Ordered response models for cyber risk

Modelli a risposta ordinale per la valutazione del cyber risk

Silvia Facchinetti and Claudia Tarantola

Abstract In the last years there have been a scholars increasing interest in cybersecurity risk measurement, data security, and privacy protection. Since quantitative loss data are rarely available, we deal with ordinal data representing experts' evaluation of the severity of the attacks. Due to the ordinal nature of the available data, it turns natural to rely on cumulative link models that allows us to express the cumulative probabilities associated with the different severity levels as a non linear function of a suitable set of explanatory variables. We evaluate the effect of each explanatory categorical variable on the risk level using the *Average Marginal Effect*. We apply our model to a real data set that includes information on serious cyber attacks occurred worldwide in 2018.

Abstract Negli ultimi anni si è registrato un interesse crescente da parte degli studiosi riguardo il problema del cyber risk e la sua misurazione. Poichè i dati quantitativi sulle perdite sono raramente disponibili, essi sono spesso rilevati su scala ordinale e riguardano il livello di gravità degli attacchi cibernetici. Risulta pertanto naturale valutare il cyber risk mediante modelli di risposta ordinale che legano la variabile gravità a variabili esplicative spesso di natura categorica. L'effetto di tali variabili viene valutato utilizzando l'AME (*Average Marginal Effect*). Il modello viene applicato a dati reali sulla gravità degli attacchi rilevati nel mondo nel 2018.

Key words: ordered response models, cyber risk, ordinal variables, Average Marginal Effect

Silvia Facchinetti
Department of Statistical sciences, Università Cattolica del Sacro Cuore, Largo Gemelli 1 - Milano, e-mail: silvia.facchinetti@unicatt.it

Claudia Tarantola
Fintech laboratory, Department of Economics and Management, University of Pavia, Strada Nuova, 65 - Pavia e-mail: claudia.tarantola@unipv.it

1 Motivation

Cyber risk is an operational risk caused by attacks on Information and Communication Technologies systems (ICT), which are gaining increasing importance, due to the globalisation of financial activities as well as to the technological advancements. (see e.g. [4], [5], [9]). Institutions should be encourage to collect data on cyber incidents in order to use statistical approaches to estimating the capital needed to cover losses due to the occurrence of cyber attacks. Companies need to identify the key vulnerable assets that are exposed to cyber risk and to implement an integrated cyber security solution and optimal investment decisions that would automate and accelerate the threat defense [8].

The main problem in cyber risk measurement is the lack of available data (see [2]). Indeed, cyber loss data are very difficult to obtain since these data are very sensitive, and it is uncommon for an institution to be willing to disclose them, since it wants to preserve its reputation and security. This lack of disclosure drastically limits the data available for the analysis of cyber risks, by regulators, research centers and by third party organisations, such as insurance companies, which may offer hedging against cyber risks.

In this context, it is quite natural to employ an ordinal measurement approach, rather than a quantitative one. Cyber risk data are often collected on an ordinal scale, in terms of severity of the attacks ordered according to the corresponding magnitude, for example critical, high and medium severity (see [6]). Due to the ordinal nature of the available data, we relay on cumulative link models that allows us to express the cumulative probabilities associated with the different severity levels as a non linear function of a suitable set of explanatory variables.

We describe the methodology and we apply our model to a real data set concerning serious cyber attacks that occurred worldwide in 2018.

2 Methodology

Let Y be the categorical variable severity with ordered levels $j = 1, 2, \ldots, k$. The cumulative link model is:

$$\text{link}[P(Y \leq j)] = \alpha_j - \mathbf{x}^T \beta, \tag{1}$$

where $\mathbf{x}$ is a p-vector of regression variables, α_j is the cut-point, β is the vector of the regression parameters and link is a suitable link function (see [1] for details).

The larger the value of $\mathbf{x}^T \beta$, the higher the probability that the response will fall in a category at the upper end of the scale. Since nonlinear link function naturally produces effects on the link scale that may be not easy to interpret, we evaluate the effect of each explanatory variable on the risk level using the so-called Marginal Effect (ME) measure, see e.g. [7]. The ME measures how a change in a specific covariate affects the response variable, holding constant the value of all the other

covariates. This measure is particularly useful because it is intuitive (the slope of the regression surface with respect to a given covariate) and can be calculated from essentially any type of covariates. The ME measure is computed differently for categorical and continuous explanatory variables. For categorical variables, they measure the discrete change. In particular, for binary data they correspond to the discrete change, and indicate how predicted probabilities change as the binary independent variable goes to from 0 to 1, holding all the other variables constant, while for factor levels they indicate how the probability of being in a category changes when we move from the baseline category to the considered one. For continuous variables, they measure the instantaneous rate of change.

Among the alternative versions of the ME (see e.g. [10]), we focus on the Average Marginal Effect (AME) that is obtained by calculating the marginal effect of a specific covariate at each of the n samples of the explanatory variables, and then averaging them.

3 Empirical analysis

We consider real data collected by the researcher of the Hackmanac Project and described in the "Italian Annual report on ICT Security in Italy 2019" [3], by the Clusit association (Associazione Italiana per la Sicurezza Informatica).

The complete data set consists of 8,417 worldwide observations on serious cyber attacks in the years 2011-2018. We consider for our analysis 1,552 cyber attacks that occurred in 2018, the most recent year. According to Clusit, an attack is considered serious if it had a significant impact on the victims in terms of economic losses, damages to reputation and/or dissemination of sensitive data. We consider a model that relates the severity, a three-category ordinal response variable that measures the severity of an attack (1 = critical, 2 = high, 3 = medium severity), to the following categorical explanatory variables:

- *Attack Technique* coded in 5 categories: 0-day, Multiple Threats, Trivial Threats, SQL Injection, Unknown;
- *Continent* coded in 6 categories: Africa, America, Asia, Australia/Oceania, Europe, Multiple Continents;
- *Type of Attack* coded in 4 categories: Cybercrime, Hactivism, Espionage/Sabotage, Information Warfare;
- *Victim* coded in 19 categories: Automotive, Chemical/Medical, Banking/Finance, Critical Infrastructures, Entertainment/News, GDO/Retail, Gov-Mil-LEAs - Intelligence, Gov.Contractors/Consulting, Health, Hospitability, Multiple Targets, Online Services/Cloud, Organization/ONG, Others, Religion, Research- Education, Security Industry, Software/Hardware Vendor, Telco.

To evaluate and interpret the effects of an explanatory variable on the response probability, we now calculate the Average Marginal Effect. Since only the highest and lowest probabilities change monotonically as the explanatory variable increases,

the most extreme outcomes are often of special interest (in our context critical vs medium severity). For this reason, in Table 1 we present the results only for extreme categories.

Fitting a cumulative logit link model, we noticed that not all the categories of the considered explanatory variables present a significant relation with severity, at level $\alpha = 0.05$. Moreover, there is an inverted effect for the medium severity attacks with respect to those with critical severity. For categorical variables with more than two possible values (as the variables in our data set), the MEs show the difference in the estimated probabilities for cases in a given category relative to the baseline one.

First consider the variable *Attack Technique*, with baseline category SQL Injection. All the categories present a significant relation with the severity. In particular:

- Multiple Threats are, on average, 0.999 more likely to generate a critical attack and -0.999 less likely to generate a medium severity attack than SQL Injection;
- Trivial Threats are, on average, 1.420 more likely to generate a critical attack and -2.446 less likely to generate a medium severity attack than SQL Injection;
- Unknown attack techniques are, on average, 1.423 more likely to generate a critical attack and -2.452 less likely to generate a medium severity attack than SQL Injection;
- 0-day is, on average, 3.062 more likely to generate a critical attack and -5.275 less likely to generate a medium severity attack than SQL Injection.

Next, we consider the variable *Continent*, with baseline category Multiple Continents. Only the category Asia show a significant relation with the response variable. This continent is, on average, 0.105 more likely to generate a critical attack than Multiple Continents. For the medium severity attacks we observe an inverted effect: Asia is, on average, -0.181 less likely to generate a medium severity attack than the reference category.

With reference to the variable *Type of Attack*, the categories Espionage/Sabotage and Information warfare present a significant relation with the severity, with respect to the baseline category Cybercrime:

- Espionage/Sabotage is, on average, 0.514 more likely to generate a critical attack and -0.403 less likely to generate a medium severity attack than Cybercrime;
- Information warfare is, on average, 0.310 more likely to generate a critical attack and -0.534 less likely to generate a medium severity attack than Cybercrime.

Finally, consider the variable *Victim*, with baseline category Automotive, the significant categories are 7: Chemical/Medical, Critical Infrastructures, Gov-Mil-LEAs-Intelligence, Gov. Contractors/Consulting, Health, Organization/ONG, Telco. In particular:

- Chemical/Medical is, on average, 1.504 more likely to generate a critical attack and -2.591 less likely to generate a medium severity attack than Automotive;
- Critical Infrastructures are, on average, 0.440 more likely to generate a critical attack and -0.757 less likely to generate a medium severity attack than Automotive;

Table 1 AME for the cumulative logit model fitted to the cyber risk data.

$ME.1 (critical severity)		effect	std.error	z.value	p.value
Attack technique	Multiple Threats	0.999	0.000	15589.237	0.000
	Trivial Threats	1.420	0.069	20.654	0.000
	Unknown	1.423	0.069	20.524	0.000
	0-day	3.062	0.135	22.672	0.000
Continent	America	0.036	0.055	0.655	0.512
	Asia	0.105	0.021	4.920	0.000
	Europe	0.028	0.020	1.401	0.161
	Australia/Oceania	0.057	0.041	1.397	0.162
	Africa	0.036	0.052	0.692	0.489
Type of attack	Espionage/Sabot.	0.514	0.073	7.073	0.000
	Hacktivism	-0.035	0.025	-1.404	0.160
	Inf. Warfare	0.310	0.038	8.096	0.000
Victim	Banking/Finance	0.210	0.110	1.913	0.056
	Chemical/Medical	1.504	0.066	22.672	0.000
	Critical Infr.	0.440	0.072	6.141	0.000
	Entert./News	0.032	0.068	0.475	0.635
	GDO/Retail	0.044	0.073	0.594	0.553
	Gov-Mil-LE-Int.	0.311	0.066	4.678	0.000
	Gov. Contr./Cons.	0.403	0.103	3.903	0.000
	Health	0.270	0.067	4.036	0.000
	Hospitability	0.029	0.073	0.400	0.689
	Multiple targets	0.058	0.066	0.883	0.377
	Online Serv/Cloud	0.085	0.067	1.256	0.209
	Organization-ONG	0.165	0.080	2.051	0.040
	Others	0.065	0.075	0.876	0.381
	Religion	-0.056	0.124	-0.456	0.649
	Research-Educ.	0.054	0.068	0.799	0.424
	Security	0.182	0.118	1.540	0.124
	SW/HW Vendor	0.092	0.068	1.352	0.176
	Telco	0.237	0.093	2.546	0.011

$ME.3 (medium severity)		effect	std.error	z.value	p.value
Attack technique	Multiple Threats	-0.999	0.000	-99521.457	0.000
	Trivial Threats	-2.446	0.071	-34.415	0.000
	Unknown	-2.452	0.072	-33.920	0.000
	0-day	-5.275	0.123	-42.942	0.000
Continent	America	-0.063	0.092	-0.687	0.492
	Asia	-0.181	0.036	-5.089	0.000
	Europe	-0.048	0.034	-1.409	0.159
	Australia/Oceania	-0.098	0.070	-1.400	0.162
	Africa	-0.062	0.089	-0.693	0.488
Type of attack	Espionage/Sabot.	-0.403	0.020	-20.164	0.000
	Hacktivism	0.061	0.043	1.405	0.160
	Inf. Warfare	-0.534	0.068	-7.842	0.000
Victim	Banking/Finance	-0.468	0.145	-3.221	0.001
	Chemical/Medical	-2.591	0.060	-42.942	0.000
	Critical Infr.	-0.757	0.126	-5.995	0.000
	Entert./News	-0.056	0.117	-0.475	0.635
	GDO/Retail	-0.075	0.126	-0.593	0.553
	Gov-Mil-LE-Int.	-0.536	0.114	-4.703	0.000
	Gov. Contr./Cons.	-0.694	0.179	-3.877	0.000
	Health	-0.466	0.114	-4.078	0.000
	Hospitability	-0.050	0.125	-0.399	0.690
	Multiple targets	-0.101	0.114	-0.881	0.378
	Online Serv/Cloud	-0.146	0.116	-1.255	0.210
	Organization-ONG	-0.284	0.138	-2.057	0.040
	Others	-0.113	0.129	-0.875	0.382
	Religion	0.097	0.213	0.456	0.649
	Research-Educ.	-0.094	0.117	-0.797	0.425
	Security	-0.314	0.204	-1.543	0.123
	SW/HW Vendor	-0.158	0.117	-1.351	0.177
	Telco	-0.407	0.159	-2.563	0.010

- Gov-Mil-LEAs-Intelligence are, on average, 0.311 more likely to generate a critical attack and -0.536 less likely to generate a medium severity attack than Automotive;
- Gov. Contractors/Consulting are, on average, 0.403 more likely to generate a critical attack and -0.694 less likely to generate a medium severity attack than Automotive;
- Health is, on average, 0.270 more likely to generate a critical attack and -0.466 less likely to generate a medium severity attack than Automotive;
- Organization/ONG is, on average, 0.165 more likely to generate a critical attack and -0.284 less likely to generate a medium severity attack than Automotive;
- Telco is, on average, 0.237 more likely to generate a critical attack and -0.407 less likely to generate a medium severity attack than Automotive.

4 Conclusion

We have proposed a novel model for cyber risk measurement, based on cumulative logit model that allows to express the cumulative probabilities associated to the different severity levels to a non linear function of a suitable set of explanatory variables. To analize the effect of the explanatory variables on the severity of cyber attack we exploit the Average Marginal Effect. An application of these technique to real data show how to interpret the significant effect of the considered categorical explicative variables on the level of severity of the attacks.

Acknowledgments

This research has received funding from the European Union's Horizon 2020 research and innovation program "FIN-TECH: A Financial supervision and Technology compliance training programme" under the grant agreement No 825215 (Topic: ICT-35-2018, Type of action: CSA).

We thanks the experts of the Hackmanac Project and Clusit for sharing the data set.

References

1. Agresti, A., and Tarantola, C. (2018). Simple Ways to Interpret Effects in Modeling Ordinal Categorical Data, Statistica Neerlandica, 72: 210-223.
2. Afful-Dadzie, A., and Allen, T.T. (2017). Data-Driven Cyber-Vulnerability Maintenance Policies, Journal of Quality Technology, 46: 234-250.
3. Antonielli, A., Bechelli, L., Bosco, F., Butti, G., et al (2019). Rapporto 2019 sulla Sicurezza ICT in Italia, Clusit.

4. Cebula, J.J., and Young, L.R. (2010). A Taxonomy of Operational Cyber Security Risks, Technical Note CMU/SEI-2010-TN-028, Software Engineering Institute, Carnegie Mellon University, 1-34.
5. Edgar, T.W., and Manz, D.O. (2017). Research Methods for Cyber Security, Elsevier.
6. Facchinetti, S., Giudici, P., and Osmetti, S.A. (2018). How to measure cybersecurity risk. In Abbruzzo A., Brentari E., Chiodi M., Piacentino D. (eds), Book of Short Papers SIS 2018, Pearson, 1643-1646.
7. Greene,W. (2008) Econometric Analysis. Upper Saddle River, NJ: Pearson Prentice Hall, 6th edn.
8. Kolfal, B., Patterson, R.A., and Yeo, M.L. (2013). Market Impact on IT Security Spending, Decision Sciences, 44: 517-556.
9. Kopp, E., Kaffenberger, L., and Wilson, C. (2017). Cyber Risk, Market Failures, and Financial Stability, IMF Working Paper WP/17/185, 1-35.
10. Long, J.S., and Freese, J. (2014). Regresson models for categorical dependent variables using Stata (3rd ed.). College Station, TX: Stata Press.