ANGHELA M. ALIZA
BSIT3B

Review of "Cyber Risk Taxonomies: Statistical Analysis of Cybersecurity Risk Classifications"

Type of statistics used (descriptive/inferential): The study primarily uses **inferential statistics**.

Justification for classification

This classification follows the study's methodology and reporting. The authors clearly state that their goal is to test how well different models can forecast the severity of cyber events across risk categories. To do this, they set up null and alternative hypotheses (for example, H0: E[S(G1,.)] = E[S(G2,.)] vs. H1: E[S(G1,.)] > E[S(G2,.)]) to compare model performance an approach that defines inferential statistics. They calculate test statistics (such as T_N_T+1 for rCRPS and rES) and compare them against critical values (1.64 for 5% significance, 2.32 for 1%) to decide whether to reject the null hypothesis. The study also runs a power analysis to check the reliability of these tests under different sample sizes, which is another key part of inferential methods. Descriptive statistics (mean, median, standard deviation, skewness, kurtosis) are included to summarize data by risk category, but these are used only as context. The study's conclusions about the effectiveness of classification are based on inferential analysis.

Analysis result

The study examines several types of cyber risk classifications: event-based (Advisen, Romanosky), operational risk-based (Eling), impact-based (Frequency & Severity, Type & Importance), and model fit-based (Tail, Body). The main aim is to measure how well models can forecast cyber events outside the sample, looking at both severity and frequency. To do this, the study uses a rolling window method and applies threshold-weighted scoring tools, namely the residual Continuously Ranked Probability Score (rCRPS) and the residual Energy Score (rES).

1. Cyber Event Severity Modeling

- **Weak Forecasting:** Risk classifications perform poorly in forecasting severity. Results are often no better than random or unclassified models.
- **Classification Results:**
  - *Type & Importance (impact-based, dynamic):* Showed some gains over no classification, but mainly at the 5% level. Performance against random classification was weaker, especially for extreme losses.
  - *Romanosky (event-based):* Delivered some significant improvements, especially in right-tail forecasts, but results were mixed.
  - *Eling (operational risk-based) & other static types:* Ineffective and may misrepresent severity.
  - *Frequency & Severity (dynamic):* Failed to improve severity forecasting.
- **Robustness Checks:** Removing extreme losses made forecasts worse, since the model used (POT-GAMLSS) already captured tail risks.
- **Takeaway:** Classifications should not be treated as main drivers of severity forecasting, though they remain useful for identifying threats and planning risk controls.

ANGHELA M. ALIZA
BSIT3B

2. Cyber Event Frequency Modeling

- **Clear Gains:** Unlike severity, frequency forecasts improve with classifications.
- **Statistical Evidence:** Chi-squared tests consistently showed better performance with risk categories across all years.
- **Insurance Impact:** Insurers may ignore classifications for severity but can use them in frequency models to refine premium pricing.

3. Power Analysis

- **Severity Limits:** Tests for severity have low statistical power, and building stronger evidence would require many more years of data.
- **Tail Weighting Helps:** Focusing on tail risks boosts statistical power even with smaller datasets.
- **Lognormal Results:** Gains in some tests came from model sensitivity to heavy tails, not real improvements from classifications. This further highlights their limits in severity forecasting.

Overall Conclusion

The study shows that cyber risk classifications help in modeling event frequency but add little value for severity forecasting. This distinction matters for both risk management and insurance, as classifications can guide premium adjustments based on frequency but should be used carefully or set aside when modeling severity. The findings also underline the weaknesses of static classifications in a changing cyber risk landscape and emphasize the importance of considering statistical power when assessing forecasting methods.

Link of the paper reviewed*: https://arxiv.org/html/2410.05297v1*