Integrated Risk Management [IMEN891G]

Cyber-Nat
: Cyber risks caused by Natural-Catastrophes

Actuarial modeling, Insurance and Risk Management Lab,
POSTECH

Keywoong Bae

목차

Contents of Research Proposal

1. Introduction to cyber risk
2. Cyber risk caused by natural catastrophes
3. Purpose of this research

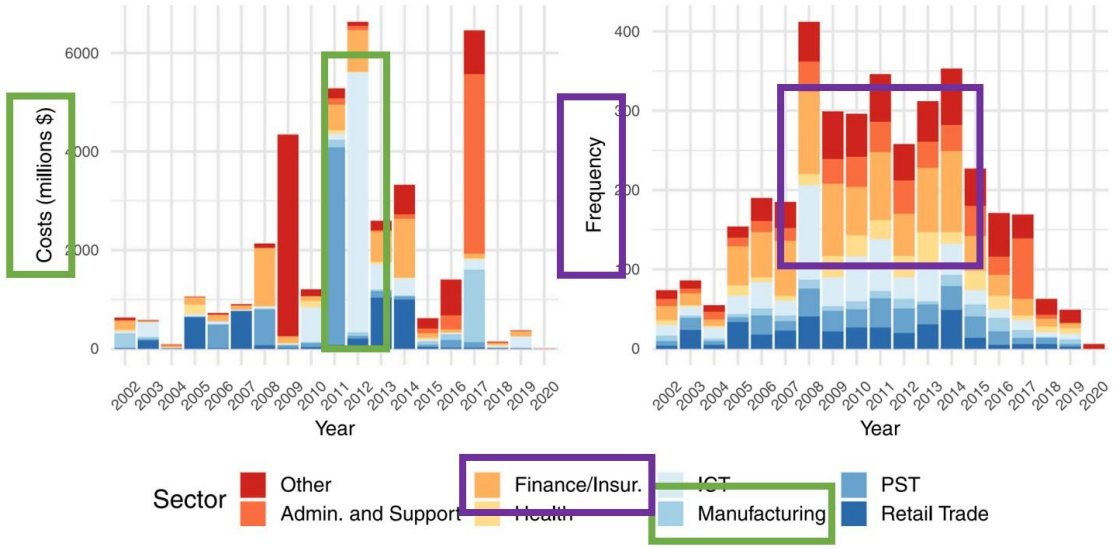
1. Introduction to cyber risk

- Cyber risk?
  - Cyber risk refers to the risks of financial loss, business interruption, and reputational damage resulting from IT system failures. [1]
  - The risk of cyber risks is increasing, and the annual damage is expected to reach $100 billion. [2]
  - The risk of exposure to cyber risks increases due to the development of IT technology, COVID-19, and the increase in digital technology.. [1,2]
  - Cyber risk is attracting attention because it is difficult to quantify and related information is lacking, and the need for related research is emerging.
- Examples of cyber risk
  - In the process of merging two financial institutions, customer deposit data is lost due to a conflict in the systems of the two institutions..
  - Health insurance agency (Oxford Health Plans Inc) incorrectly upgrades its computer systems, resulting in claims processing system failure. This led to long-term delays in distributing bills and delayed payment to doctors who provided medical services.
  - An employee of a financial institution embezzled funds from customer accounts and traded unauthorized options between 1993 and 1998. This securities fraud resulted in a loss of $2.8 million.

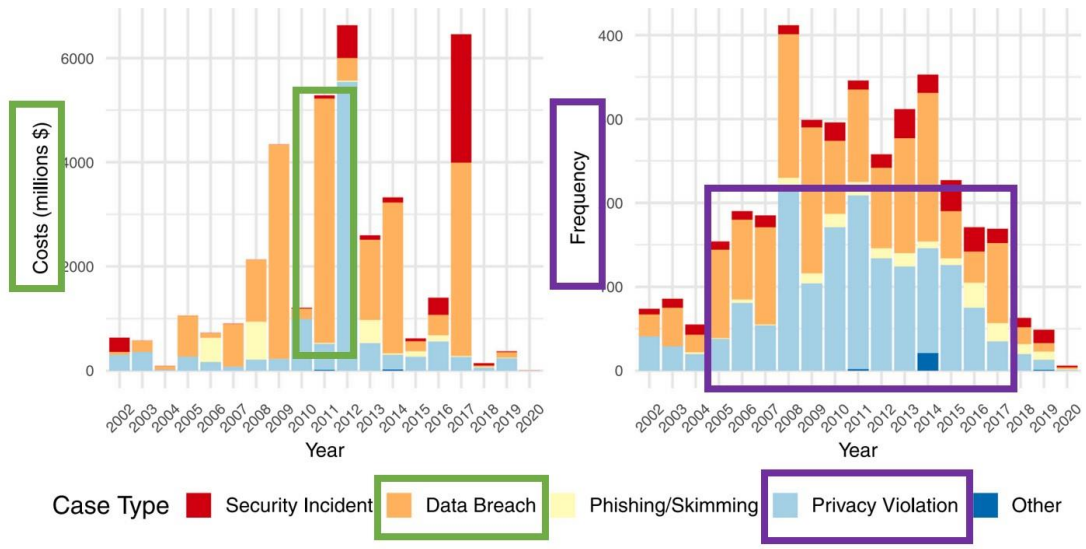
[1] Aldasoro, Iñaki, et al. "The drivers of cyber risk." Journal of Financial Stability 60 (2022): 100989.
[2] Eling, Martin, et al. "The economic impact of extreme cyber risk scenarios." North American Actuarial Journal 27.3 (2023): 429-443.

1. Introduction to cyber risk

- The influence of cyber risks is growing
  - The areas and types of cyber risks are very diverse..
    - Representative fields of occurrence include the financial and manufacturing fields, and types of occurrence include data breach and privacy violation..



Frequency and cost of cyber incidents across sectors



Frequency and cost of cyber incidents by case type

- Typically, <u>cyber risk cases caused by natural catastrophes</u> are quite **remarkable** types.

2. Cyber risk caused by natural catastrophes

- Cyber risks caused by natural catastrophes

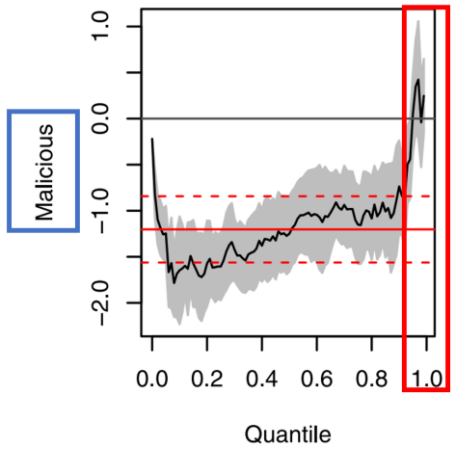
HURRICANE LAURA LIVE UPDATES

Hurricane Laura Knocks Out Power For Hundreds Of Thousands In Louisiana And Texas

AUGUST 27, 2020 · 12:01 PM ET

By Samantha Raphelson

2. Cyber risk caused by natural catastrophes

- The meaning of cyber-nat risk caused by natural catastrophes
  - Loss distribution due to operational risk and cyber risk is characterized by a heavy tail.. [1]
  - 

  - Cyber-nat risk is also a type of cyber risk, and although it occurs very rarely, it has the characteristic of large losses when it occurs..
  - This corresponds to a tail distribution, and prior understanding of the potential damage from this risk is very important..
  - By analyzing cyber risks caused by unexpected natural disasters, it is possible to set appropriate insurance rates and effectively distribute risks.
  - We analyze cyber incidents caused by natural disasters and discover their characteristics and causes to prepare in advance..

[1] Aldasoro, Iñaki, et al. "The drivers of cyber risk." Journal of Financial Stability 60 (2022): 100989.

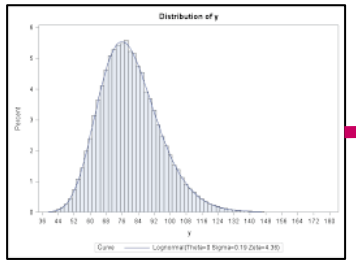
3. Purpose of this research

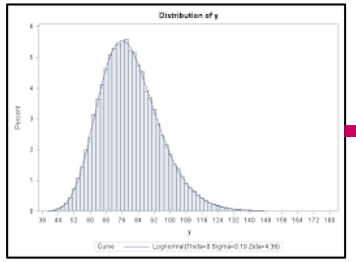
- 연구의 목적
  - Based on cyber risks and natural catastrophes, we summarize their characteristics through statistical analysis.
  - Based on the analysis of cyber risks and natural disaster risks, determine how they are related to cyber risks arising from natural disasters.
  - Focusing on 12 cyber-nat risks, we observed how cyber risks occur due to natural disasters and established scenarios and risk management strategies for them.

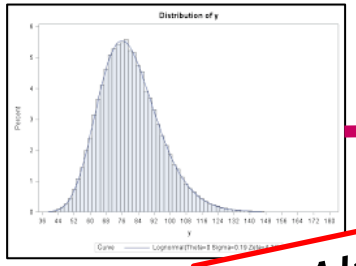
[1] Aldasoro, Iñaki, et al. "The drivers of cyber risk." Journal of Financial Stability 60 (2022): 100989.

3. Purpose of this research

Step 1. Loss Distribution Approach







~ 11/12

Step 2. Loss Factor Analysis & Prediction

- Linear Regression
- Generalized Linear Model (GLM)
- Generalized Additive Model (GAM)
- Machine Learning / Deep Learning methods

- MSE
- RMSE
- Etc..

~ 11/26

Step 3. Scenario analysis

Case I. Typhoon
….

Case II. Hurricane
….

Case III. Forest Fire
….

~ 12/3

목차

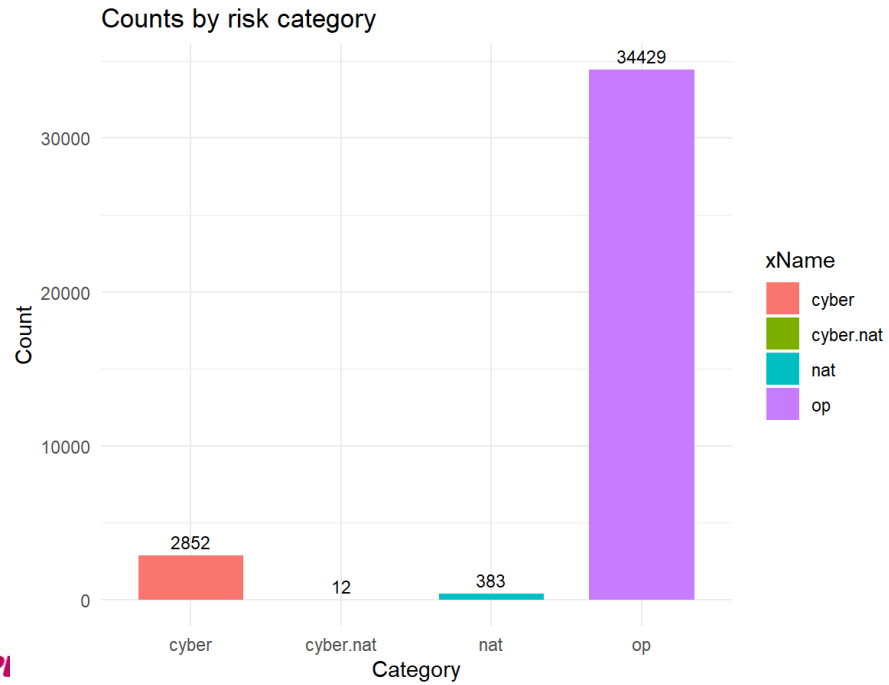
Contents of Research Results

1. Datasets Risk Type Classification
2. Empirical Distribution Fitting of Cyber risk (frequency, severity)
3. Empirical Distribution Fitting of Natural Catastrophes (frequency, severity)
4. Variable Selection with Lasso Regression
5. GLM regression and Factor analysis (loss severity)
6. Conclusion

1. Datasets Risk Type Classification

SAS OpRisk Global data

- It is a dataset that collects operational losses and is the largest dataset in the risk management field. ^**1**
- A dataset that covers risks in a variety of cases and focuses on risks with losses exceeding $100,000 ($ 0.1 M).
- From March 1, 1971 to April 1, 2021, a total of 37652 events are covered.
- Since this dataset isn't divided, so we classified this dataset into four categories: Cyber, Nat-cat, Cyber-nat, Operational risk
- With the "description" column of this dataset, we used two classifying methodologies: (1) Keyword based, (2) BERT classification
- In this study, we plan to classify events into four types of risks and conduct statistical analysis for each case.



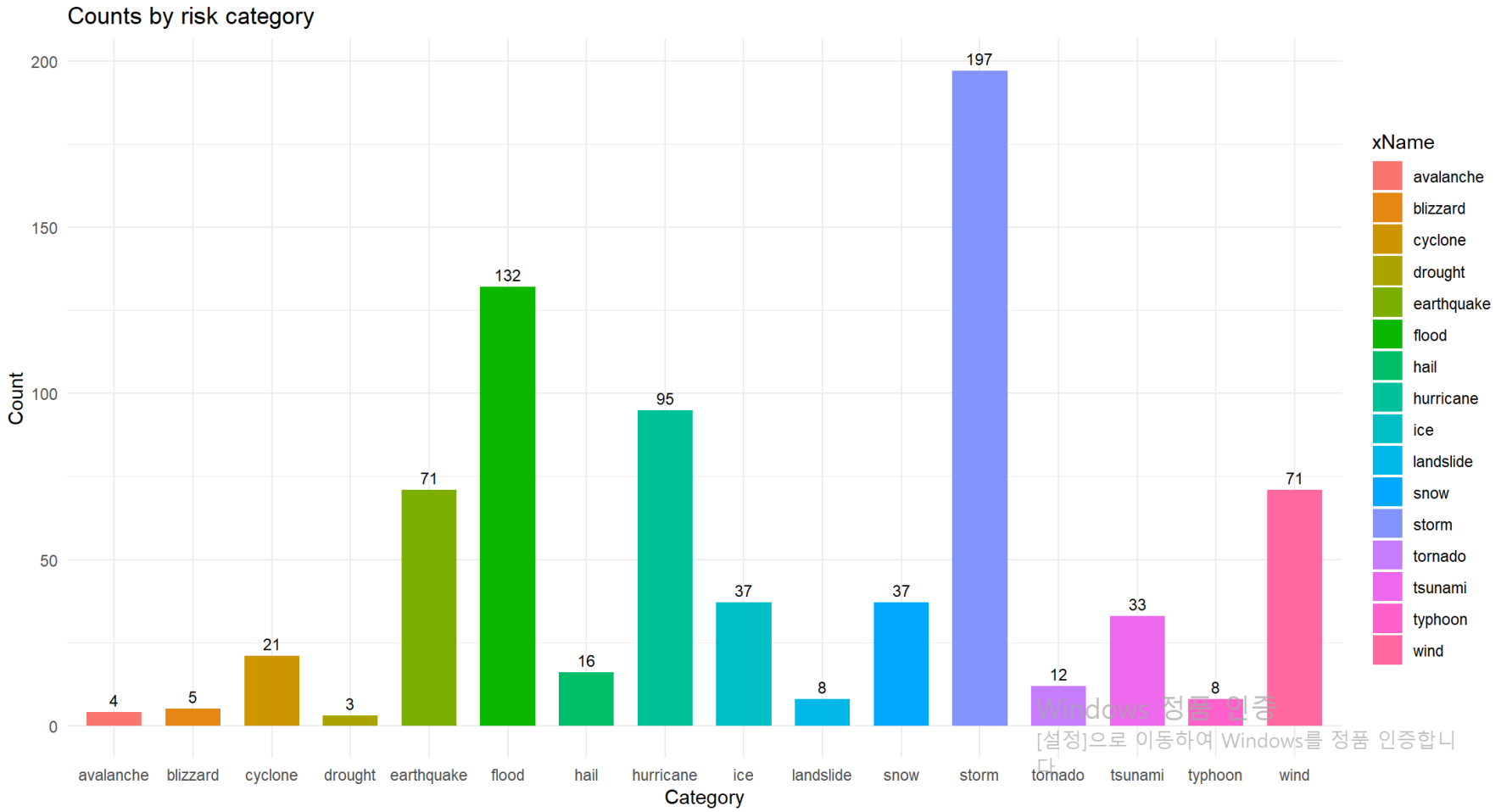
Risk Type	count
Cyber risk	2852
Natural Catastrophes	383
Operational risk	34429
Cyber nat	12

[1] Eling, Martin and Ibragimov, Rustam and Ning, Dingchen, Time Dynamics of Cyber Risk (July 2, 2023).

1. Datasets Risk Type Classification

SAS OpRisk Global data

- The SAS dataset includes a total of 16 natural disaster-related events, details of which are as follows.



Cyber-Nat: Cyber risks caused by Natural-Catastrophes

2-1. Empirical Distribution Fitting of Cyber risk (Loss frequency)

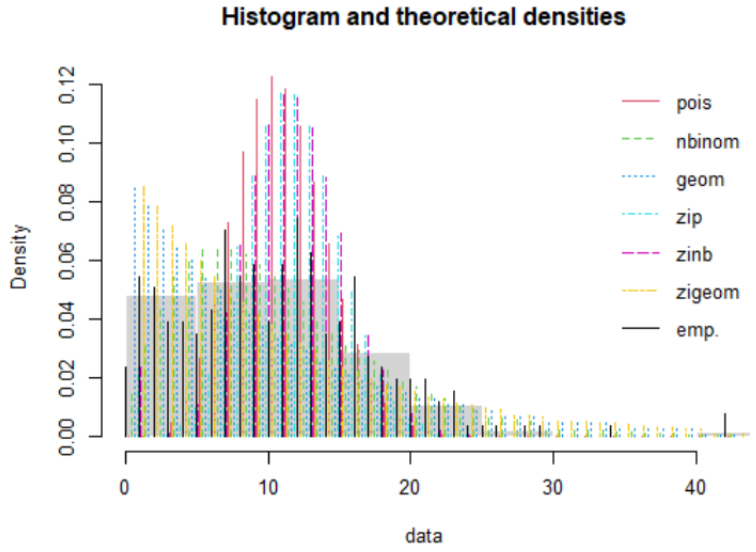
- Frequency Distribution of Cyber risk
  - Loss frequency is typically modeled either by a Poisson or a negative binomial distribution.
  - Based on this distributions, we used additional distributions (geometric distribution, and zero-inflated discrete distributions).
  - We used Chi-squared test for fitting: $\chi^2 = {}^{(n-1)s^2}/_{\sigma^2}$
- Table of Chi-squared test result
  -

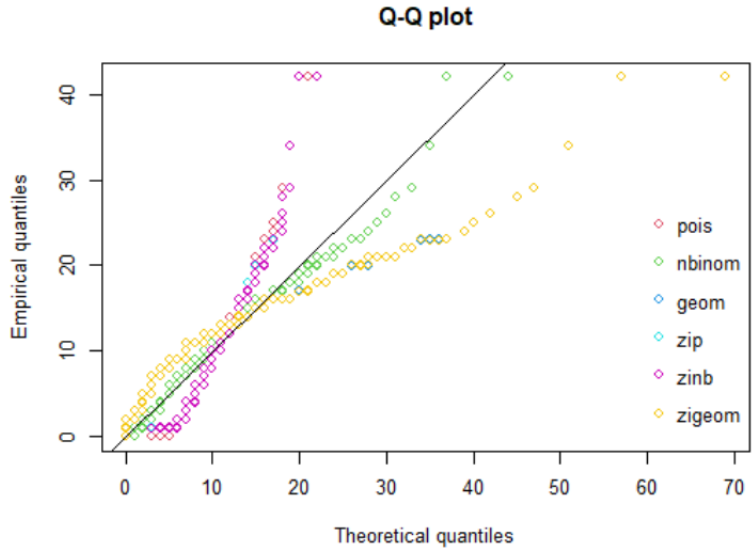
	pois	nbinom	geom	zip	zinb	zigeom
χ^2-test	6494.578	**24.05419**	83.97248	810.1257	756.7123	83.97048
P-value	0	**0.00421645**	8.320e-14	1.4056e-168	4.3752e-158	8.328e-14
AIC	2208.341	**1685.127**	1749.131	2137.739	2125.177	1749.131
BIC	2211.886	**1692.217**	1752.676	2144.829	2135.812	1752.676

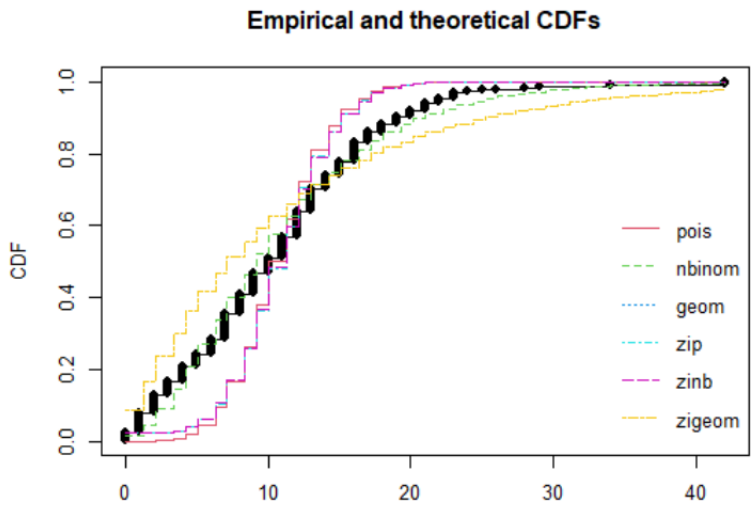
Cyber-Nat: Cyber risks caused by Natural-Catastrophes

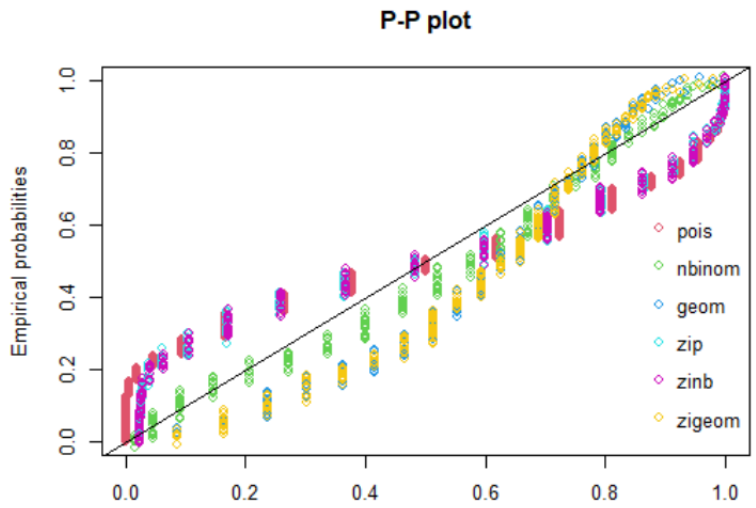
2-1. Empirical Distribution Fitting of Cyber risk (Loss frequency)

- Plot of Chi-squared test result









Cyber-Nat: Cyber risks caused by Natural-Catastrophes

2-2. Empirical Distribution Fitting of Cyber risk (Loss severity)

- Severity Distribution of Cyber risk
  - For the loss severity, we fit the data to Weibull, gamma, and log-normal distribution, which is commonly used in actuarial literatures.
  - We used Kolomogorov-Smirov test for fitting: $D = \max_{1 \le i \le n} |F(x_i) - \frac{i}{n}|$
- Table of KS-squared test result
  -

Gofstat results	weibull	gamma	log-normal
KS test	0.1538082	0.2258349	**0.07036063**
AIC	17549.58	19006.04	**16565.41**
BIC	17561.49	19017.95	**16577.32**

Cyber-Nat: Cyber risks caused by Natural-Catastrophes

2-2. Empirical Distribution Fitting of Cyber risk (Loss severity)

- Estimated distributions of Cyber loss frequency and severity
  - Cyber loss frequency: Negative-binomial ($size = 2.536241, \mu = 10.665716$)
  - Cyber loss severity: Log-normal distribution ($\mu = 1.664924, \sigma = 2.106628$)

Cyber-Nat: Cyber risks caused by Natural-Catastrophes

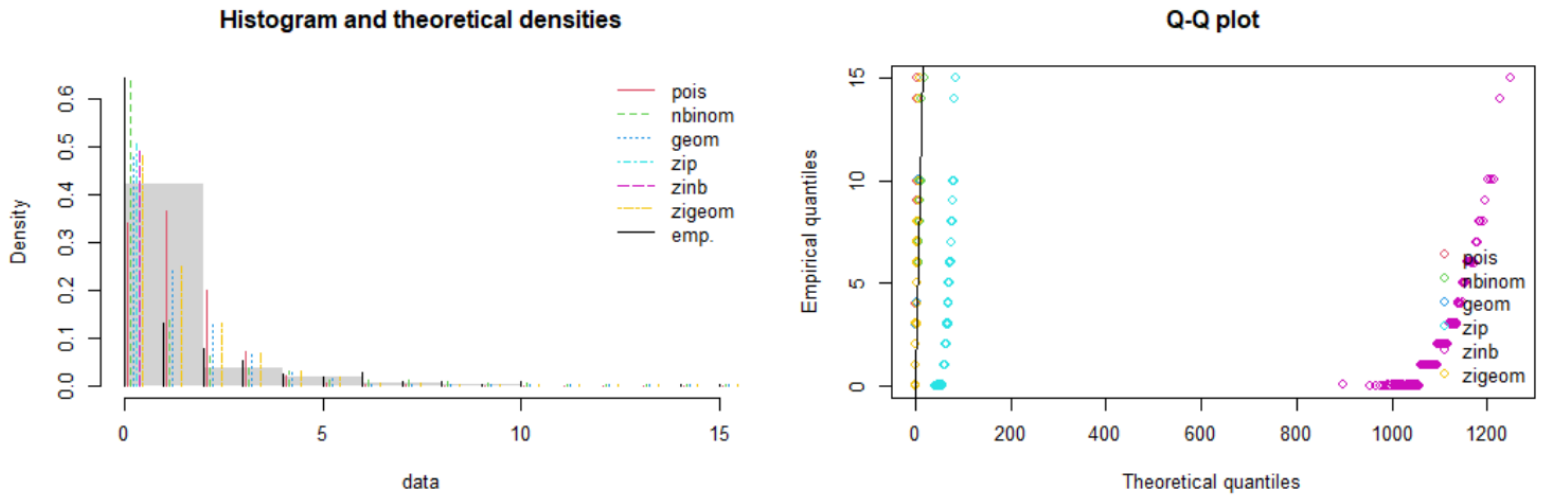
3-1. Empirical Distribution Fitting of Natural Catastrophes (Loss frequency)

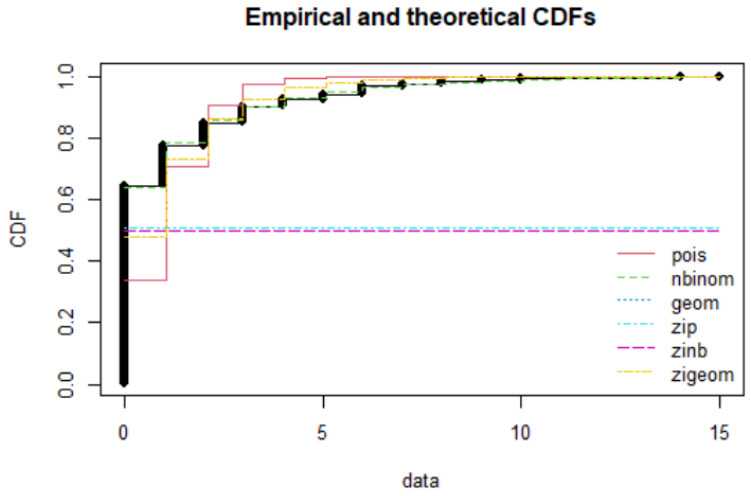
- Frequency Distribution of Natural Catastrophes
  - For a total of 383 natural disaster risk cases, a chi-squared test was performed to estimate the frequency distribution.
  - Like cyber risk, natural disaster risk was also estimated for six discrete distributions: Poisson, negative-binomial, geometric, zero-inflated poisson, zero-inflated negative-binomial, and zero-inflated geometric
- Table of Chi-squared test result
  -

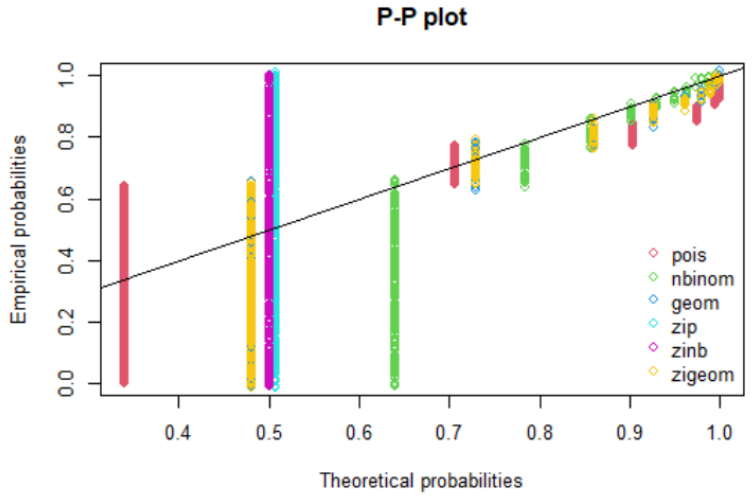
	pois	nbinom	geom	zip	zinb	zigeom
P-Value	5.2270e-115	**0.6678481**	5.5536e-14	0	0	5.5536e-14
AIC	1383.137	957.9549	1030.885	14032.11	154681.4	1030.885
BIC	1387.015	965.7104	1034.762	14039.86	154693.0	1034.762

3-1. Empirical Distribution Fitting of Natural Catastrophes (Loss frequency)

- Plot of Chi-squared test result







3-2. Empirical Distribution Fitting of Natural Catastrophes (Loss severity)

- Severity Distribution of Cyber risk
  - For a total of 383 natural disaster risk cases, a KS-test was performed to estimate the severity distribution.
  - Like cyber risk, natural disaster risk was also estimated for three continuous distributions.
- Table of KS-test result
  -

Gofstat results	weibull	gamma	log-normal
KS test	0.07131231	0.1732938	**0.02686458**
AIC	3933.814	4065.805	**3865.712**
BIC	3941.726	4073.717	**3873.624**

Cyber-Nat: Cyber risks caused by Natural-Catastrophes

3-2. Empirical Distribution Fitting of Cyber risk (Loss severity)

- Estimated distributions of Natural catastrophes loss frequency and severity
  - Nat-cat loss frequency: Negative-binomial ($size = 0.283164, \mu = 1.081545$)
  - Nat-cat loss severity: Log-normal distribution ($\mu = 2.883738, \sigma = 2.012829$)

4. Variable Selection with Lasso Regression (loss severity)

- Lasso Regression analysis
  - Lasso regression is a type of regularized linear regression that prevents the model from being overfitted by adding constraints on linear regression coefficients.
  - Using the Lagrangian coefficient λ, it is embedded in the objective function of the optimization problem and optimized together.
  - When constraints are added, some of the unimportant predictors depending on their size are made 0 and excluded from the model.
- As a result of Lasso regression analysis, **cyber risk** showed no significant relationship with sub.risk.disaster, sub.risk.thefts, region.europe, and sector.manufactoring.
- **Nat-Cat risk** was shown to be unrelated to sub.risk.systems and sector.manufactoring.

Cyber-Nat: Cyber risks caused by Natural-Catastrophes

5. GLM regression and Factor analysis (loss severity)

- GLM regression of cyber risk

```
Call:
glm(formula = scope, family = gaussian(link = "identity"), data = new.cyber)

Coefficients: (1 not defined because of singularities)
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                1.755e+00  2.002e-01   8.764  < 2e-16 ***
cyber.assets               1.625e-07  7.813e-08   2.079 0.037690 *
cyber.employees           -1.248e-07  5.386e-07  -0.232 0.816836
cyber.revenue              1.159e-06  1.292e-06   0.897 0.370030
cyber.legal.liability      4.555e-03  4.379e-04  10.401  < 2e-16 ***
sub.risk.customer.intake   4.350e-02  1.331e+00   0.033 0.973932
sub.risk.account.management -3.037e-01  1.987e-01  -1.528 0.126605
sub.risk.disaster          1.263e+00  3.869e-01   3.264 0.001112 **
sub.risk.improper.business 1.016e+00  1.262e-01   8.051 1.24e-15 ***
sub.risk.monitoring        1.193e-01  2.269e-01   0.526 0.598939
sub.risk.product.flaws     1.478e+00  4.508e-01   3.278 0.001059 **
sub.risk.selection         1.310e-01  9.434e-01   0.139 0.889558
sub.risk.suitability       1.029e+00  1.777e-01   5.790 7.90e-09 ***
sub.risk.customer                 NA         NA      NA       NA
sub.risk.systems           1.218e+00  1.799e-01   6.770 1.59e-11 ***
sub.risk.systems.security  2.893e-01  1.155e-01   2.505 0.012304 *
sub.risk.transaction       4.224e-01  1.838e-01   2.299 0.021606 *
sub.risk.unauthorized     -1.186e-01  1.320e-01  -0.899 0.368821
sub.risk.vendors           1.084e+00  2.100e+00   0.516 0.605889
region.north.america      -2.990e-01  9.644e-02  -3.100 0.001953 **
region.asia               -1.600e-01  1.224e-01  -1.307 0.191247
region.other.americas     -1.814e-01  3.513e-01  -0.516 0.605661
region.africa             -2.688e-01  2.508e-01  -1.072 0.283873
region.other              -1.868e-01  2.228e-01  -0.839 0.401812
sector.financial.services -1.501e+00  1.847e-01  -8.129 6.68e-16 ***
sector.retail.trade       -3.666e-01  3.149e-01  -1.164 0.244462
sector.professional       -1.796e-01  3.678e-01  -0.488 0.625355
sector.information        -7.491e-01  2.221e-01  -3.373 0.000754 ***
sector.arts               -1.061e+00  7.320e-01  -1.449 0.147526
sector.mining             -1.217e+00  6.054e-01  -2.010 0.044489 *
sector.construction        3.242e-01  9.560e-01   0.339 0.734518
sector.accommodation      -2.181e-01  4.941e-01  -0.441 0.658979
sector.management         -1.350e+00  1.887e+00  -0.715 0.474413
sector.utilities          -1.008e+00  3.874e-01  -2.603 0.009293 **
sector.administrative     -6.646e-01  4.483e-01  -1.483 0.138308
sector.transportation     -2.469e-01  3.869e-01  -0.638 0.523452
sector.public              2.873e-02  9.502e-01   0.030 0.975877
sector.wholesale.trade    -1.004e+00  8.582e-01  -1.170 0.242056
sector.health             -6.206e-01  4.509e-01  -1.376 0.168794
sector.other              -3.507e+00  1.889e+00  -1.857 0.063415 .
sector.agriculture        -1.437e+00  1.100e+00  -1.307 0.191418
sector.real.estate        -1.589e+00  9.592e-01  -1.657 0.097666 .
sector.non.profit         -2.069e+00  1.343e+00  -1.541 0.123506
sector.educational        -3.755e-01  8.612e-01  -0.436 0.662844
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

회귀분석 결과	변수 명	P-value
0.0045	`cyber.legal.liability`	***
1.1016	`sub.risk`, Improper Business or Market Practices	***
1.029	`sub.risk`, Suitability, Disclosure & Fiduciary	***
1.218	`sub.risk`, Systems	***
-1.501	`sector`, Financial Services	***
-0.749	`sector`, Information	***
1.263	`sub.risk`, Disasters and Other Events	**
1.478	`sub.risk`, Product Flaws	**
-0.299	`region`, North America	**
-1.008	`sector`, Utilities	**
0.00000162	`cyber.assets`	*
0.289	`sub.risk`, Systems Security	*
0.422	`sub.risk`, Transaction Capture, Execution & Maintenance	*
-1.217	`sector`, Mining	*
-3.507	`sector`, Other Services (except Public Administration)	.
-1.589	`sector`, Real Estate, Rental and Leasing	.

Cyber-Nat: Cyber risks caused by Natural-Catastrophes

5. GLM regression and Factor analysis (loss severity)

- GLM regression of natural catastrophes

```
Call:
glm(formula = scope, family = gaussian(link = "identity"), data = new.nat)

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)               1.562e+00  1.487e+00   1.050   0.2944
nat.assets                1.521e-06  1.047e-06   1.453   0.1472
nat.employees             4.104e-07  1.241e-06   0.331   0.7410
nat.revenue               5.435e-06  3.942e-06   1.379   0.1689
nat.legal.liability       1.067e-02  6.099e-03   1.750   0.0811 .
sub.risk.disaster         1.142e+00  1.442e+00   0.792   0.4287
region.north.america      6.568e-01  3.638e-01   1.805   0.0719 .
region.asia               2.940e-01  3.991e-01   0.737   0.4618
region.other.americas     1.125e+00  8.425e-01   1.336   0.1825
region.africa            -4.869e-01  2.075e+00  -0.235   0.8146
region.other              4.075e-01  5.733e-01   0.711   0.4778
sector.construction       9.893e-01  1.080e+00   0.916   0.3605
sector.financial.services -9.582e-01  5.148e-01  -1.861   0.0636 .
sector.utilities          -4.663e-01  3.445e-01  -1.354   0.1768
sector.mining             -1.790e-01  4.440e-01  -0.403   0.6871
sector.public              6.834e-01  1.050e+00   0.651   0.5155
sector.wholesale.trade    -1.303e+00  1.070e+00  -1.218   0.2241
sector.information        -9.245e-01  5.340e-01  -1.731   0.0844 .
sector.transportation     -1.867e-01  5.560e-01  -0.336   0.7372
sector.real.estate        -2.001e-01  1.464e+00  -0.137   0.8913
sector.retail.trade       -3.109e+00  1.649e+00  -1.885   0.0603 .
sector.agriculture        -2.165e+00  1.051e+00  -2.059   0.0403 *
sector.arts               -1.143e+00  2.028e+00  -0.564   0.5734
sector.accommodation      -1.345e+00  1.453e+00  -0.926   0.3553
sector.health             -4.895e-01  2.027e+00  -0.242   0.8093
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

회귀분석 결과	변수 명	P-value
-2.165	`sector`, Agriculture, Forestry, Fishing and Hunting	*
0.01067	`nat.legal.liability`	.
0.656	`region`, North America	.
-0.989	`sector`, Financial Services	.
-0.9245	`sector`, Information	.
-3.109	`sector`, Retail Trade	.

6. Conclusion

- Improvement of this research.
  - Due to time limitations and personal capacity issues, I was not able to achieve all of the plans in the proposal announced at the beginning of the semester.
- Future works I have to do
  - GLM frequency analysis: Just as we did GLM regression analysis for severity, we need to conduct GLM frequency analysis to extract factors that affect frequency.
  - Annual loss: Using the distributions estimated previously, we need to estimate annual losses.
  - Based on the extracted factors and estimated annual losses, we have to establish a scenario hypothesis.