A simple statistical methodology for predicting cyber-attacks*

Rong Chen† Giuseppe Corvasce‡ Ruey S. Tsay§

Abstract

The paper proposes a simple statistical methodology for predicting and forecasting cyber-attacks. The procedure models the duration, defined as the number of days from one cyber-attack to the next ones, relying on the ARMA(1,1)/PARCH(1,1) specification, with the aim to also describe the evolution of the conditional volatility of the duration. The empirical analysis is based on disruptive, exploitive as well as mixed and undetermined cyber-attacks perpetrated by criminals and reports in sample and out-of-sample results of the statistical methodology, showing several metrics of performance and goodness of fit.

Keywords: Time-series models, cyber-attacks, out-of-sample results, metrics of performance

*The authors would like to thank Yakov Amihud, Viral Acharya, Frank Diebold, Stephen Brown, Matthew Richardson, Robert C. Merton, Richard Roll, Edward Amoroso, Giuseppe Arbia, Victor Ng, Charles Harry, Nancy Gallagher as well as several participants of the conferences and seminars organized by the GARP New York and Chicago Chapters, the University of Alberta, the Department of Finance at the University of Luxembourg, Harvard University, the University of Maryland, New York University, The Wharton School at the University of Pennsylvania, the Allied Social Science Association. The authors also thank the NYPD, the Federal Bureau of Investigation (FBI), the Interpol, Accademia Militare di Modena, Accademia Militare "Nunziatella", Accademia Militare Teuliè, Arma dei Carabinieri, Polizia di Stato, the FedPol in Switzerland, Ministero dell'Interno, Ministero della Difesa, Direzione Investigativa Antimafia, Direzione Distrettuale Antimafia, the association UNUCI, the Center for Financial Statistics and Risk Management at Rutgers University, its advisors and the staff members. The authors also thank several consultants, advisors, hedge/private equity and family funds as well as banks located in New York, New Jersey, Washington, Philadelphia, London, Hong Kong, France, Italy and Switzerland for comments received.

†Rutgers University - Department of Statistics, School of Arts and Sciences; Address: Hill Center, 110 Frellinghuysen Road, Piscataway, NJ 08854-8019, USA. Email: rongchen@stat.rutgers.edu

‡Rutgers University - Center for Financial Statistics and Risk Management; Address: Hill Center 110 Frelinghuysen Road Piscataway, NJ 08854.
University of Pennsylvania - Wharton School; Address: 3733 Spruce Street, Philadelphia, PA 19104.6340.
Harvard University - Massachusetts Hall, Cambridge, MA 02138 - USA
Society for Financial Studies; Mailing Address: PO Box 28494 Raleigh, NC 27611.
Email: giuseppecorvasce@gmail.com.

§University of Chicago - Booth School of Business - Econometrics and Statistics. Address: 5807 S. Woodlawn Avenue, Chicago, IL 60637, U.S.A. Email: ruey.tsay@chicagobooth.edu

1. Introduction

Cyber-attacks are a significant threat in the digital age and necessitates the development of reliable methods able to predict and forecast them. They can be defined as intentional efforts to steal, expose, alter, disable, or destroy data, applications, or other assets through unauthorized access to a network, computer system or digital device. By one estimate, cybercrime will cost the world economy USD 10.5 trillion, by the end of the year 2025.

Cyber-attacks can disrupt, damage, destroy businesses and can be considerably more costly than others. Some of the most common types of cyber-attacks include: malware, social engineering, denial-of-service attacks, account compromise, man-in-the-middle attacks and supply chain attacks. In particular, ransomware attacks are a subcategory of malware attacks, evolved in the last few years. They have the aim to limit a computer system's access or encrypt files in exchange for payment, generating data loss, financial and business risks, and damage to reputation.

This paper proposes a simple statistical methodology for predicting cyber events. The statistical approach models the duration, defined as the number of days from one cyber event to the next ones, relying on the AutoRegressive Moving Average model (ARMA), with one own lagged variable and one moving average term as well as on the Power Autoregressive Conditional Heteroskedastic model (PARCH), introduced by Ding et al. (1993), for describing the evolution of the conditional variance of the duration, as a metric of uncertainty.

The empirical analysis relies on the series of the durations related to cyber events, classified and collected by the Center for International & Security Studies at the University of Maryland (Harry and Gallagher 2018) and discusses the summary and descriptive statistics, the estimation results of the statistical methodology, the metrics for evaluating the goodness of fit of the model as well as the forecasting results, based on 50% of the durations for the cyber events.

The selected statistical models for the durations related to cyber events, based on Akaike, Schwartz and Hannan-Quinn information criterions, are respectively the following: ARMA(1,1)

with PARCH(1,1) variance and normal errors, for DISRUPTIVE cyber-attacks; ARMA(1,1) with PARCH(1,1) variance and generalized errors distribution, for EXPLOITIVE cyber-attacks; ARMA(1,1) with PARCH(1,1) variance and normal errors, for MIXED and UNDETERMINED cyber-attacks.

The empirical results also report the percentages of correctly forecasted cyber-attacks, based on the comparisons of the realized durations and the rounded one-step ahead forecasts (static forecasts) of the durations. The out-of-sample analysis is performed on 50% of the observations. In particular, rounding the static forecasts of the durations approximate them to the closest values. The standard rounding procedure for the number of the forecasted days involves reducing the number of days by significant decimal digits, while retaining the general magnitude or size of the original number[1].

The percentage of correctly forecasted DISRUPTIVE cyber-attacks is equal to 18.83%; whereas, the percentage respectively increases to 55.15% and 33.03% for EXPLOITIVE as well as MIXED and UNDETERMINED cyber-attacks.

The rest of the paper is organized as follows: Section 2 discusses the data and provides some summary statistics. Section 3 reviews the statistical methodology. In Section 4 are reported the empirical results; whereas, Section 5 concludes the paper.

2. Data and Descriptive Statistics

Following the taxonomy provided by Harry and Gallagher (2018), that also considers the studies proposed by Howard and Longstaff (1998) and Kjaerland (2005), cyber events are classified by types. In particular, the database focuses on n. 1189 disruptive, n. 3490 exploitive as well as n. 1284 mixed and undetermined cyber events perpetrated by criminals, although the motivations behind cyber-attacks can also be politicals and personals. The

[1]The result of the rounding is less accurate than the original number, but simplifies the calculations and provides an approximative value of the number of days.

data are from the CISSM Cyber Attacks Database (https://cissm.umd.edu/cyber-events-database), collected by the University of Maryland and for the period from January 1st, 2014 to February 15th, 2024[2].

Disruptive cyber events are cyber-attacks that cause a significant disruption to the normal operations of a system, service, or network. These events can impact an organization's ability to produce, deliver, or communicate, leading to downtime, reduced productivity, and financial losses. Disruptive effects can be classified into five sub-categories: Message Manipulation, External Denial of Service, Internal Denial of Service, Data Attack, and Physical Attack.

Exploitive cyber events are cyber-attacks designed to steal information rather than to disrupt operations. Hackers may be seeking customer data, intellectual property, classified national security information, or sensitive details about the organization itself. While the tactics or techniques used by malicious actors may change regularly, the location from which they get that information does not.

Mixed and undetermined cyber events are also cyber-attacks, where criminal actors also use techniques such as double extortion, which involves data exfiltration (exploitative) and encryption (disruptive).

Criminally motivated attackers seek financial gain through monetary theft, data theft, or business disruption. Cybercriminals may hack into a bank account to steal money directly or use social engineering scams to trick people into sending money to them. Hackers may steal data and use it to commit identity theft or sell it on the dark web or hold it for ransom. They may mainly use ransomware or other tactics to hold data or devices hostage until a company pays. However, according to the most recent statistics, 32 percent of cyber incidents involved the theft and sale of data rather than encryption for extortion.

[Please Insert Table 1 around here]

[2]The taxonomy proposed by Harry and Gallagher (2018) is based on reported cyber events and does not consider unreported attacks, although these events might happen more frequently.

Table 1 reports the summary and the descriptive statistics of the durations, across the analyzed event types. On average, the number of days from one cyber disruptive event to the next is equal to 3.1068; whereas, the standard deviation is equal to 5.2472. The exploitive as well as the mixed and undetermined cyber events are more frequent than the disruptive events. On average, the exploitive cyber events occur almost every day and the mixed and undetermined cyber events result every three day. The timing uncertainty, measured by the standard deviation decreases to 1.7106 for the exploitive cyber events and reaches the value of 15.6771, for mixed and undetermined cyber events.

[Please Insert Figure 1 around here]

Figure 1 shows the evolution of the durations for the analyzed event types and for the investigated period. The maximum number of days, from one cyber event to the next one, is equal to 73 for disruptive cyber-attacks. This level decreases to 43 for exploitive cyber-attacks and increases to 401 for mixed and undetermined cyber events. Whereas, the minimum number of days is equal to 0, across event types.

3. The Model

This section proposes the statistical approach for modeling the dynamic of the duration (D), defined as the number of days (N) from one cyber event to the next ones. The model relies on the AutoRegressive Moving Average model (ARMA), with one own lagged variable and one moving average term as well as on the Power Autoregressive Conditional Heteroskedastic model (PARCH), introduced by Ding et al. (1993), for describing the evolution of the conditional variance of the duration, as a metric of uncertainty.

The series of the duration D, related to the cyber-attack (τ), if it is stationary, can be defined in the following way:

$$D_\tau = N_\tau - N_{\tau-1} \tag{1}$$

$$D_\tau = a\,(1) + a\,(2) \cdot D_{\tau-1} + \epsilon_\tau + a\,(3) \cdot \epsilon_{\tau-1} \tag{2}$$

where, $a\,(1)$ is the constant of the mean equation; whereas, $a\,(2)$ and $a\,(3)$ respectively depict the autoregressive and the moving average component of the mean equation. The terms $\epsilon_1, ..., \epsilon_\tau$ are i.i.d. white noise error terms, that are commonly normal random variables. Further, the conditional variance of D, related to the cyber-attack τ can be written as follows:

$$\left(\sqrt{\sigma_\tau^2}\right)^{a(8)} = a\,(4) + a\,(5) \cdot \left(abs\left(\epsilon_{\tau-1}\right) - a\,(6) \cdot \epsilon_{\tau-1}\right)^{a(8)} + a\,(7)\left(\sqrt{\sigma_{\tau-1}^2}\right)^{a(8)}, \tag{3}$$

where, $a\,(4)$ is the constant term of the deterministic variance process; $a\,(5)$ is the coefficient for the ARCH component; $a\,(6)$ is the asymmetry coefficient for the deterministic variance process; $a\,(7)$ is the coefficient of the GARCH component; $a\,(8)$ is the coefficient for the power transformation term inclusive of any positive value.

The statistical approach incorporates the evolution of the conditional variance and relies on the standard class of ARCH model (Engle 1982), which specifies the use of a squared term in that the conditional variance is specified as a function of past squared residuals and variances. It captures the tendency of the duration for the cyber events volatility to cluster, meaning periods of high volatility tend to be followed by more high volatility, and periods of low volatility tend to be followed by more low volatility. The PARCH model also incorporates the Taylor (1986) and Bollerslev (1986) GARCH model, which relates the conditional standard deviation as a function of past lagged absolute residuals and standard deviations.

4. Empirical Results

The estimation results rely on the Berndt-Hall-Hall-Hausman (BHHH) algorithm (Berndt et al. 1974) that is a numerical optimization algorithm similar to the Newton-Raphson algorithm, but it replaces the observed negative Hessian matrix with the outer product of the gradient. The optimization algorithm begins at an initial estimate for the optimal values and proceeds iteratively to get better estimates at each stage, till when there is a convergence for finding the solutions. For simplicity, the maximum number of iterations is fixed to n. 500 and the convergence rate to 1e-06. The step method is based on the Levenberg-Marquardt algorithm (Levenberg 1944; Marquardt 1963) that is more robust than the Gauss-Newton algorithm, since it allows to derive solutions even if the algorithm starts very far o from the final minimum. In cases with multiple minima, the algorithm converges to the global minimum only if the initial guess is already somewhat close to the final solution.

The statistical models for the durations related to cyber events are respectively the following: ARMA(1,1) with PARCH(1,1) variance and normal errors, for DISRUPTIVE cyber-attacks; ARMA(1,1) with PARCH(1,1) variance and generalized errors distribution, for EXPLOITIVE cyber-attacks; ARMA(1,1) with PARCH(1,1) variance and normal errors, for MIXED and UNDETERMINED cyber-attacks.

[Please Insert Table 2 around here]

Table 2 reports the estimation results for the statistical models. The coefficients of the approaches for the durations of disruptive, exploitive as well as mixed and undetermined cyber-attacks are all statistically significant at 1% level. The signs of the moving average components related to the mean equations are negatives. Regarding the conditional variance processes, the coefficient for the ARCH component is equal to 0.095 for the duration of disruptive cyber-attacks. It increases to 0.562 for exploitive cyber events and decreases to 0.050 for the duration of mixed and undetermined cyber-attacks.

The asymmetric response of volatility to positive and negative dramatic variations of the duration is also evident for disruptive as well as mixed and undetermined cyber-attacks. The coefficients that depict the asymmetry effect of the models respectively reach a level of -0.832 for the duration of disruptive cyber-attacks and -0.942 for the duration of mixed and undetermined cyber-attacks. This level is positive and equal to 0.629 for the duration of exploitive cyber-attacks. Further, the coefficients that depict the persistence of the variance component for disruptive as well as mixed and undetermined cyber-attacks are respectively equal to 0.926 and 0.939.

[Please Insert Table 3 around here]

The results related to the goodness of fit for the models are reported in Table 3. In particular, the proportion of the variance in the dependent variable that is explained by the independent variables, that also accounts for the number of predictors, is equal to 0.153 for the duration of disruptive cyber-attacks. The adjusted R-squared reaches a level of 0.259 for the duration of mixed and undetermined cyber events and decreases to 0.020 for exploitive cyber attacks.

Further, the metrics for the selection of the models, within the approaches characterized by GARCH effects, respectively reveal how ARMA(1,1) with PARCH(1,1) variance and normal errors, ARMA(1,1) with PARCH(1,1) variance and generalized errors distribution as well as ARMA(1,1) with PARCH(1,1) variance and normal errors are the best models for the durations of disruptive, exploitive and mixed and undetermined cyber events. In particular, the Schwartz Information Criterions are respectively equal to 5.147, 3.224 and 3.885.

[Please Insert Figure 2 around here]

Figure 2 respectively shows the evolution of the conditional volatility for the durations.

For disruptive cyber events, the mean of the conditional standard deviation is equal to 3.873; it respectively reaches a maximum level of 17.89 and a minimum level of 0.73, with a skewness that is equal to 1.38. The chart also shows the evolution of the conditional volatility for the duration of exploitive as well as mixed and undetermined cyber-attacks. In particular, the mean and the skewness of the conditional standard deviations are respectively equal to 1.50 and 4.81 for exploitive cyber-attacks; whereas, these values are respectively equal to 2.92 and 7.01 for mixed and undetermined cyber-attacks.

4.1 Out-of-sample results

This subsection reports the results of the out-of-sample forecasts of the model, with the aim to evaluate the performance on the test data. Out-of-sample forecasting involves testing a model's predictive accuracy on data it hasn't been trained on, providing a more realistic assessment of its generalizability. For the purpose of the analysis, the durations of cyber attacks are splitted into a training set, based on 50% of the durations, used to develop the model and a separate test set, based on the remaining observations, used for evaluation. Out-of-sample forecasts are considered more trustworthy than in-sample forecasts because they avoid overfitting to the training data and better reflect real-time information.

[Please Insert Table 4 around here]

Table 4 contains the metrics of performance (Root Mean Square Error, Mean Absolue Error, Theil Inequality Coefficient, Bias Proportion and Variance Proportion), based on the test sample of observations. The table also reports the percentage of correctly forecasted cyber attacks, based on the comparison of realized durations and the rounded one-step ahead forecasts (static forecasts) of the durations. In particular, the Theil Inequality Coefficient, that quantifies the discrepancy between the actual values and the forecast values, is respectively equal to 0.462 for the duration of disruptive cyber-attacks, 0.668 for the duration of

exploitive cyber-attacks and 0.468 for the duration of mixed and undetermined cyber-attacks. The percentage of correctly forecasted durations is respectively equal to 18.83% for disruptive events type, 55.15% for exploitive events type and 33.02% for mixed and undetermined events type.

5. Conclusion

Cyber-attacks are exponentially increasing, enhancing the necessity to design more relevant prediction models and approaches due to several limitations and assumptions of mathematical modeling techniques such as Bayesian Joint Probability, Monte Carlo Simulations and Probability Distributions to provide alternative methods for cyber-attack forecasting.

This paper is a pioneering contribution in the cybersecurity literature that proposes a simple statistical methodology for predicting and forecasting cyber events. The statistical approach models the duration, defined as the number of days from one cyber event to the next ones, relying on the AutoRegressive Moving Average model (ARMA), with one own lagged variable and one moving average term as well as on the Power Autoregressive Conditional Heteroskedastic model (PARCH), introduced by Ding et al. (1993), for describing the evolution of the conditional variance of the duration, as a metric of uncertainty. While Deep learning and machine learning (ML) algorithms have been widely used to predict cyber threats, the role of time series methodologies able to predict and forecast cyber-attacks is becoming prominent.

Some References

Berndt, E., Hall, B., Hall, R., and Hausman, J., 1974. Estimation and Inference in Nonlinear Structural Models. *Annals of Economic and Social Measurement*, 3(4), 653-665.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307-327.

Ding Z., Granger, C.W.J. and Engle, R.F., 1993. A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1, 83-106.

Engle, R.F., 1982. Autoregressive Conditional Heteroskedasticity with Estimates of Variance of United Kingdom Inflation, *Econometrica*, 50, 987-1008.

Harry, C., and Gallagher, N., 2018. Classifying Cyber Events. *Journal of Information Warfare*, 17(3), 17-31.

Howard, J. and Longstaff, T., 1998. A Common Language for Computer Security Incidents, *Technical Report*, Sandia National Laboratories.

Kjaerland, M., 2005. A taxonomy and comparison of computer security incidents from the commercial and government sectors. *Computers and Security*, 25(7), 522–538.

Levenberg, K., 1944. A Method for the Solution of Certain Non-Linear Problems in Least Squares. *Quarterly of Applied Mathematics*, 2(2), 164-168.

Marquardt, D., 1963. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *SIAM Journal on Applied Mathematics*, 11(2), 431-441.

Schwert, W., 1990, Stock volatility and the crash of '87. *Review of Financial Studies*, 3(1), 77-102.

Smith, J. M., & Vasquez, J., 2014. Monte Carlo Methods for Cybersecurity Risk Analysis. *International Journal of Computer Science & Information Security*, 12(7), 20-29.

Shevchenko, P. V., Jang, J., Malavasi, M., Peters, G. W., Sofronov, G., and Truck, S., 2023. The Nature of Losses from Cyber-Related Events: Risk Categories and Business Sectors. *Journal of Cybersecurity*, 9(1).

Taylor, S.J., 1986. Modeling Financial Time Series. *John Wiley & Sons*, New York.

Table 1.
Summary Statistics

The table contains the summary statistics (mean, median, max., min. std. dev, skewness and kurtosis) for the so called **DURATION**, the number of days from one cyber attack to the next ones. The table respectively contains the descriptive statistics for n. 1189 DISRUPTIVE cyber attack events, n. 3490 EXPLOITIVE cyber attack events as well as n. 1284 MIXED and UNDETERMINED events. The data are from the **CISSM Cyber Attacks Database** (https://cissm.umd.edu/cyber-events-database), collected by the University of Maryland and for the period from **January 1st, 2014** to **February 15th, 2024**.

Descriptive Statistics	**CYBER ATTACK**		
	DISRUPTIVE	**EXPLOITIVE**	**MIXED and UNDETERMINED**
Mean	3.1068	1.0590	2.8108
Median	1.0000	1.0000	1.0000
Max.	73.0000	43.0000	401.0000
Min.	0.0000	0.0000	0.0000
Std. dev.	5.2472	1.7106	15.6771
Skewness	4.6914	6.4103	17.0356
Kurtosis	40.1965	115.6759	371.3736

Table 2.
Estimation Results

The table contains the estimation results of the econometric procedures that depict the evolutions of the called **DURATIONS**, the number of days from one cyber attack to the next ones for **DISRUPTIVE** (Panel 2.1), **EXPLOITIVE** (Panel 2.2) as well as **MIXED** and **UNDETERMINED** (Panel 2.3) event types. **Panel 2.1** reports the estimation results for the ARMA(1,1) with PARCH(1,1) volatility and Normal Errors. a(1) is the constant of the mean equation; whereas, a(2) and a(3) respectively depict the autoregressive and the moving average component of the mean equation; a(4) is the constant term of the deterministic variance process; a(5) is the coefficient for the ARCH component; a(6) is the asymmetry coefficient for the deterministic variance process; a(7) is the coefficient of the GARCH component; a(8) is the coefficient for the power term. **Panel 2.2** reports the estimation results for the ARMA(1,1) with PARCH(1,1) volatility and generalized errors distribution (GED). b(1) is the constant of the mean equation; whereas, b(2) and b(3) respectively depict the autoregressive and the moving average component of the mean equation; b(4) is the constant term of the deterministic variance process; b(5) is the coefficient for the ARCH component; b(6) is the asymmetry coefficient for the deterministic variance process; b(7) is the coefficient for the power term; b(8) is the parameter that depicts the generalized errors distribution. **Panel 2.3** reports the estimation results for the ARMA(1,1) with PARCH(1,1) volatility and Normal Errors. c(1) is the constant of the mean equation; whereas, c(2) and c(3) respectively depict the autoregressive and the moving average component of the mean equation; c(4) is the constant term of the deterministic variance process; c(5) is the coefficient for the ARCH component; c(6) is the asymmetry coefficient for the deterministic variance process; c(7) is the coefficient of the GARCH component; c(8) is the coefficient for the power term. The table also reports the estimated values (Est. Values), z-statistics (z-Stat.) and the statistical significance (p-values).

Panel 2.1: **DISRUPTIVE** cyber attacks

Coefficients	**ARMA(1,1)/PARCH(1,1)**		
	Est. Values	***z*-Stat.**	**p-values**
a(1)	1.929	30.304	0.000
a(2)	0.980	322.435	0.000
a(3)	-0.867	-80.840	0.000
a(4)	0.015	6.608	0.000
a(5)	0.095	17.337	0.000
a(6)	-0.832	-24.792	0.000
a(7)	0.926	213.333	0.000
a(8)	0.305	6.173	0.000

Panel 2.2: **EXPLOITIVE** cyber attacks

Coefficients	ARMA(1,1)/PARCH(1,1) with GED		
	Est. Values	**z-Stat.**	**p-values**
b(1)	23.584	20.462	0.000
b(2)	1.000	59528.65	0.000
b(3)	-0.980	-935.39	0.000
b(4)	1.234	10.622	0.000
b(5)	0.562	4.283	0.000
b(6)	0.629	12.817	0.000
b(7)	1.228	3.886	0.000
b(8)	0.712	41.958	0.000

Panel 2.3: **MIXED and UNDETERMINED** cyber attacks

Coefficients	ARMA(1,1)/PARCH(1,1)		
	Est. Values	**z-Stat.**	**p-values**
c(1)	1.208	7.785	0.000
c(2)	0.956	96.326	0.000
c(3)	-0.873	-58.583	0.000
c(4)	0.041	9.865	0.000
c(5)	0.050	7.560	0.000
c(6)	-0.942	-9.127	0.000
c(7)	0.939	134.206	0.000
c(8)	0.643	7.181	0.000

Table 3.
Goodness of fit of the models

The table reports the metrics of goodness of fit (R-squared, Adjusted R-squared, Akaike Information Criterion, Schwartz Criterion, Hannan-Quinn Criterion) for the models that depict the evolution of the duration, defined as the number of days from one cyber attack to the next ones for **DISRUPTIVE**, **EXPLOITIVE** as well as **MIXED** and **UNDETERMINED** event types**.** The selected econometric models for the durations are respectively the following: ARMA(1,1) with PARCH(1,1) volatility and Normal errors, for DISRUPTIVE cyber attacks; ARMA(1,1) with PARCH(1,1) volatility and GED, for EXPLOITIVE cyber attacks; ARMA(1,1) with PARCH(1,1) volatility and Normal Errors, for MIXED and UNDETERMINED cyber attacks.

Metrics of goodness of fit	**CYBER ATTACK**		
	DISRUPTIVE	**EXPLOITIVE**	**MIXED and UNDETERMINED**
R-squared	0.154	0.020	0.260
Adj. R-squared	0.153	0.020	0.259
Akaike Information Criterion	5.113	3.210	3.852
Schwartz Information Criterion	5.147	3.224	3.885
Hannan-Quinn Criterion	5.126	3.215	3.865

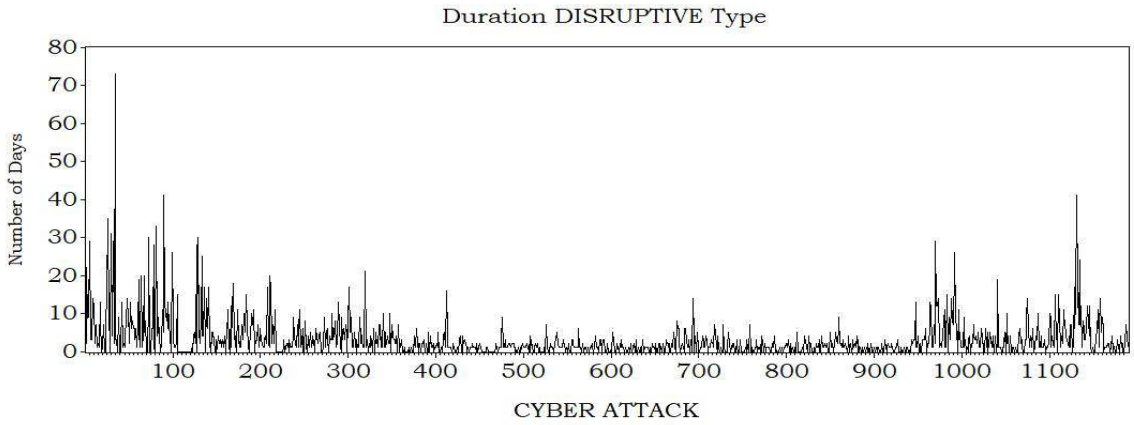
Table 4.
Forecasts of the Durations for the cyber attacks

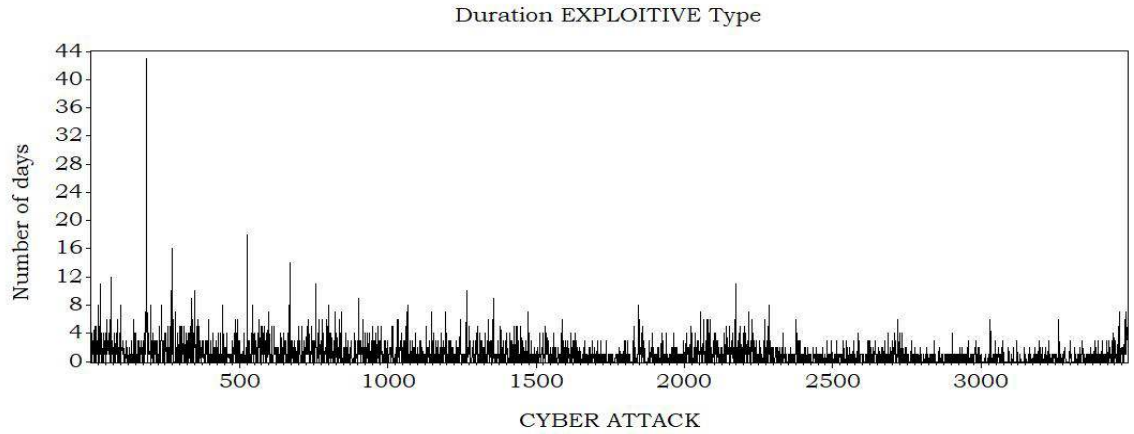
The table contains the out-of-sample results based on **50%** of the observations for **DISRUPTIVE**, **EXPLOITIVE** as well as **MIXED** and **UNDETERMINED** cyber attack event types. It shows the Root Mean Square Error (RMSE), the Mean Absolute Error (MAE), the Theil Inequality Coefficient, the Bias Proportion and the Variance Proportion. The table also reports the percentage of correctly classified cyber attacks, based on the comparison of realized durations and the rounded one-step ahead forecasts (static forecasts) of the durations. The data are from the **CISSM Cyber Attacks Database**, collected by the University of Maryland and for the period from **January 1st, 2014 to February 15th, 2024**.

Metrics of Performance	**CYBER ATTACK**		
	DISRUPTIVE	**EXPLOITIVE**	**MIXED and UNDETERMINED**
RMSE	3.551	1.199	1.207
MAE	2.095	0.707	0.864
Theil Inequality Coeff.	0.462	0.668	0.468
Bias Proportion	0.002	0.130	0.007
Variance Proportion	0.382	0.472	0.579
% Correctly Forecasted	**18.83%**	**55.15%**	**33.02%**

Figure 1.
Durations

The figure shows the durations, defined as the number of days from one cyber attack to the next ones for DISRUPTIVE, EXPLOITIVE as well as MIXED and UNDETERMINED event types. The data are from the **CISSM Cyber Attacks Database** (https://cissm.umd.edu/cyber-events-database), collected by the University of Maryland and for the period from **January 1st, 2014** to **February 15th, 2024**.





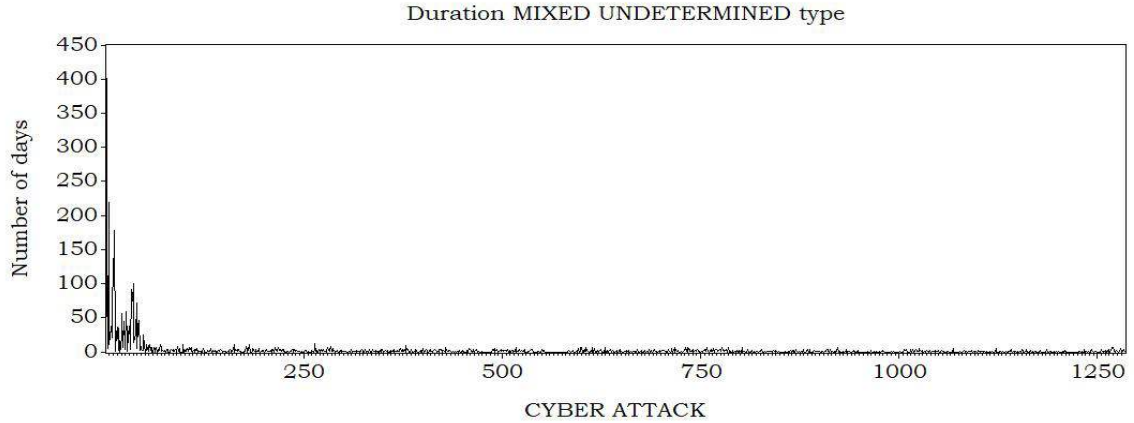


Figure 2.
Conditional Volatility of the Durations

The figure shows the evolution of the conditional volatility for the duration of DISRUPTIVE, EXPLOITIVE as well as MIXED and UNDETERMINED cyber attacks. The left axis reports the duration, defined as the level of the conditional standard deviation; whereas, the bottom axis reports the cyber attacks. The data are from the **CISSM Cyber Attacks Database** (https://cissm.umd.edu/cyber-events-database), collected by the University of Maryland and for the period from **January 1st, 2014** to **February 15th, 2024**.