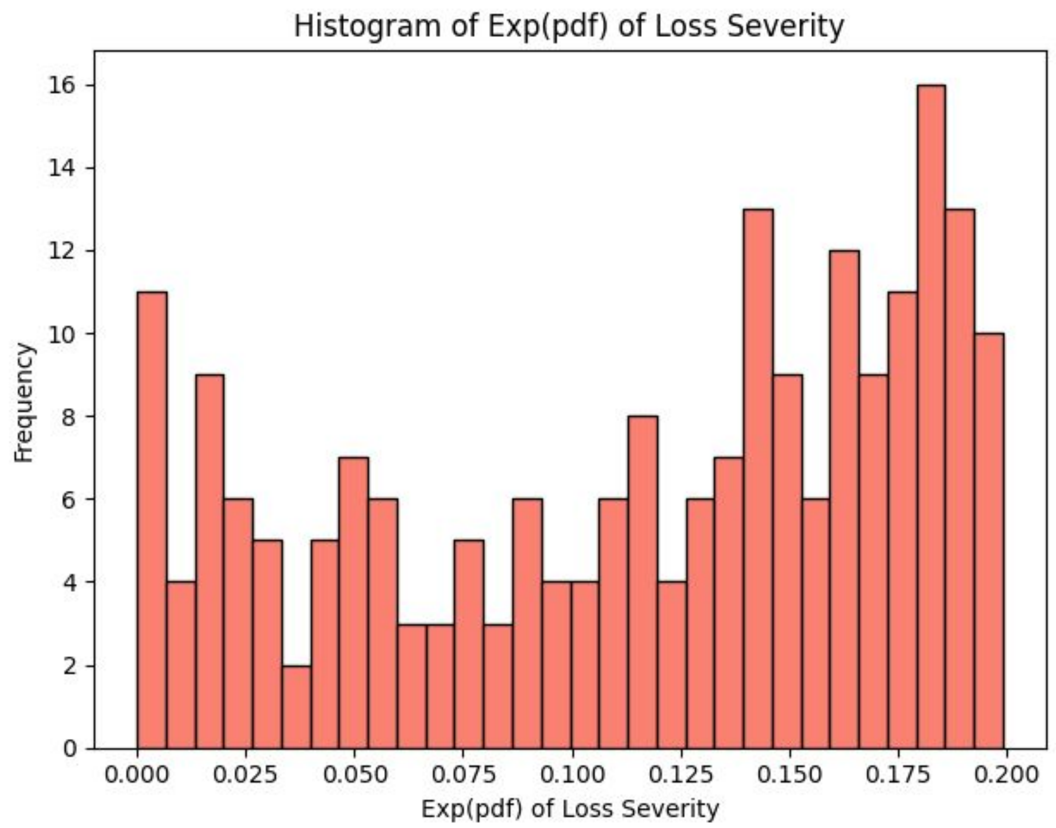
- As evidenced by both the statistical and visual diagnostics, the goodness-of-fit analysis reveals significant deviations from the theoretical distribution.
- Deviations from the straight line may suggest overdispersion (variance > mean).
- An exact frequency distribution is less consequential for loss distribution, even in practice.

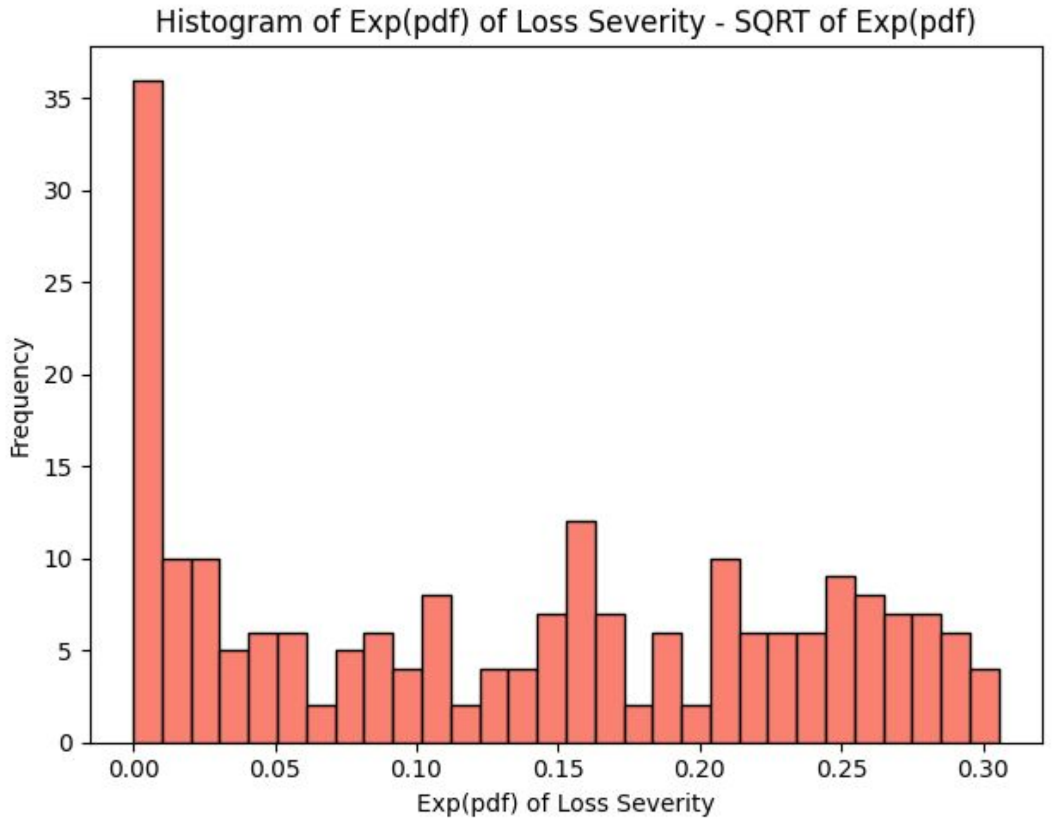
Loss Severity Visualisation











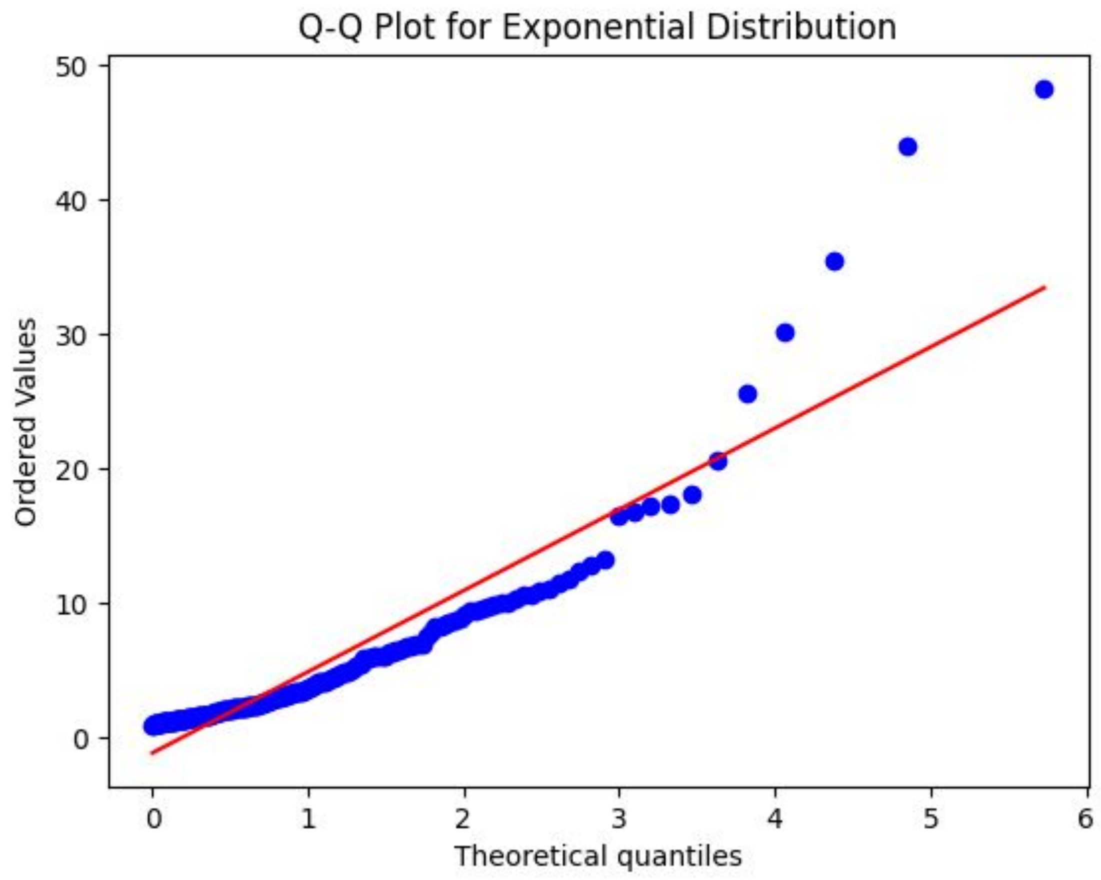
Exponential Distribution

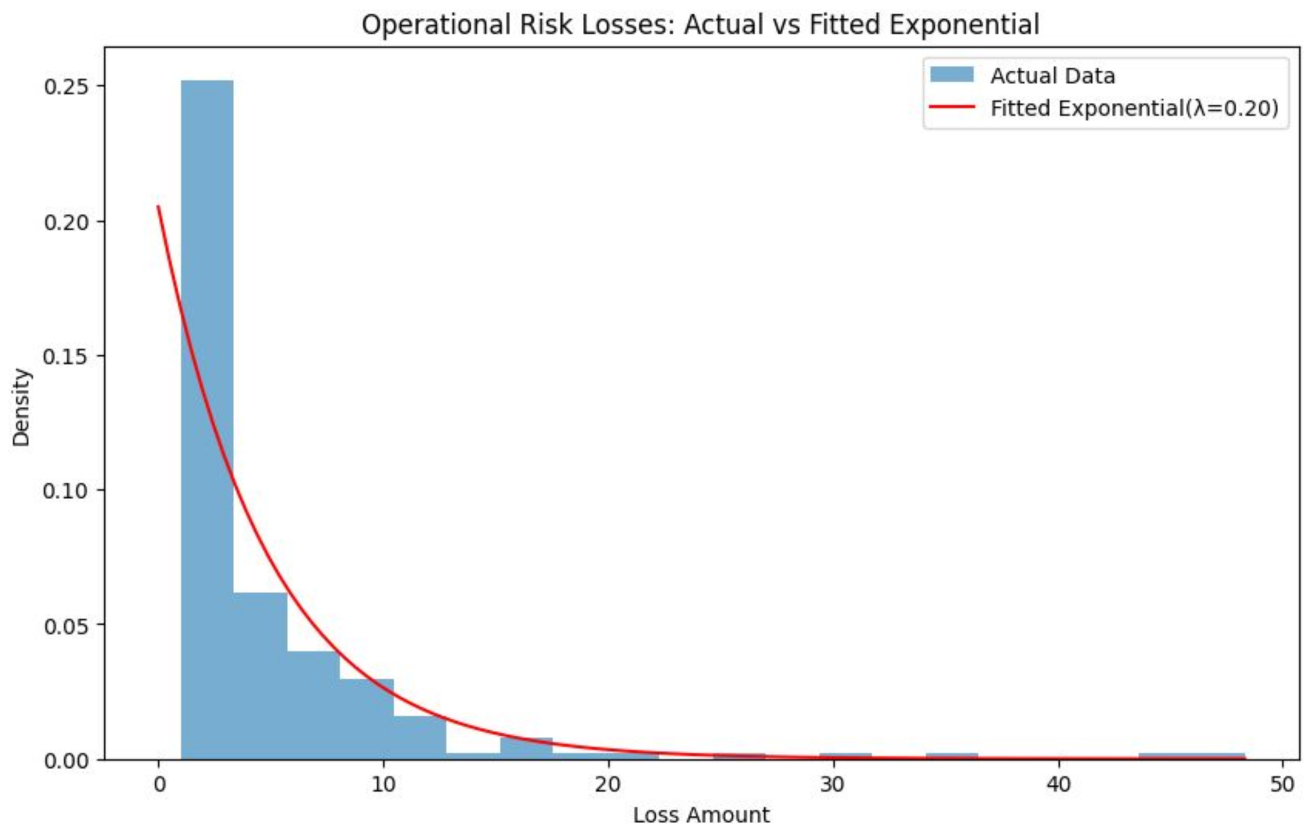
- Summary statistics
- Mean, SD, skew
- Goodness of Fit tests - Anderson-Darling; Kolmogorov-Smirnov
- QQ-plot for exp distribution
- Actual vs Fitted Exponential -
  - by producing key statistical measures including an estimated lambda - 0.20, fitted distribution mean of 4.88, variance of 23.81, and a 12.88% probability of losses exceeding 10

```
DataFrame with Exponential
Distribution PDF values:
       Amount    exp_pdf
0    1.941270   0.156292
1    2.178559   0.147017
2    2.593938   0.132088
3    1.829758   0.160850
4    2.328230   0.141453
..        ...        ...
208  1.405350   0.179448
209  3.359607   0.108427
210  6.113823   0.053306
211  1.611940   0.170140
212  6.825473   0.044371

[213 rows x 2 columns]
```

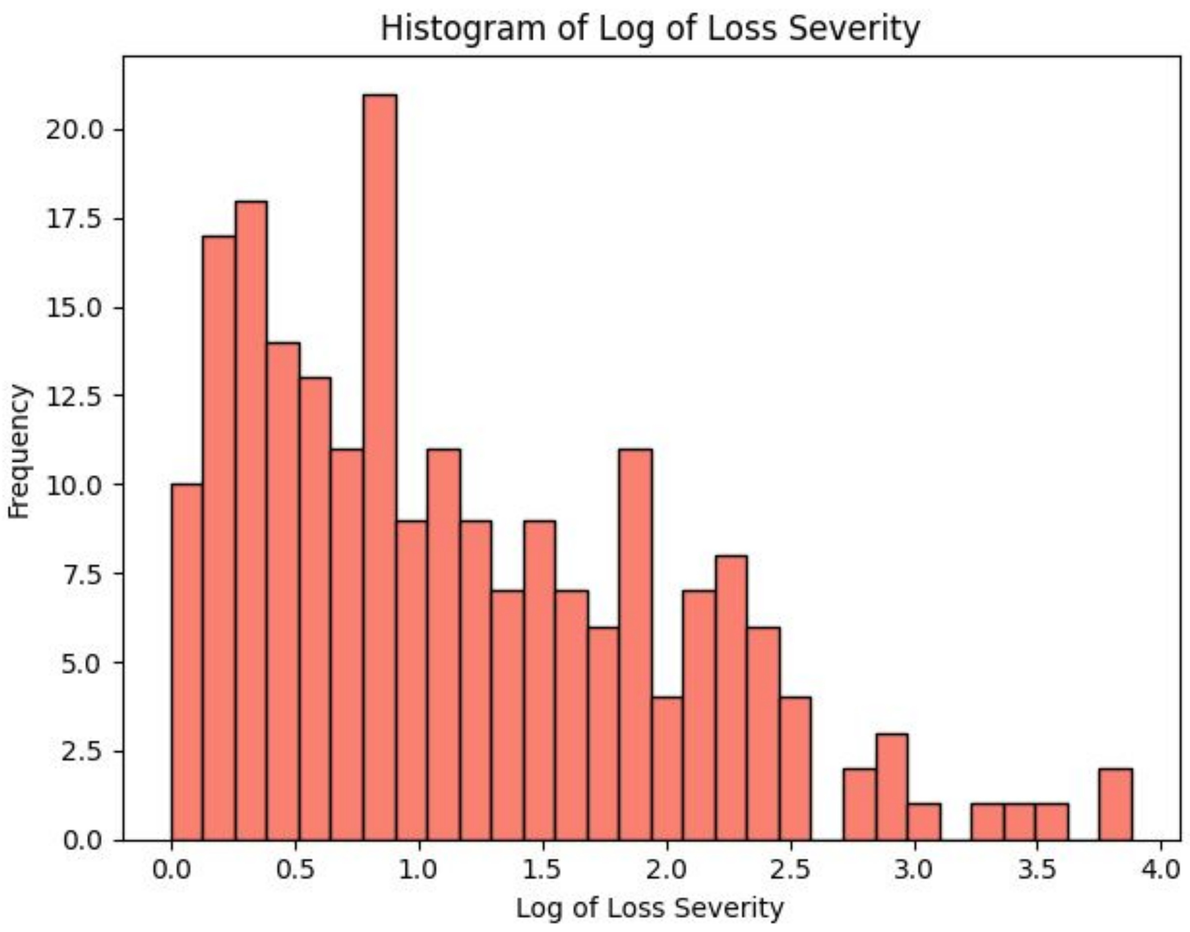
```
Anderson-Darling (AD) Test for Exponential
Distribution:
If the AD test statistic is greater than the critical
value, reject the null hypothesis that the data comes
from an exponential distribution.
AD statistic: 8.776875861481358
Critical values: [0.919 1.075 1.337 1.601 1.952]
Significance levels: [15.  10.   5.   2.5  1. ]

Kolmogorov-Smirnov (KS) Test for Exponential
Distribution:
KS statistic: 0.17798298123156747
p-value: 2.2519055092391495e-06
If the p-value is below the threshold (e.g., 0.05),
reject the null hypothesis that the data follows an
exponential distribution.
```

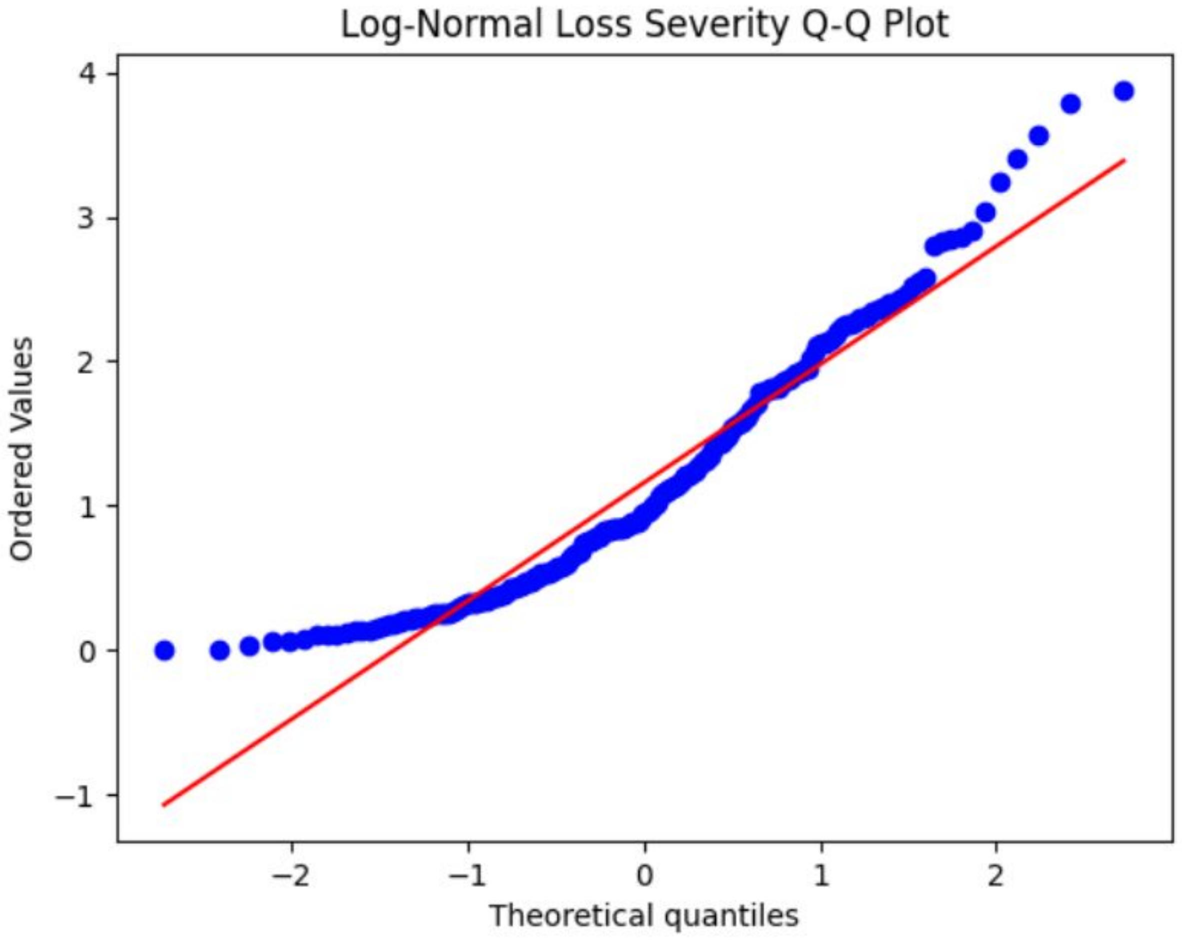
	mu	sig	skew
0	1.070352	1.257332	1.421599





Log-Normal Distribution





- 3 distinct normality tests reveal significant deviations from normality.
- Despite the log-transformation improvements to the distribution's shape, loss severity maintains significant non-normal characteristics.

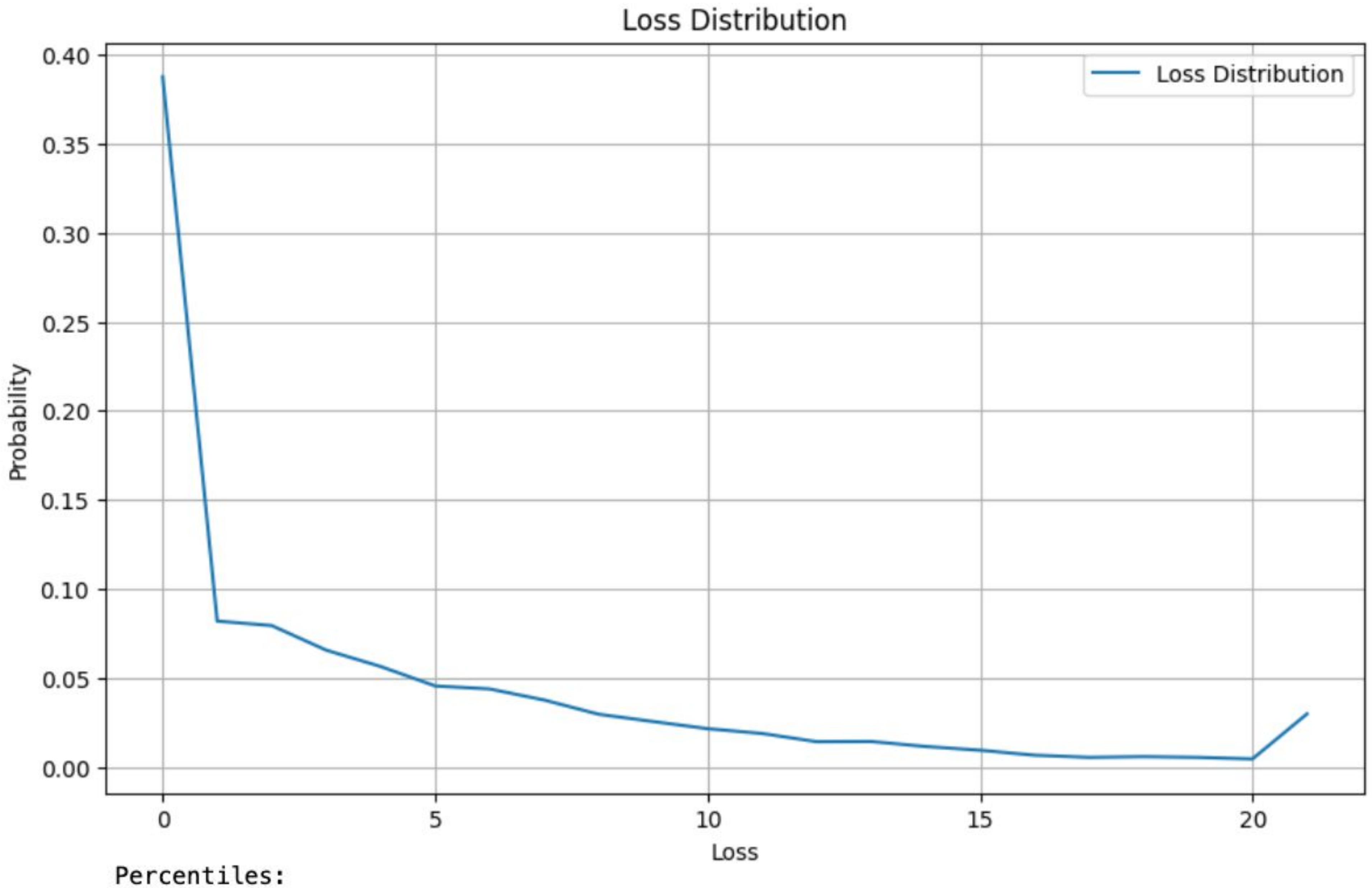
Loss Severity Goodness-of-fit test

```
Anderson-Darling statistic: 3.918077907215462
Critical values: [0.566 0.644 0.773 0.902 1.072]
Significance levels: [15.  10.   5.   2.5  1. ]
Reject null hypothesis of normality if test
statistic
is greater than the critical value at a chosen
significance level
```

```
KS statistic: 0.4819715547651897
p-value: 4.756873059485521e-46
Reject null hypothesis of normality if
p-value < 0.05
```

```
Shapiro-Wilk statistic: 0.932827675062277
p-value: 2.570577385689424e-08
Reject null hypothesis of normality if
p-value < 0.05
```

Poisson-LogNormal Loss Distribution



Percentiles:

50th percentile: 1.78

75th percentile: 6.23

90th percentile: 12.44

95th percentile: 17.85

99th percentile: 34.62

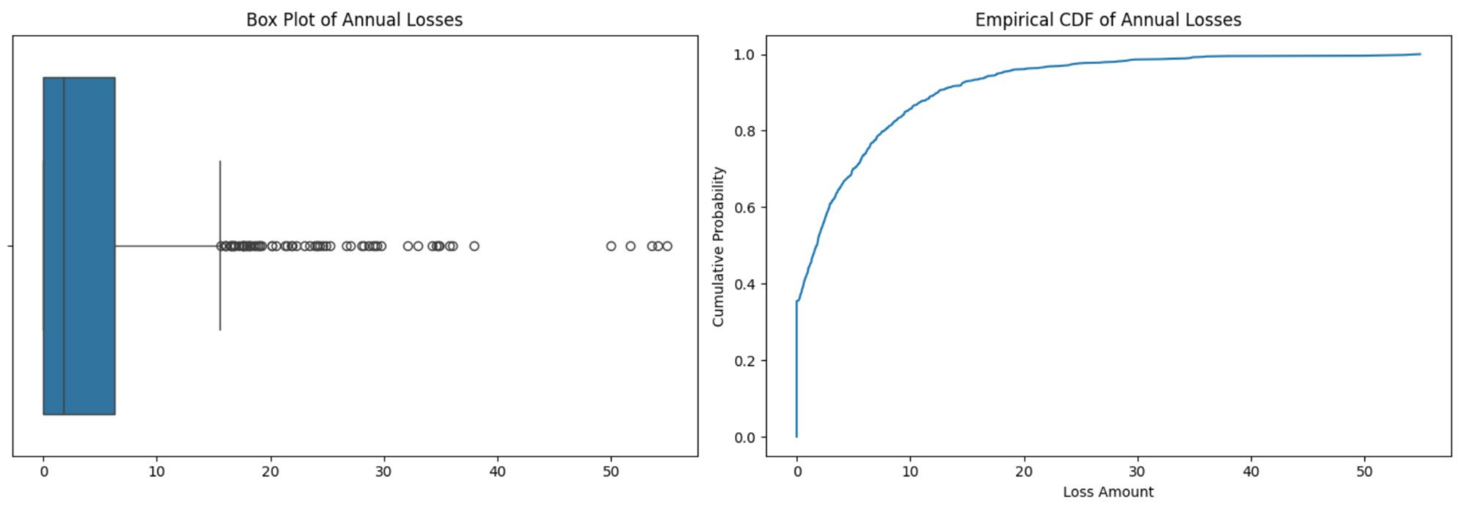
Monthly Stress Loss (90th percentile): 12.49

Annual Stress Loss: 149.93

Estimate Stability (95% Confidence Intervals):

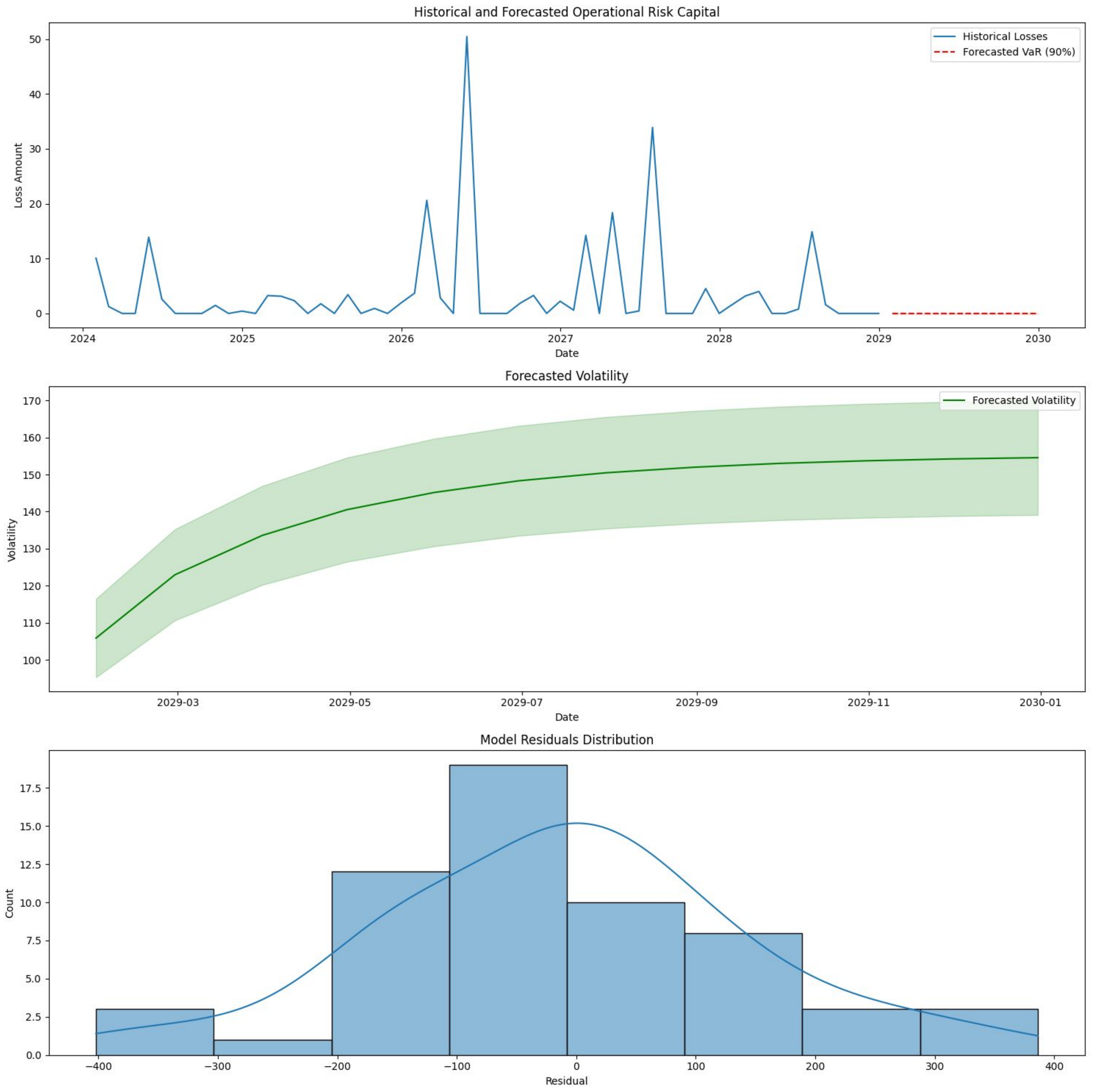
VaR_95: (15.98, 20.15)

ES_95: (24.51, 32.59)



- Highly skewed distribution of operational losses with a mean of \$4.5m and standard deviation of \$7.18m, indicating substantial variability.
- The risk percentiles also show potential for significant tail risk (escalating risk levels)..
- The box-plot shows that there are outliers (beyond the whiskers).

GARCH (1,1) Prediction

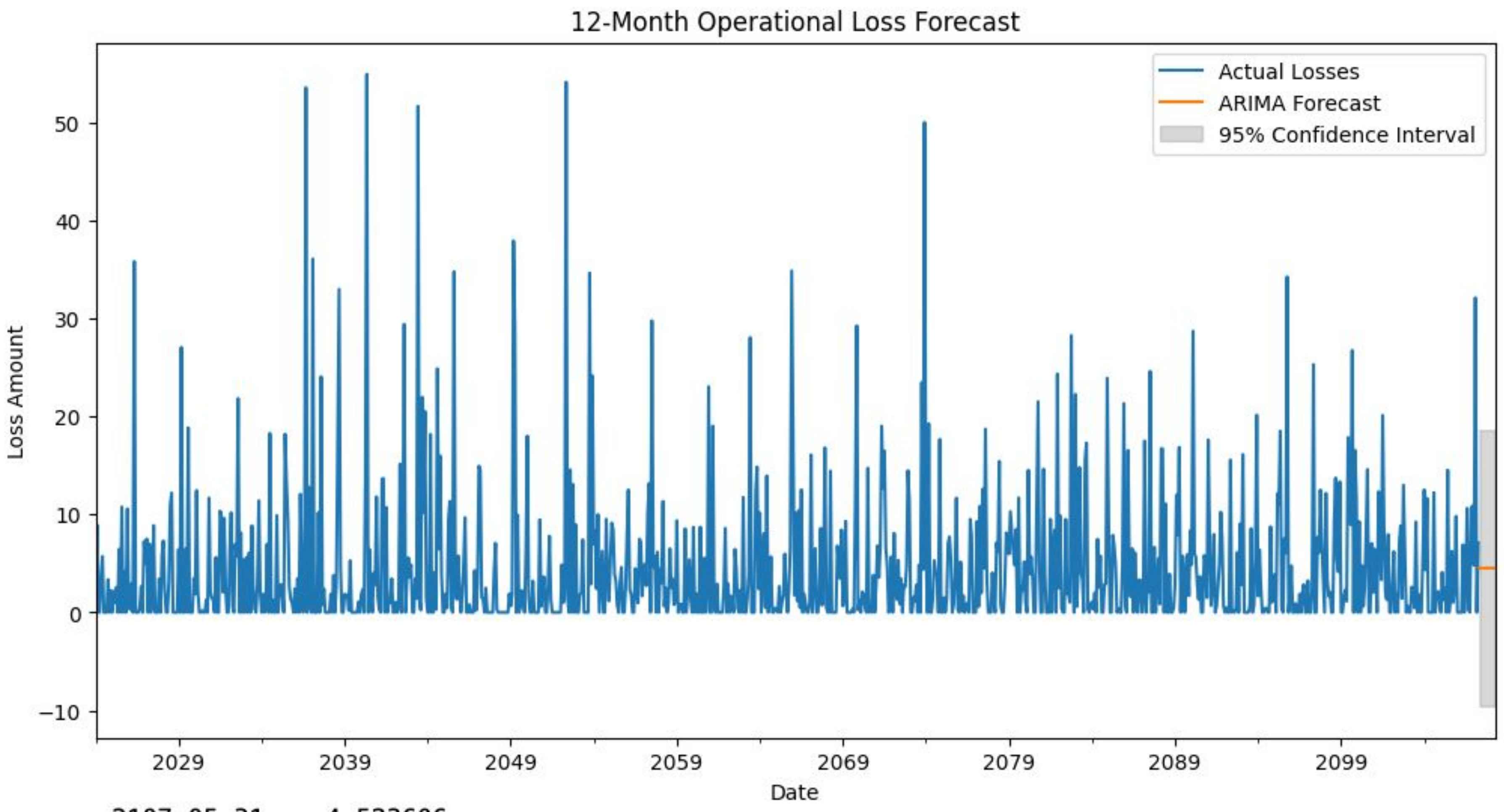



```
Operational Risk Capital Forecast Analysis
--------------------------------------------------

Monthly Forecasts:
    Month  Forecasted_VaR  Volatility
0       1            0.00      105.88
1       2            0.00      122.96
2       3            0.00      133.58
3       4            0.00      140.52
4       5            0.00      145.16
5       6            0.00      148.31
6       7            0.00      150.47
7       8            0.00      151.96
8       9            0.00      152.99
9      10            0.00      153.71
10     11            0.00      154.20
11     12            0.00      154.55

Forecast Summary Statistics:
Average Forecasted VaR: $0.00
Maximum Forecasted VaR: $0.00
Minimum Forecasted VaR: $0.00
VaR Range: $0.00

Volatility Analysis:
Average Forecasted Volatility: 14285.78%
Maximum Forecasted Volatility: 15454.63%
Volatility Trend: Increasing
```

```
GARCH Model Summary:
--------------------------------------------------
                     Constant Mean - GARCH Model Results
==============================================================================
Dep. Variable:                          y   R-squared:                   0.000
Mean Model:                 Constant Mean   Adj. R-squared:              0.000
Vol Model:                          GARCH   Log-Likelihood:           -378.085
Distribution:      Standardized Student's t   AIC:                       766.171
Method:                Maximum Likelihood   BIC:                       776.559
                                            No. Observations:               59
Date:                    Wed, Nov 13 2024   Df Residuals:                   58
Time:                            21:37:26   Df Model:                        1
                                 Mean Model
==========================================================================
                 coef    std err          t      P>|t|  95.0% Conf. Int.
--------------------------------------------------------------------------
mu             7.8373     14.821      0.529      0.597 [-21.211, 36.886]
                              Volatility Model
==============================================================================
                 coef    std err          t      P>|t|       95.0% Conf. Int.
------------------------------------------------------------------------------
omega       7299.3554   4782.337      1.526      0.127 [-2.074e+03,1.667e+04]
alpha[1]       0.3716      0.174      2.131  3.309e-02    [2.982e-02,  0.713]
beta[1]        0.3259      0.284      1.147      0.252     [ -0.231,  0.883]
                                 Distribution
============================================================================
                 coef    std err          t      P>|t|     95.0% Conf. Int.
----------------------------------------------------------------------------
nu           189.5979     99.358      1.908  5.636e-02 [ -5.140,3.843e+02]
============================================================================
```

- Poor explanatory power based on Adjusted R-squared at 0.000.
- High uncertainty in mean estimate at the 95% confidence interval.
- The model may be unable to capture long -term volatility dynamics.
- Further investigation into zero VaR forecasts.

ARIMA Prediction



2107-05-31	4.523606
2107-06-30	4.518208
2107-07-31	4.518197
2107-08-31	4.518197
2107-09-30	4.518197
2107-10-31	4.518197
2107-11-30	4.518197
2107-12-31	4.518197
2108-01-31	4.518197
2108-02-29	4.518197
2108-03-31	4.518197
2108-04-30	4.518197

- **ARIMA Model Implementation**
  - We Applied ARIMA(1,1,1) model for 12-month operational loss forecasting
  - Monthly frequency data starting from January 2024

- **Historical Loss Patterns**
  - High volatility in actual losses (blue line)
  - Multiple significant spikes reaching 50 units

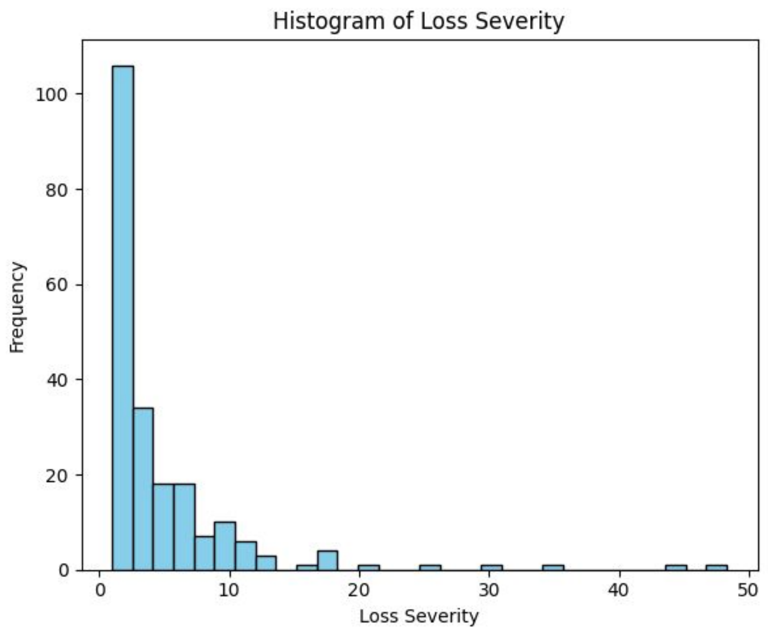
- **Forecast Characteristics**
  - Predicts relatively stable future trend around 5 units
  - Conservative estimates compared to historical volatility
  - Reflects increasing uncertainty in longer-term predictions

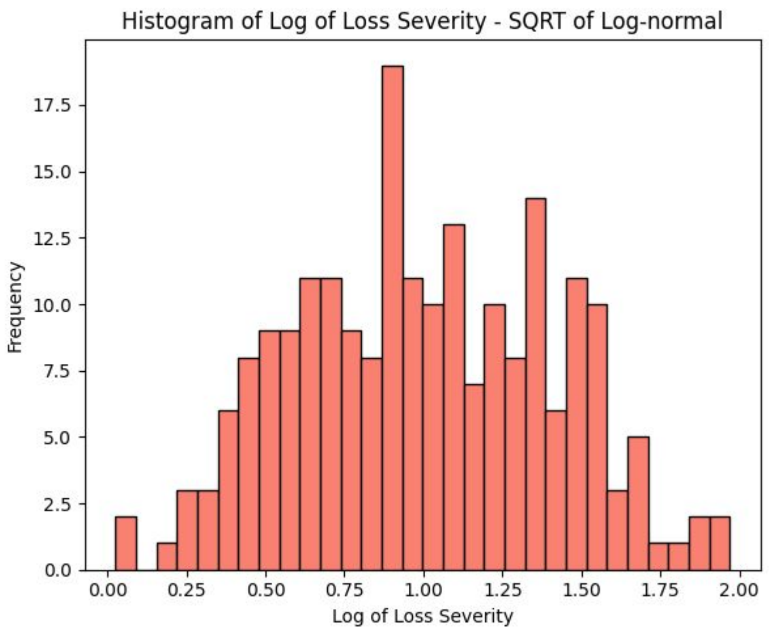
- **Model Limitations**
  - May underestimate potential extreme loss events
  - Assumes some level of predictable pattern in losses
  - Struggles to capture the sporadic nature of operational risk

- **Key Implications**
  - Need for additional stress testing scenarios
  - Potential for model enhancement using alternative approaches

Limitations and Recommendations for further analysis

- Model useability limited to this dataset.
- Modelling after data transformation - eg. increasing threshold, removing outliers,
- Other Transformations techniques on Log-normal or Exponential distribution: Power Transformations, Stabilizing Transformations, Johnson Transformation System, Generalized Pareto Transform, Severity-Specific Scaling, Composite Transformation.
- Limitations of ILD - ILD data available only from past 20 years. Need to Consider a combination of historic loss data and scenario analysis for modelling of individual risks
    - *'...most organisations have only been systematically collecting operational risk data for 15 years or less. This is unlikely to be sufficient to model low frequency, high impact operational risks faced by firms'. (Kelliher POJ, Acharyya M, Couper A, et al. 2017)*
- **Limitations of Lognormal Distribution:** The lognormal distribution may not accurately capture the extreme tails of operational risk data, where very large losses can occur infrequently. *(Peters, G. W., Shevchenko, P., Young, M., & Yip, W. (2017))*
- Other models: Negative binomial, SQRT of Log-Normal distribution.





Conclusion

This project successfully applied the Loss Distribution Approach (LDA) to forecast operational losses for the next 12 months. By analyzing the historical loss data from the Internal Loss Database (ILD), we identified key trends and patterns in the frequency and severity of losses.

The analysis involved careful model selection, parameter estimation, and simulation techniques. While the lognormal distribution is a common choice, we recognized its limitations in capturing extreme tail events. To address this, we explored alternative distributions like the α-stable distribution, which can better accommodate rare but severe losses.

The resulting loss distribution provides valuable insights into the potential range of future losses and associated risks. This information can inform risk management strategies, capital allocation decisions, and regulatory compliance efforts. However, it's crucial to acknowledge the inherent uncertainties in forecasting and to continuously monitor and update the model to adapt to changing risk profiles.

REFERENCES

- Peters, G. W., & Sisson, S. A. (2009). Bayesian Inference, Monte Carlo Simulation, and Operational Risk. *Journal of Operational Risk*, 3(3), 3-33.
- Dutta, K. C., & Perry, J. (2012). A Tale of Tails: An Empirical Analysis of Loss Distribution Models for Estimating Operational Risk. *Journal of Operational Risk*, 5(3), 39-78.
- Peters, G. W., Shevchenko, P., Young, M., & Yip, W. (2017). Analytic Loss Distributional Approach Models for Operational Risk from Compound Processes and Implications for Capital Allocation. *Journal of Operational Risk*, 11(2), 3-33.
- Kuketayev, A. (2024). Lectures. *Quantitative Risk Management*, George Washington University, October 9 2024,October 16, 2024, October 23,2024.
- Hugen, J. (2016). *Risk Management and Insurance: Principles and Practice*. Springer.
- Frachot, A., Georges, P., & Roncalli, T. (2001). Loss Distribution Approach for Operational Risk. *Journal of Operational Risk*, 1(1), 21-41.
- Kelliher POJ, Acharyya M, Couper A, et al. (2017). Good practice guide to setting inputs for operational risk models. British Actuarial Journal. 22(1):68-108. doi:10.1017/S1357321716000179
- Hull, John C.. (2015). Risk management and financial institutions (4). Hoboken: John Wiley.
- El Adlouni, S., Ezzahid, E., & Mouatassim, Y. (2011). Mixed distributions for loss severity modelling with zeros in the operational risk losses. International Journal of Applied Mathematics & Statistics, 21, 96-109.

Thank you!