International Journal of Statistics and Applications 2021, 11(4): 79-83
DOI: 10.5923/j.statistics.20211104.01

Adversarial Risk Analysis Structural Equation Model Estimators of Big Data

Boakye Agyemang[1,2,*], Bashiru I. I. Saeed[3], Albert Luguterah[2], Samuel Baffoe[1], Daniel Mbima[1]

[1]Department of Applied Mathematics, Koforidua Technical University, Faculty of Applied Sciences and Technology, Koforidua, Ghana
[2]University for Development Studies, Faculty of Mathematical Sciences, Department of Statistics, Navrongo, Ghana
[3]Tamale Technical University, Faculty of Applied Sciences and Technology, Tamale, Ghana

Abstract This paper seeks to discover the estimators to support decisions in relation to adversarial risk analysis structural equation models of big data using the concept of adversarial risk analysis structural equation modelling (ARA-SEM). A comparative literature examination of hundred (100) related articles published between 2007 and 2020 was employed in the study for the analysis. The results reveal twenty-four (24) eminent estimators discovered to be used in the modelling process. The study concludes that these estimators must be assessed with regard to their properties of statistical significance, unbiasedness, consistency and as well as efficiency in the modeling process of the model.

Keywords Estimates, Estimators, Adversarial Risk Analysis (ARA), Structural Equation Model (SEM), and Adversarial Risk Analysis-Structural Equation Model (ARA-SEM)

1. Introduction

According to ISO (31000:20095) risk refers to the “effect of uncertainty on defined objectives” using probability of occurrence of event as the measurement framework. The assessment and analysis of statistical risk is very crucial and critical in making informed and intelligent decisions particularly for optimum decision science (Boakye et al, 2021). This is the surest way of avoiding overly ambitious or skyrocketing forecasting or otherwise as far as the application of decision science is concerned.

Big data as the non-conventional strategies and innovative technologies applied by businesses and organizations to capture, manage, process, and make sense of huge amount of data requires special or new inductive approach in dealing with such data (Reed, 2017). This is as a result of the big characteristics big data which introduces statistical risks and adversaries known as intelligent opponents whose main aims are to intentionally pose threats to systems in an intelligent manner that is very difficult to detect (Zhaohao, 2018; IBM, 2015). The ISF (2016) corroborates through specifying several adversarial threats within its catalogue which has attracted the attention of researchers such as Rios et al (2009), Banks et al (2015), Ibrahim et al (2015) Kantarcioglu and Xi (2016), among others delving into adversarial risk analysis.

* Corresponding author:
agyemang.boakye@ktu.edu.gh (Boakye Agyemang)
Received: Aug. 4, 2021; Accepted: Sep. 6, 2021; Published: Nov. 15, 2021
Published online at http://journal.sapub.org/statistics

The approaches used by the above-mentioned researchers deals widely with adversarial risk analysis (ARA) to model the intentions and strategic behaviour of adversaries in the cybersecurity domain in particular and as a two-player game in ordinary data rather than big data.

Furthermore, Ibrahim et al (2015) reveal that big data analysis requires a combination of analytical techniques and technologies that include new applications to derive benefit or insights from such data. The aim of this paper therefore is to develop methods to support decisions in relation to adversarial risk analysis of big data by particularly determining some adversarial risk estimators to be derived from big data analysis using adversarial risk analysis structural equation modelling. This would be achieved through the development and application of new a inductive statistical concept to infer laws regarding the determination of the estimators, the models and as well as their fit indices regarding adversarial risks in big data (Billings, 2013).

2. Methodology

The main methodology adopted in this study is the exploratory literature survey of related works particularly with respect to adversarial risk and structural equations models. This is as a result of the fact that the modelling approach combines risk analysis in statistics, the theory of games, and structural equation models to derive interactional adversarial risk problem formulations, and also to provide a model that will extend existing risk analysis models. A total of hundred (100) related works published between 2007 and 2020 were reviewed in the determination of the estimators coupled with some proposed assumptions as listed below. The approach has been utilized by Amado, A., Cortez, P., Paulo, R. and Sérgio, M. (2017) in their work on research trends on big data in marketing, a text mining and topic modeling based literature analysis.

The following conditions therefore must be set out to perform adversarial risk assessment:

i. there exist disruptive events or actions or decisions e_1, e_2, ..., e_∞;
ii. events and decisions are assumed to be independent;
iii. these actions or decisions are interactive in nature;
iv. the probabilities of occurrence exist, and may be uncertain, i.e $p(e_i|ai) = q_i(ai)$.
v. the costs ($c_{i's}$) are conditioned on the occurrence of $e_{i's}$ and decision which are also typically random and the assessment may be a distribution.
vi. there is the presence of adversaries which changes the risk analysis probabilities from being conditioned on one event to being conditioned on two or more events.
vii. the distribution of the adversarial risk estimates must be normally distributed.

3. Results, Discussions and Findings

3.1. Estimators in Adversarial Risk Analysis Modelling of Big Data

Since the adversarial risk analysis combined with structural equation modelling is a virtually new concept which is being proposed by this very research, it was intended to explore some possible estimates that could spur the debate and the research in this area as already indicated in the methodology.

The following estimates have therefore been identified, denoted and defined as summarised in Tables 1a to h.

Table 1a. Estimators of Adversarial Risk Analysis of Big Data (ARA-SEM)

S/N	Estimate	Notation or symbol	Definition or meaning	Purpose
1	Probabilities	$p(e_i\|ai)$	The probabilities of occurrence exist, and may be uncertain	Examine the possibilities of respective events occurring
2	Cost density	$\pi(ci\|ai)$	Measures the uncertain cost that results from each decision.	Used to modelled through the density p(cja).
3	Cost utility	$\cup$ (ci)	It represents the decision maker's utility function with respect to cost.	To examine whether the utility u(c) of the cost is decreasing and typically Nonlinear.
4	Maximum expected utility	ψ	It is a measure of the decision that maximizes the expected utility	Used to derive the value of the decision.

Table 1b. Estimators of Adversarial Risk Analysis of Big Data (ARA-SEM)

S/N	Estimate	Notation or symbol	Definition or meaning	Purpose
5	Risk maximum expected utility	ψ_r	It is a measure of the decision that maximizes the expected utility in the presence of risk	Used to derive the value of the decision in the presence of risk.
6	The multiple risk maximum expected utility	ψ_m	Measures the utilities associated with choices or decisions that within elements of M	Help in the derivation of the value of the decisions in multiples.
7	Factor/Agent correlations	R_A , R_F , R_{IO}	Measures the relationship between the factors or agents or variables.	Useful in exploring interactional opponents relationships
8	Adversarial coefficient of correlation	R_{AR}	Measure the relationship among the factors, opponents or variables in the presence of adversaries.	Bases for informed decision, action and strategy.

Table 1c. Estimators of Adversarial Risk Analysis of Big Data (ARA-SEM)

S/N	Estimate	Notation or symbol	Definition or meaning	Purpose
9	Adversarial coefficient of determination	R_{AR}-Square	It measures the magnitude or degree of relationship in adversaries.	Bases for informed decision, action and strategy.
10	Interactional Risk (Adversarial risk) estimate	$S.E_{AR}$ (Risk)	It measures the magnitude or degree of risk in adversaries.	Bases for informed decision, action and strategy.
11	Adversarial risk significance	P_{AR}-value	Measures the level of significance of adversarial risk.	Bases for informed decision, action and strategy.
12	Adversarial Risk Level	ARL	Measures levels of risk associated with adversarial actions.	Bases for informed decision, action and strategy.

Table 1d. Estimators of Adversarial Risk Analysis of Big Data (ARA-SEM)

S/N	Estimate	Notation or symbol	Definition or meaning	Purpose
13	Adversarial Risk Chi-square	AR$\chi 2$	It is the measure of the Chi-Square value for assessing the overall model fit. It particularly measures the magnitude of discrepancy between the sample and fitted covariances matrices' (Hu and Bentler, 1999: 2)	It measures the goodness of fit a model
14	Adversarial Risk Root mean square error of approximation	ARRMSEA	It is an informative model fit indices which measures fitness of model with unknown but optimally chosen parameter estimates to fit the populations covariance matrix.	Assessing model adequacy or fitness
15	Adversarial Risk Goodness-of-fit statistic	ARGFI	It is for measuring the proportion of variance that is accounted for by the estimated population covariance (Tabachnick and Fidell, 2007).	It is a fit index

Table 1e. Estimators of Adversarial Risk Analysis of Big Data (ARA-SEM)

S/N	Estimate	Notation or symbol	Definition or meaning	Purpose
16	Adversarial Risk Root mean square residual	ARRMR	The RMR is the square root of the difference between the residuals of the sample covariance matrix and the hypothesised covariance model	It is also a form of fit index a model.
17	Adversarial Risk Adjusted Goodness of Fit Index	ARAGFI	It is a modified form of GFI which adjusts the GFI based upon degrees of freedom, with more saturated models reducing fit (Tabachnick and Fidell, 2007)	Model fitness
18	Adversarial Risk Normed-fit index	ARNFI	It is the assessment of the model through comparism of the $\chi 2$ value of the model to the $\chi 2$ of the null model	Model fitness

Table 1f. Estimators of Adversarial Risk Analysis of Big Data (ARA-SEM)

S/N	Estimate	Notation or symbol	Definition or meaning	Purpose
19	Adversarial Risk Comparative fit index	ARCFI	Is the modified form of the NFI which takes into account sample size (Byrne, 1998) that performs well even when sample size is small (Tabachnick and Fidell, 2007). The assumption is that all latent variables are uncorrelated for the null or independence model and compares the sample covariance matrix with this null model	Model fitness and selection
20	Adversarial Risk Incremental Fit Index	ARIFI	Incremental fit index does not use the chi-square in its raw form but compare the chi- square value to a baseline model	Model fitness and selection

Table 1g. Estimators of Adversarial Risk Analysis of Big Data (ARA-SEM)

S/N	Estimate	Notation or symbol	Definition or meaning
21	Adversarial Risk Parsimony Goodness-of-Fit Index	ARPGFI	Parsimony Goodness-of-Fit Index (PGFI) based on the GFI by adjusting for loss of degrees of freedom
22	Adversarial Risk Parsimonious Normed Fit Index	ARPNFI	This adjusts for degrees of freedom based on the NFI (Mulaik et al 1989)
23	Adversarial Risk Akaike Information Criterion	ARAIC	Is a type of parsimony fit index which adjusts for sample size (Akaike, 1974), and compare non-nested or non-hierarchical models estimated with the same data and indicates to the researcher which of the models is the most parsimonious.

Table 1h. Estimators of Adversarial Risk Analysis of Big Data (ARA-SEM)

S/N	Estimate	Notation or symbol	Definition or meaning	Purpose
24	Adversarial Risk Consistent Akaike Information Criterion	ARCAIC	Another form of parsimony fit index which adjusts for sample size	Model reliability and validity

3.2. Discussions and Findings

Twenty-four (24) estimates of adversarial risk analysis structural equation modelling have been deduced from existing literature, the modelling process and as well as the model assumptions giving rise to entirely new discovery propounding approach to propel adversarial risk analysis modelling especially in big data. From literature, fourteen (14) of the estimators were obtained from Schreiber et al (2010) and remaining ten (10) also based on the assumptions and the modelling process of the ARA-SEM model.

This result is consistent with Schreiber et al (2010) particularly on the fit indices, however, it gives an enhancement to the outcome and recommendation of Schreiber, et al (2010) in their work on reporting structural equation modelling and confirmatory factor analysis results primarily due to the addition of the adversarial estimators. The results further imply that since the estimators are consistent with previous studies, they are therefore adopted in the analysis of adversarial risk analysis structural equation models and other related models as well.

4. Conclusions and Recommendations

The study concludes first and foremost that since the estimates are consistent with previous studies, they are therefore adopted in the analysis of adversarial risk analysis structural equation models and other related models as well.

Finally, the paper recommends based on the conclusion that the estimators of the discovered estimates should be explored in order to provide the path for their use in the proposed modelling approach including their statistical properties of significance, unbiasedness, consistency and as well as efficiency in this modeling process of the ARA-SEM model. This recommendation is supported by Adepoju (2007) that properties of estimators should be of interest to researchers after they have deduced them for the typical samples.

REFERENCES

[1] Abdul-Aziz1, A. R., Bashiru, I. I. S. and Luguterah, A. (2019). Structural Equation Modelling with Mediation Variable: *The Perspective of Service Quality on Customer Loyalty of Commercial Banks in Ghana*.

[2] Adepoju, A. A. (2007). Comparative Performance of the Limited Information in a Two Equations Structural Model. Science Focus. Vol 12(2), Pp. 122-129.

[3] Amado, A., Cortez, P., Paulo, R. and Sérgio, M. (2017). Research trends on Big Data in Marketing: A text mining and topic modeling based literature analysis. European Research on Management and Business Economics. http://dx.doi.org/10.1016/j.iedeen.2017.06.002.

[4] Arce, D., and Sandler, T. (2007). ‘‘Terrorist Signalling and the Value of Intelligence,’’ British Journal of Political Science, 37, 573–586. Arnold Publishing.

[5] Banks, D., and Anderson, S. (2006), ‘‘Game Theory and Risk Analysis in the Context of the Smallpox Threat,’’ in Statistical Methods in Counterterrorism, eds. A. Wilson, G. Wilson, and D. Olwell, Springer, New York, pp. 9–22.

[6] Bedford, T., and Cooke, R. (2001). Probabilistic Risk Analysis: Foundations and Methods, Cambridge, U.K: Cambridge University Press.

[7] Billings S. A. (2013). "*Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains*". Wiley.

[8] Boakye, A., Bashiru, I. I. S., Luguterah, A. Baffoe, S. and Mbima, D. (2021) Modeling adversarial risk in big data. International Journal of Science and Research (IJSR). Volume 10, Issue 11, Pg 585-589.

[9] Dalvi, N. N., Domingos, P. M., Mausam, S. K. Sanghai, and Verma, D. (2004). Adversarial classification. In 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, Washington, USA, pages 99–108.

[10] Dedić, N.; Stanier, C. (2017). "Towards Differentiating Business Intelligence, Big Data, Data Analytics and Knowledge Discovery". Heidelberg: Springer International Publishing.

[11] Echos, L. (2013). Big Data car Low-Density Data? La faible densité en information comme facteur discriminant. Sebastopol, CA: O'Reilly Media, p. 1.

[12] French, S., and Rios Insua, D. (2000). Statistical Decision Theory, London: Hodder G. Wilson, and D. Olwell, Springer, New York, pp. 9–22.

[13] Hausken, K. (2002). ‘‘Probabilistic Risk Analysis and Game Theory,’’ Risk Analysis, Homelan Security, ed. Voeller, New York: Wiley.

[14] Hooper, D., Coughlan, J., Mullen, M. (2008). Structural Equation Modelling: Guidelines for Determining Model Fit. Electronic Journal of Business Research Methods, 6(1), 53-60.

[15] Ibrahim; Targio Hashem, Abaker; Yaqoob, Ibrar; Badrul Anuar, Nor; Mokhtar, Salimah; Gani, Abdullah; Ullah Khan, Samee (2015). "big data" on cloud computing: Review and open research issues". Information Systems. 47: 98–115. doi:10.1016/j.is.2014.07.006.

[16] ISO (31000:20095) http://en.wikipedia.org/wiki/ISO_31000.

[17] Kakushadze, Z. and Yu, W. (2016). Statistical Risk Models. Quantigic Solutions LLC. Centre for Computational Biology, Duke- NUS Medical School 8 College Road, Singapore.

[18] Kantarcioglu, M., Xi, B., and Clifton, C. (2011). Classifier evaluation and attribute selection against active adversaries. Data Min. Knowl. Discov., 22(1-2): 291–335, 2011.

[19] Kantarcioglu, M., Xi, B., and Clifton, C. (2016). Adversarial Data Mining: Big Data Meets Cyber Security. Vienna, Austria ACM ISBN 978-1-4503-4139-4/16/10.

[20] Liengaard, B. Sharma, P.N. Hult, G.T.M. Jensen, M.B. Sarstedt, M. Hair, J.F. and Ringle, C.M. (2020) Model Fit. Electronic Journal of Business Research Methods Volume 6

Issue 1. (53-60).

[21] Medvedev, V., Kurasova, O., Bernatavi, J., Treigys, P., Marcinkevi, V., Dzemyda, G. (2017). A new web-based solution for modelling data mining processes. Simulation Modelling Practice and Theory 76 (2017) 34–46. Elsevier.

[22] O'Hagan, A., Buck, C., Daneshkhah, A., Eiser, J., Garthwaite, P., Jenkinson, D., Oakley, J., and Rakow, T. (2006). Uncertain Judgements: Eliciting Experts' of the Smallpox Threat,'' in Statistical Methods in Counterterrorism, eds. A. Wilson.

[23] Pappas, I. O., Kourouthanassis, P. E., Giannakos, M. N., and Lekakos, G. (2017). The interplay of online shopping motivations and experiential factors on personalized e-commerce: A complexity theory approach. Telematics and Informatics, Elsevier. Vol. 34 Pg. 730–742.

[24] Rasoolimanesh, S. M., Ringle, C. M., Sarstedt, M., Olya, H. (2021). The combined use of symmetric and asymmetric approaches: partial least squares-structural equation modeling and fuzzy-set qualitative comparative analysis. International Journal of Contemporary Hospitality Management. Emerald Publishing Limited, Pp. 0959-6119.

[25] Reed, J. (2017), Data Analytics: Applicable Data Analysis to Advance Any Business Using the Power of Data Driven Analytics.

[26] Rios, D. I., Rios, J. & Banks, D. (2009) Adversarial Risk Analysis, Journal of the American Statistical Association, 104:486, 841-854, DOI: 10.1198/jasa.2009.0155.

[27] Schreiber, J. B., Amaury, N., Stage, F. K., Barlow, E. A. and Jamie, K. (2010). Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review, The Journal of Educational Research, Vol 99, Issue 6, Pg 323-338.

[28] Snijders, C.; Matzat, U.; Reips, U.-D. (2012). Big Data': Big gaps of knowledge in the field of Internet". International Journal of Internet Science. 7: 1–5. Springer-Verlag.

[29] Wirthmann A, Karlberg, M., Kovachev B., Reis F. (2015) Structuring risks and solutions in the use of big data sources for producing official statistics –Analysis based on a risk and quality framework. Working Paper, CONFERENCE OF EUROPEAN STATISTICIANS Workshop on Statistical Data Collection: Riding the Data Deluge 29 April – 1 May, Washington D.C., United States of America.

[30] Xindong, W., Xingquan, Z., Gong-Qing, W. and Wei, D. (2014). Data Mining with Big Data. IEEE Transactions on Knowledge and Data Engineering, Vol. 26, No. 1.

[31] Zhaohao, S. (2018). 10 Bigs: Big Data and Its Ten Big Characteristics, BAIS No. 17010, PNG University of Technology.