보험연구원 산학세미나

A Brief Review on Cyber Risk Research and Spatial Features of Cyber Risk Interdependency

2023. 2. 17

포항공과대학교 산업경영공학과

정광민 교수

Definition of Cyber Risk

Two aspects towards the definition

1

세계경제포럼(World Economic Forum) 정의

→ 사이버 위협(cyberthreat) 조직/기관의 가치 있는 자산에 영향을 끼침으로써 궁극적으로 심각한 결과를 유발하는 손실 사건의 실현 가능성 (Probable loss event that materializes when a cyberthreat affects an asset of value and results in a material impact on an organization)

✓ **물리적 사이버 리스크(Physical cyber risk)** : 하드웨어 또는 소프트웨어의 핵심 기술 기반시설상 발생하는 리스크
✓ **정보화 사이버 리스크(Informational cyber risk)** : 데이터 또는 디지털 정보의 유출 또는 파손 리스크
✓ **인지적 사이버 리스크(Cognitive cyber risk)** : 사이버 공간상 개인 또는 집단의 지식, 가치, 믿음, 인식 등의 훼손을 유발하는 리스크

2

Biener, Eling and Wirfs (2015) 정의

→ 정보 및 정보 시스템 상의 기밀성, 가용성 또는 완전성에 부정적 영향을 초래하는 (정보기술자산으로의) 운영 리스크 (Operational risks to information and technology assets that have consequences affecting the confidentiality, availability or integrity of information or information systems)

Definition of cyber risk

Classification of Cyber Risks

1 CRO(Chief Risk Officer) Forum (2016) 분류

사고유형	근본원인	리스크 동인	결과유형
• 시스템 미작동/오용 • 데이터 보안실패 • 데이터 통합/가용성 저해 • 악의적 침해	• 인적위험 • 시스템 및 기술 실패 • 내부 프로세스 실패 • 외부사건	• 국가단위 공격 • 사이버 범죄조직 • 해커집단 • 핵티비스트(Hacktivists) • 내부자	• 사업휴지 • 데이터 손실 • 절도/사기 • 랜섬웨어 또는 사이버 상 갈취 • 개인정보유출 • 평판 손실 • 규제 또는 사법비용 / 과징금 또는 벌금 • 물리적 자산 피해 등

2 Zeller and Scherer (2022) 분류

	개별 사건(Idiosyncratic events)		시스템적 사건(Systemic events)	
	공격유형	예상결과	공격유형	예상결과
데이터유출	표적 데이터 절도	개별 실수에 의한 (의도치 않은) 데이터 유출	광범위한 악성 소프트웨어/피싱에 의한 데이터 절도	클라우드 서비스 공급자(CSP)에 의한 의도치 않은 데이터 유출
사업휴지	표적 디도스/ 랜섬웨어 공격	IT 시스템 미작동 등에 의한 네트워크 장애	광범위한 랜섬웨어 공격	클라우드 서비스 중단에 의한 업무장애(예, 결제시스템 마비)
절도/사기/갈취 등	임원급 내부자와 외부 공격자의 결탁에 의한 표적 정보 절도	관리자 소홀에 의한 데이터베이스 손상	광범위한 랜섬웨어 공격	클라우드 내 보관 중인 데이터 유출

Definition of cyber risk

Classification of Cyber Risks

3

Cebula and Young (2014) + 정광민(2021) 분류
→ 사이버 리스크를 포괄하는 디지털 운영 리스크의 3단계 분류 접근법

✓ **대분류(Core category)** : Basel III 운영리스크에서 제안하는 4개의 위험분류
(인적위험, 시스템 및 기술실패, 내부프로세스 실패, 외부사건)
✓ **소분류(Sub category)** : 각 대분류 요소별 이질적 특성을 갖는 리스크 동인을 묶기 위한 기준점
✓ **리스크 요인(Risk factor)** : 디지털 전환 및 사이버 공간 상 위험 손실사건의 원인을 설명할 수 있는 세부요인

4가지 위험유형 (대분류)

- 인적위험
- 시스템 및 기술 실패
- 내부 프로세스 실패
- 외부사건

→

대분류 위험 유형별
3 – 4가지 위험유형
세분화

(소분류: 전체 13가지 유형)

→

소분류 위험 유형별
3 – 6가지 핵심 리스크
동인 세분화

(리스크요인/동인:
전체 53가지 유형)

Definition of cyber risk

Classification of Cyber Risks

3

Cebula and Young (2014) + 정광민(2021) 분류

→ 사이버 리스크를 포괄하는 디지털 운영 리스크의 3단계 분류 접근법

바젤 분류 구조

4가지 원인	**7가지 손실유형**
• 인적위험 • 시스템 및 기술 실패 • 내부 프로세스 실패 • 외부사건	• 내부사취 • 외부사취 • 고용 및 사업장 안전 • 고객, 상품, 영업실무 • 유형자산 손실 • 시스템 장애 • 집행전달, 처리절차

- **바젤 분류는 광범위한 정의 하 4가지 "원인"을 규정 (인적위험, 시스템 및 기술 실패, 내부 프로세스 실패, 외부사건)**
- **손실사건에 따라 유형을 분류하여 자기자본 산출에 초점**
- **바젤 분류에는 디지털 리스크에 관한 이해 제고를 위한 분류의 체계화/세분화 취약**

- 디지털 운영리스크 이해 제고를 위한 **4가지 원인별 손실 유형의 명확한 세분화** 필요
- 세부 리스크 동인 이해를 위한 **핀셋 분류 필요**
  (전사적 디지털 운영리스크 관리체계 확립을 위한 **Action plan 개발 효율성**)
- 상대적으로 더 자주, 더 큰 파급력을 가진 **리스크 동인에 관한 통계적 이해 제고** 필요

How cyber risk research has progressed over the last decade

Two aspects on cyber risk research

Risk Engineering

❑ **Risk prediction**

- ***Detection of malicious attacks*** (e.g., Okutan et al., 2017; Husak et al., 2018)
- ***Proactively prediction – attack projection, intention recognition, intrusion prediction, network security situation forecasting*** (e.g., Bilge et al., 2017; Xu et al., 2017; Subroto and Apriyana, 2019)

❑ **Risk modeling**

- ***Statistical loss model*** (e.g., Edwards et al., 2016; Eling and Loperfido, 2017; Eling and Jung, 2018)
- ***Extreme risk model*** (e.g., Wheatley et al., 2016; Eling and Wirfs, 2019; Jung, 2021; Malavasi et al., 2022)

Risk Management

❑ **Risk mitigation (Self-protection) & retention**

- ***Optimal investment on cybersecurity*** (e.g., Gordon and Loeb, 2002; Wang, 2019; Krutilla et al., 2021)
- ***Enterprise cyber risk management & risk capital management*** (e.g., Boehme et al., 2019; Eling and Schnell, 2020)

❑ **Risk transfer (cyber insurance)**

- ***Cyber insurance market analysis*** (e.g., Eling and Schnell, 2016; Romanosky, 2016; Pooser et al., 2018; Romanosky et al., 2019; Xie et al., 2020; Cole and Fier, 2021)
- ***Cyber insurance rate-making*** (e.g., Yang et al., 2020; Eling, Jung and Shim, 2022)

How cyber risk research has progressed over the last decade

Literature on cyber risk engineering

Risk prediction

- ***Graph models***
  - Bayesian network to forecast cyber incidents (Okutan et al., 2017); Graphical presentation of cyber attack scenarios (Husak et al., 2018); Markov time-varying model (Li et al., 2020)
- ***Time series (attack arrival)***
  - ARMA-GARCH or copula-GARCH (Chen et al., 2015; Xu et al., 2017)
- ***Machine learning approach***
  - Random Forest classifier (Bilge et al., 2017); Neural Networks (Subroto and Apriyana, 2019)

Risk modeling

- ***Loss distribution***
  - Negative binomial approach (Edwards et al., 2016); Tweedie approach (Eling and Jung, 2022)
- ***Loss dependency with copulas***
  - Elliptical family copulas (Boehme and Kataria, 2006); Archimedean copulas (Herath and Herath, 2007); Vine copulas (Eling and Jung, 2018; Peng et al.,2018)
- ***Extreme value theory***
  - Power-law based EVT (Wheatley et al., 2016); Block maxima with ARMA-GARCH (Jung, 2021)

How cyber risk research has progressed over the last decade

Literature on cyber risk management

Risk mitigation & retention

- ***Optimal investment on cybersecurity***
  - Optimal level of cybersecurity investment with cost-benefit difference maximization (Gordon and Loeb, 2002); Optimal level between cybersecurity investment and cyber insurance (Wang, 2019)
- ***Enterprise cyber risk management and cyber risk capital***
  - Top-down or bottom-up approach by risk management process (Boehme et al., 2019); Cyber risk capital requirement under Solvency II, US RBC and SST (Eling and Schnell, 2020)

Risk transfer (cyber insurance)

- ***Cyber insurance market analysis***
  - Status quo analysis on the US cyber insurance market (Romanosky et al., 2019); Determinants of cyber insurance participation and current performance (Xie et al., 2020)
- ***Cyber insurance rate-making***
  - Cyber insurance pricing for cyber-physical power systems under insurer insolvency (Yang et al., 2020); Quantile-based rate-making by industry, firm size and security level (Eling, Jung and Shim, 2022)

What is missing in the literature

Research motivation

Aspect 1 :

- Spatial features of internet systems?
- Critical internet infrastructures feature physical spatial network systems (Tranos, 2013; Schmidtke, 2018) .
- Internet use delays rely on physical distances measured by roundtrip time (Schmidtke, 2018).
- In addition, telecommunication firms may decide to construct internet networks in agglomeration economies for profitability (Malecki, 2002; Priemus, 2007).

Aspect 2 :

- Socio-economic features may appear to address cyber risk event frequency (Park et al., 2019; Chen et al., 2021).
- The Social Disorganization Theory (SDT) can support this potential appearance.
- But, a regional level analysis on data breach occurrence is limited.
- Spatio-temporal patterns of such features may exist in regional clusters of the cyber risk landscape.

What is missing in the literature

Relevant literature review

	Aspect 1: Static cyber loss analysis			Aspect 2: Spatial/Socio-economic analysis on cyber risks			
	Eling and Loperfido (2017)	Eling and Wirfs (2019)	Jung (2021)	Khey and Sainato (2013)	Park et al. (2019)	Chen et al. (2021)	**Present study**
Data	PRC (2005 - 2015)	SAS OpRisk (1995 – 2014)	• Cowbell Cyber (2005 - 2018) • PRC (2005 - 2018)	PRC (2005 - 2012)	State-level data from multiple sources (2004 - 2010)	China Judgements Online database (2014 - 2018)	• PRC (2005 - 2018) • Social Determinants of Health Database (2009 - 2018)
Sample size	2,266	26,541	21,555	3,226	355	6,106	• 5,748 (PRC); • 32,245 (SDOH)
Method	• Multi-dimensional scaling • Multiple factor analysis for contingency tables • Goodness-of-fit	• Loss distribution approach • Dynamic extreme value theory	• Generalized Extreme Value distribution • Time series analysis	Moran's I statistics	Panel regression	• Morans' I statistics • Generalized additive model	• Morans' I statistics • Spatial lag/error models
Focus of study	Distribution fitting of cyber risk and risk measurement	Distribution fitting of cyber risk and firm-specific characteristics for extremes	Statistical features of extreme cyber losses	Spatial cluster analysis of data breaches	Relationship between socio-economic factors and cybercrimes	Spatio-temporal pattern of cyber frauds in China	Comprehensive spatial analysis of data breach events
Main findings	• Clusters exist by types of data breaches • Skew-normal distribution is optimal for cyber severity	• log-normal distribution might over-estimate cyber losses • The larger the firm size, the more exposed to extreme losses	• Threshold-based estimation might underestimate extreme losses • The cost of a smaller breach is larger than the cost of larger breach	Breaches tend to occur within particular geo-clusters	Income, degree of education, poverty rate, inequality make the Internet penetration be more related with cyber crime	The distribution of cyber fraud events is affected by the regional economy and population	• Spatial dependency exists in terms of county-level • Population and income are generally related with cyber risk

What need to be addressed in this study

Key research questions

Question 1:

Do data breaches have a spatial pattern in the U.S.?

1. If so, which regions are more exposed to data breach risks
2. Whether there is a regional cluster in data breach event frequency
3. What risk types or industries appear to be more affected by such clusters

Question 2:

What socio-economic factors address the occurrence of data breaches?

1. How can the size of cyber risk exposures address the occurrence of data breaches?
2. What industrial features may address the occurrence?

Contributions

- We explore spatial dependency between states / counties of the U.S. and spatial impacts of socio-economic factors in the frequency of data breaches.
- This exploration is carried out with a dataset combining data breach risk data with geo-graphical information and socio-economic data, the combination that has not been used in the literature

Methodology

Moran's I (Anselin, 1995; Darmofal, 2015)

- Global Moran's I
  - A single value that measures global spatial autocorrelation ($-1 \le I \le 1$)
  - $I = \frac{N}{S}\frac{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}(y_i-\bar{y})(y_j-\bar{y})}{\sum_{i}^{N}(y_i-\bar{y})^2}$, where w_{ij} is an element of $N \times N$ weight matrix with N as the number of regions, S is the sum of the weights, y_i is observation at i^{th} region
    - $I \approx 1$: similar values within the region
    - $I \approx -1$: dissimilar values within the region
    - $I \approx 0$: no spatial autocorrelation exists over all areas

- Local Moran's I
  - A single value that measures local spatial autocorrelation of single region ($-1 \le I_i \le 1$)
  - $I_i = \sum_{j}^{J_i} w_{ij}(y_i - \bar{y})(y_j - \bar{y})$, where J_i is the neighborhood set of area
  - Each region can be defined as a hot or cold spot depending on neighboring regions

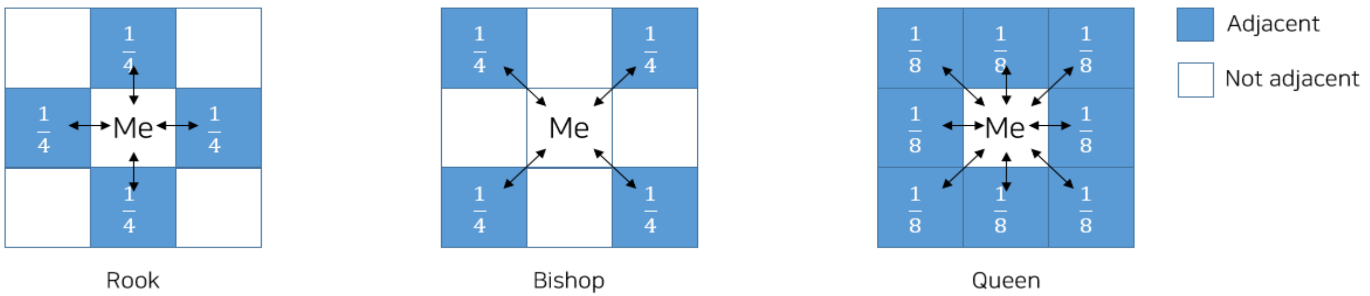
	Region	Neighbor
High-High(HH)	Higher	High
High-Low(HL)	Higher	Low
Low-High(LH)	Lower	High
Low-Low(LL)	Lower	Low

→ Local Moran's I identifies local cluster/spatial outliers

Methodology

Spatial Weight Matrix (W) (Fischer and Wang, 2011)

- It designs the way to impose weights between adjacent regions
- There are several ways to construct a spatial weight matrix (defining the concept of contiguity)
- We use the Queen contiguity to model possible adjacent sources of autocorrelation



Spatial panel regression (Elhorst, 2014)

- Spatial Autoregressive Model (SAR) considers endogenous interaction effects
- Spatial Error Model (SEM) considers potential impacts of those variables in the error term
- Spatial Autoregressive Combined Model (SAC) offers both endogenous and error interaction effects by incorporating the spatial autocorrelation in the response variable and the spatial correlation with latent factors

Methodology

Spatial panel regression (Elhorst, 2014)

- Spatial Autoregressive Model (SAR) considers endogenous interaction effects
- Spatial Error Model (SEM) considers potential impacts of those variables in the error term
- Spatial Autoregressive Combined Model (SAC) offers both endogenous and error interaction effects by incorporating the spatial autocorrelation in the response variable and the spatial correlation with latent factors

SAR	SEM	SAC
$y_{it} = \rho \sum_{k=1}^{N} w_{ik} y_{kt} + \boldsymbol{x_{it}\beta} + \mu_i + \xi_t + \epsilon_{it}$	$y_{it} = \boldsymbol{x_{it}\beta} + \mu_i + \xi_t + u_{it}$ $u_{it} = \lambda \sum_{j=1}^{N} w_{ij} u_{it} + \epsilon_{it}$	$y_{it} = \rho \sum_{k=1}^{m} w_{ik} y_{kt} + \boldsymbol{x_{it}\beta} + \mu_i + \xi_t + u_{it}$ $u_{it} = \lambda \sum_{j=1}^{N} m_{ij} u_{it} + \epsilon_{it}$

where μ_i, ξ_i are spatial specific effects that control for all time-invariant variable or spatial-invariant variable

Data description

Empirical cyber risk data

- **Data provider:** Privacy Rights Clearinghouse (PRC)
- **Sample size:** 9,034 from 2005 to 2019
- **Provided information:** Year of breach, Date made public, State, City, Latitude, Longitude, Breach type, Industry type, Total records, Company, Description of incident, Information source
- **Types of data breach**

Type	Summary	Type	Summary
CARD	Debit/credit card fraud	PORT	Loss of portable device(s)
HACK	Hacking by outside/malware	STAT	Stationary computer loss
INSD	Insider of the organization	DISC	Unintended disclosure of data
PHYS	Physical damage/loss	UNKN	Unknown

- **Types of organization**

Type	Summary	Type	Summary
BSF	Financial services	GOV	Government, utility
BSR	Retailers	MED	Healthcare/medical service provider
BSO	Other businesses	NGO	Non-profit organization
EDU	Educational institution	UNKN	Unknown

Data description

Empirical cyber risk data

- **Data provider:** Social Determinants of Health (SDOH)
- **Collected information:** Population, Income, Wholesale, Retail, Finance, Education, Administrative, Armed forces
- **Variables used in the study**

Variable	Description
Population	Population of region
Income ($)	Per capita income (in dollars, inflation-adjusted to file data each year)
Wholesale (%)	Percentage of the employed working in wholesale trade
Retail (%)	Percentage of the employed working in retail trade
Finance (%)	Percentage of the employed working in finance and insurance, real estate, and rental and leasing
Education (%)	Percentage of the employed working in educational services, and healthcare and social assistance
Administrative (%)	Percentage of the employed working in public administration
Armed forces (%)	Percentage of the employed working in armed forces

Data description

Summary statistics (County-level)

	Mean	Std	Min	Median	Max
Breach frq (State-level)	7.92	15.67	0.00	3.00	164.00
Breach frq	0.14	1.23	0.00	0.00	154.00
Population	98,191.14	314,909.20	41.00	26,003.00	10,105,720.00
Income ($)	23,773.12	6,323.85	5,327.00	23,126.50	72,832.00
Wholesale (%)	2.46	1.21	0.00	2.37	30.56
Retail (%)	11.44	2.46	0.00	11.53	41.67
Finance (%)	4.64	1.95	0.00	4.33	22.82
Education (%)	22.83	4.67	2.02	22.49	52.65
Administrative (%)	5.80	3.34	0.00	4.87	48.33
Armed forces (%)	0.32	1.66	0.00	0.05	81.25

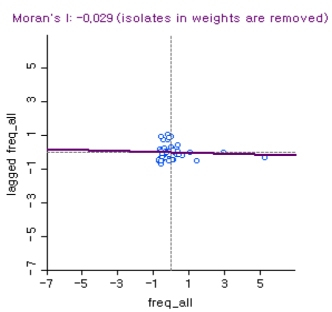
Aspect 1 : Spatial loss clusters

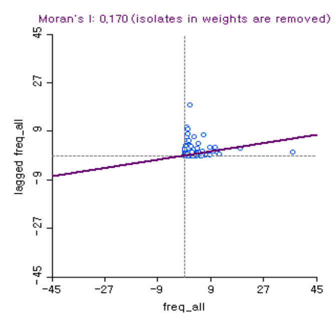
Global Moran's I

Year	State-level spatial statistics	County-level spatial statistics
2005	-0.046	0.076***
2006	-0.021	0.130***
2007	-0.005	0.096***
2008	-0.047	0.071***
2009	0.053	0.088***
2010	-0.025	0.164***
2011	-0.024	0.167***
2012	0.036	0.192***
2013	-0.025	0.185***
2014	-0.032	0.033***
2015	-0.016	0.234***
2016	-0.048	0.290***
2017	-0.035	0.196***
2018	-0.072	0.018**
Entire period	-0.029	0.170***

Note: *, **, and ***, indicate significance level of 10%, 5%, and 1%.

State-level spatial dependency	County-level spatial dependency



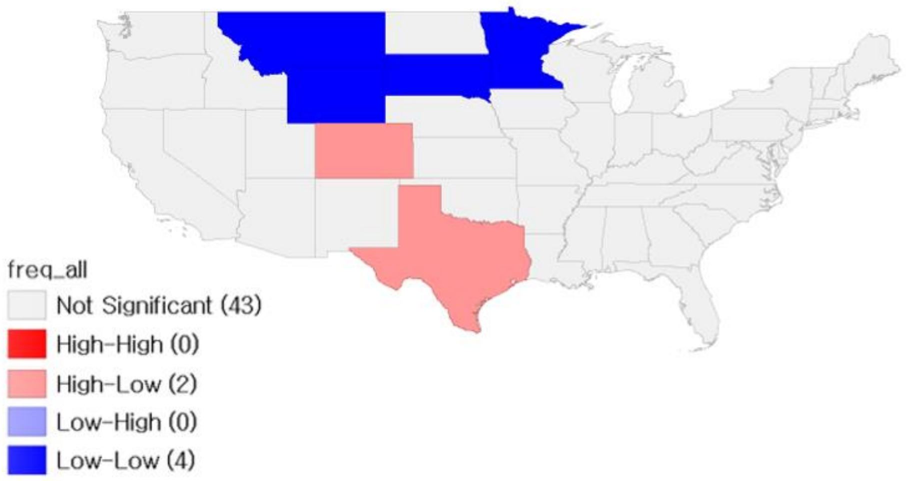


- There is no statistical evidence on spatial dependency across states.
- There is a significant evidence on the dependency across counties at the 1% significance level.
- Relatively more exposed counties and less exposed counties are clustered respectively.

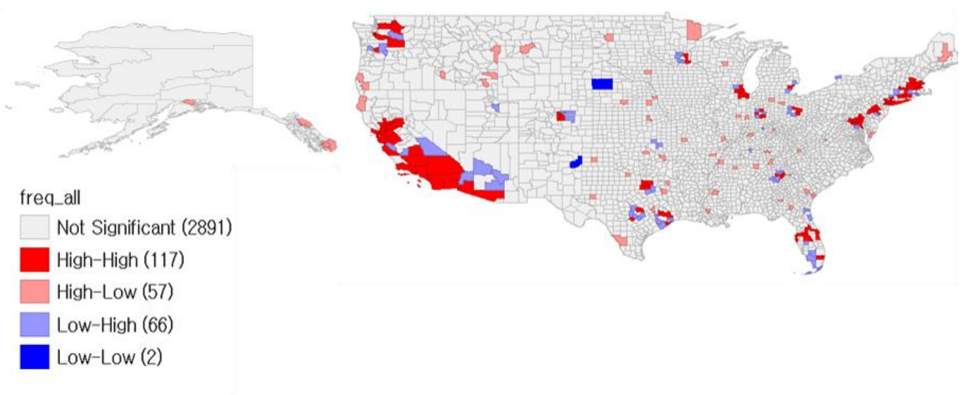
Aspect 1 : Spatial loss clusters

Local Moran's I

State-level



County-level



- Northern states (part of mid-west and west divisions) overall tend to be less exposed to data breach events, categorized as low-low areas at the state-level.
- Texas and Colorado states are found to be high-low areas (more exposed to data breaches, whereas their neighbors are less exposed).
- West and east coast regions are more exposed to data breach events at the county-level.

Aspect 2 : Spatial socio-economic drivers on cyber risks

Spatial panel analysis on all counties

	Dependent variable: Data breach event frequency at the county level			
	OLS	**SAR**	**SEM**	**SAC**
Constant	-5.140*** (0.493)	-5.592*** (0.504)	-5.594*** (0.504)	-5.545*** (0.502)
ln (Population)	0.223*** (0.008)	0.223*** (0.008)	0.223*** (0.008)	0.222*** (0.008)
ln (Income)	0.359*** (0.048)	0.405*** (0.050)	0.405*** (0.050)	0.401*** (0.049)
Wholesale	-2.350*** (0.813)	-2.505*** (0.813)	-2.509*** (0.813)	-2.508*** (0.808)
Retail	-5.209*** (0.409)	-5.159*** (0.409)	-5.155*** (0.409)	-5.096*** (0.407)
Finance	4.679*** (0.591)	4.384*** (0.594)	4.388*** (0.594)	4.436*** (0.591)
Education	-0.918*** (0.207)	-0.875*** (0.207)	-0.875*** (0.207)	-0.878*** (0.206)
Administration	0.785** (0.309)	0.786** (0.309)	0.786** (0.309)	0.788** (0.308)
Armed force	-2.121*** (0.632)	-2.169*** (0.632)	-2.171*** (0.632)	-2.173*** (0.629)
ρ (Spatial autocorrelation)	-	-0.005 (0.011)	-	-0.129** (0.050)
λ (Spatial error dependency)	-	-	0.002 (0.011)	0.125*** (0.047)
Loglikelihood	-37,910.13	-37,895.69	-37,895.77	**-37,892.46**
AIC	75,838.26	75,825.39	75,825.54	**75,820.92**
Adj R^2	0.078	0.080	0.080	0.077
Observation	21,931	21,931	21,931	21,931

Note: We take the transformation of natural logarithm for two continuous variables (population and income). *, **, and *** indicate significance levels of 10%, 5%, and 1%, respectively.

- Population and average income level of a county are positive and significant at the 1% confidence level.
- Financial industry and public administration sector are positive and significant in explaining the data breach frequency.
- The other industries (wholesale, retail, education and armed force) are negative and significant.
- Spatial coefficients of the SAC are all significant (interpretation in the next slide).

Aspect 2 : Spatial socio-economic drivers on cyber risks

Spatial panel analysis on all counties

Spatial effects of the SAC model

	Direct effect	Indirect effect	Total effect
ln (Population)	**0.223**	**-0.026**	0.197
ln (Income)	**0.402**	**-0.047**	0.356
Wholesale	-2.515	0.293	-2.222
Retail	-5.110	0.595	-4.515
Finance	**4.448**	**-0.518**	3.930
Education	-0.880	0.103	-0.778
Administration	**0.790**	**-0.092**	0.698
Armed force	-2.179	0.254	-1.925

- Population and average income level of a region have **positive direct effects**, but **negative indirect effects**

  → A county with large population or high income is more likely to be exposed to data breach events itself, however, neighboring regions may have less likelihood of such events

- Financial industry and public administration sector also have **positive direct effects**, but **negative indirect effects**

  → Counties with higher proportion of the financial industry or public administration sector tend to be more exposed to data breach events themselves, but to have less spatial impacts on neighboring regions

Aspect 2 : Spatial socio-economic drivers on cyber risks

Spatial panel analysis on California

	Dependent variable: Data breach event frequency at the county level			
	OLS	SAR	SEM	SAC
Constant	-33.523*** (8.708)	-39.196*** (8.646)	-35.564*** (8.684)	-44.815*** (8.127)
ln (Population)	1.898*** (0.130)	1.923*** (0.128)	1.919*** (0.130)	1.851*** (0.122)
ln (Income)	2.490*** (0.818)	3.058*** (0.832)	2.693*** (0.819)	3.708*** (0.750)
Wholesale	-112.40*** (21.914)	-112.51*** (21.666)	-115.33*** (21.710)	-109.30*** (20.157)
Retail	-36.668*** (8.677)	-35.485*** (8.471)	-36.750*** (8.487)	-28.412*** (7.892)
Finance	-38.516*** (10.362)	-45.725*** (10.594)	-43.398*** (10.569)	-47.938*** (9.767)
Education	-18.935*** (4.944)	-18.853*** (4.816)	-18.571*** (4.858)	-20.444*** (4.411)
Administration	1.639 (3.944)	1.310 (3.838)	0.964 (3.850)	3.343 (3.578)
Armed force	-43.695** (18.294)	-38.714** (17.924)	-41.497** (17.991)	-44.381*** (16.083)
ρ (Spatial autocorrelation)	-	-0.121* (0.065)	-	-0.531*** (0.096)
λ (Spatial error dependency)	-	-	-0.071 (0.076)	0.446*** (0.086)
Loglikelihood	-996.511	-990.243	-991.607	**-986.080**
AIC	2,011.023	2,014.486	2,017.215	**2,008.159**
Adj R^2	0.479	0.506	0.500	0.473
Observation	406	406	406	406

Note: We take the transformation of natural logarithm for two continuous variables (population and income). *, **, and *** indicate significance levels of 10%, 5%, and 1%, respectively.

- Population and average income level of a county are positive and significant at the 1% confidence level.
- Financial industry is negative and significant, whereas public administration sector is positive but insignificant.
- The other industries (wholesale, retail, education and armed force) are negative and significant.
- Spatial coefficients of the SAC are all significant (interpretation in the next slide).

Aspect 2 : Spatial socio-economic drivers on cyber risks

Spatial panel analysis on California

Spatial effects of the SAC model

	Direct effect	Indirect effect	Total effect
ln (Population)	**1.957**	**-0.748**	1.209
ln (Income)	**3.920**	**-1.499**	2.421
Wholesale	-115.563	44.183	-71.380
Retail	-30.039	11.485	-18.554
Finance	-50.684	19.378	-31.306
Education	-21.615	8.264	-13.351
Administration	3.535	-1.351	2.183
Armed force	-46.923	17.940	-28.923

- Population and average income level of a region have **positive direct effects**, but **negative indirect effects**

  → A county with large population or high income is more likely to be exposed to data breach events itself, however, neighboring regions may have less likelihood of such events

- Financial industry has **negative direct effects**, but **positive indirect effects**

  → Counties with higher proportion of the financial industry tend to be less exposed to data breach events themselves, but to have higher spatial impacts on neighboring regions

Aspect 2 : Spatial socio-economic drivers on cyber risks

Spatial panel analysis on Hacking risk type

	Dependent variable: Data breach event frequency at the county level			
	OLS	**SAR**	**SEM**	**SAC**
Constant	-3.088*** (0.273)	-3.066*** (0.279)	-3.067*** (0.279)	-3.043*** (0.278)
ln (Population)	0.097*** (0.004)	0.097*** (0.004)	0.097*** (0.004)	0.097*** (0.004)
ln (Income)	0.245*** (0.027)	0.243*** (0.027)	0.243*** (0.027)	0.242*** (0.027)
Wholesale	-1.647*** (0.450)	-1.635*** (0.450)	-1.637*** (0.450)	-1.646*** (0.448)
Retail	-2.403*** (0.227)	-2.408*** (0.227)	-2.407*** (0.227)	-2.378*** (0.226)
Finance	1.978*** (0.327)	1.989*** (0.329)	1.990*** (0.329)	2.021*** (0.328)
Education	-0.451*** (0.114)	-0.457*** (0.115)	-0.457*** (0.115)	-0.460*** (0.114)
Administration	0.374** (0.171)	0.371** (0.171)	0.371** (0.171)	0.373** (0.170)
Armed force	-1.169*** (0.350)	-1.164*** (0.350)	-1.164*** (0.350)	-1.169*** (0.348)
ρ (Spatial autocorrelation)	-	-0.005 (0.011)	-	-0.134** (0.058)
λ (Spatial error dependency)	-	-	-0.000 (0.011)	0.128** (0.054)
Loglikelihood	-24,942.42	-24,937.05	-24,937.17	**-24,934.32**
AIC	**49,902.83**	49,908.1	49,908.33	49,904.64
Adj R^2	0.056	0.057	0.057	0.054
Observation	21,931	21,931	21,931	21,931

Note: We take the transformation of natural logarithm for two continuous variables (population and income). *, **, and *** indicate significance levels of 10%, 5%, and 1%, respectively.

- Population and average income level of a county are positive and significant at the 1% confidence level.
- Financial industry and public administration sector are positive and significant in explaining the data breach frequency.
- The other industries (wholesale, retail, education and armed force) are negative and significant.
- Spatial coefficients of the SAC are all significant (interpretation in the next slide).

Aspect 2 : Spatial socio-economic drivers on cyber risks

Spatial panel analysis on Hacking risk type

Spatial effects of the SAC model

	Direct effect	Indirect effect	Total effect
ln (Population)	**0.097**	**-0.012**	0.085
ln (Income)	**0.242**	**-0.029**	0.213
Wholesale	-1.651	0.200	-1.451
Retail	-2.385	0.289	-2.096
Finance	**2.028**	**-0.246**	1.782
Education	-0.461	0.056	-0.405
Administration	**0.374**	**-0.045**	0.329
Armed force	-1.173	0.142	-1.031

- Population and average income level of a region have **positive direct effects**, but **negative indirect effects**

  → A county with large population or high income is more likely to be exposed to data breach events itself, however, neighboring regions may have less likelihood of such events

- Financial industry and public sector have **positive direct effects**, but **negative indirect effects**

  → Counties with higher proportion of the financial industry or public administration tend to be more exposed to hacking events themselves, but to have less spatial impacts on neighboring regions

AIRM RESEARCH GROUP

Conclusion

Research questions		Findings
1) Do data breaches have a spatial pattern in the U.S.?	→	✓ There is no statistical evidence on spatial dependency across states. ✓ At the county-level, spatial autocorrelation exists.
2) What socio-economic factors address the occurrence of data breaches?	→	✓ Larger or richer counties can be more exposed to data breach events themselves, but their neighboring counties may less experience such events. ✓ Counties next to larger or richer counties in California are less likely to be exposed to data breach events. ✓ Counties adjacent to richer counties tend to more experience hacking events.

Further implications

- Businesses in a region with relatively large population or high-income level may need to be more regulated with respect to cybersecurity enhancement.
- Financial industry concentrated regions (i.e., local-level financial hubs) or those with critical public infrastructures (or governmental agencies) should be incentivized to enhance cyber risk management.

References

1) 정광민. (2021). 금융산업의 디지털 전환과 운영리스크: 은행과 보험산업 중심으로. 보험연구원 연구보고서 2021-07.
2) Anselin, L. (1995). Local indicators of spatial association-LISA. *Geographical Analysis*, *27*(2), 93-115.
3) Biener, C., Eling, M., and Wirfs, J. H. (2015). Insurability of cyber risk: An empirical analysis. The Geneva Papers on Risk and Insurance-Issues and Practice, 40(1), 131-158.
4) Bilge, L., Han, Y., and Dell'Amico, M. (2017). Riskteller: Predicting the risk of cyber incidents. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, pages 1299-1311.
5) Boehme, R., Laube, S., and Riek, M. (2019). A fundamental approach to cyber risk analysis. Variance, 12(2), 161-185.
6) Cebula, J. J., and Young, L. R. (2014). A taxonomy of operational cyber security risks version 2. Software Engineering Institute: Carnegie Mellon University.
7) Chen, S., Gao, C., Jiang, D., Hao, M., Ding, F., Ma, T., Zhang, S., and Li, S. (2021). The spatiotemporal pattern and driving factors of cyber fraud crime in China. *ISPRS International Journal of Geo-Information*, *10*(12), 802.
8) Chong, W. F., Feng, R., and Jin, L. (2021). Holistic principle for risk aggregation and capital allocation. *Annals of Operations Research*, 1-34.
9) Cole, C. R., and Fier, S. G. (2021). An empirical analysis of insurer participation in the US cyber insurance market. North American Actuarial Journal, 25(2), 232-254.
10) CRO Forum. (2016). Concept paper on a proposed categorization methodology for cyber risk. London: CRO Forum.
11) Darmofal, D. (2015). *Spatial analysis for the social sciences*. New York: Cambridge University Press.
12) Edwards, B., Hofmeyr, S., and Forrest, S. (2016). Hype and heavy tails: A closer look at data breaches. Journal of Cybersecurity, 2(1), 3-14.
13) Elhorst, J. P. (2014). *Spatial econometrics: from cross-sectional data to spatial panels*. Heidelberg: Springer.
14) Eling, M., and Schnell, W. (2016). What do we know about cyber risk and cyber risk insurance?. Journal of Risk Finance, 17(5), 474-491.
15) Eling, M., and Schnell, W. (2020). Capital requirements for cyber risk and cyber risk insurance: An analysis of Solvency II, the U.S. risk-based capital standards, and the Swiss Solvency Test. North American Actuarial Journal, 24(3), 370{392.
16) Eling, M., Jung, K., and Shim, J. (2022). Unraveling heterogeneity in cyber risks using quantile regressions. *Insurance: Mathematics and Economics, 104*, 222-242.
17) Eling, M., and Loperfido, N. (2017). Data breaches: Goodness of fit, pricing, and risk measurement. *Insurance: Mathematics and Economics*, *75*, 126-136.
18) Eling, M., and Wirfs, J. H. (2019). What are the actual costs of cyber risk events?. *European Journal of Operational Research*, *272*(3), 1109-1119.
19) Eisenbach, T. M., Kovner, A., and Lee, M. J. (2021). Cyber risk and the US financial system: A pre-mortem analysis. Journal of Financial Economics. 145(3), 802-826
20) Fischer, M. M., and Wang, J. (2011). Spatial data analysis: models, methods and techniques. Springer Science & Business Media.
21) Gordon, L. A. and Loeb, M. P. (2002). The economics of information security investment. ACM Transactions on Information and System Security (TISSEC), 5(4), 438-457.
22) Husak, M., Komarkova, J., Bou-Harb, E., and Celeda, P. (2018). Survey of attack projection, prediction, and forecasting in cyber security. IEEE Communications Surveys & Tutorials, 21(1), 640-660.
23) Jung, K. (2021). Extreme data breach losses: An alternative approach to estimating probable maximum loss for data breach risk. North American Actuarial Journal, 25(4), 580-603.

References

24) Khey, D. N., and Sainato, V. A. (2013). Examining the correlates and spatial distribution of organizational data breaches in the United States. Security Journal, 26(4), 367-382.
25) Krutilla, K., Alexeev, A., Jardine, E., and Good, D. (2021). The benefits and costs of cybersecurity risk reduction: A dynamic extension of the Gordon and Loeb model. Risk Analysis, 41(10), 1795-1808.
26) Malavasi, M., Peters, G. W., Shevchenko, P. V., Truck, S., Jang, J., and Sofronov, G. (2022). Cyber risk frequency, severity and insurance viability. Insurance: Mathematics and Economics, 106, 90-114.
27) Malecki, E. J. (2002). Hard and soft networks for urban competitiveness. Urban Studies, 39(5-6), 929-945.
28) Okutan, A., Yang, S. J., and McConky, K. (2017). Predicting cyber attacks with Bayesian networks using unconventional signals. In Proceedings of the 12th Annual Conference on Cyber and Information Security Research.
29) Park, J., Cho, D., Lee, J. K., and Lee, B. (2019). The economics of cybercrime: The role of broadband and socioeconomic status. ACM Transactions on Management Information Systems (TMIS), 10(4), 1-23.
30) Pooser, D. M., Browne, M. J., and Arkhangelska, O. (2018). Growth in the perception of cyber risk: evidence from US P&C insurers. The Geneva Papers on Risk and Insurance-Issues and Practice, 43, 208-223.
31) Priemus, H. (2007). The network approach: Dutch spatial planning between substratum and infrastructure networks. European Planning Studies, 15(5), 667-686.
32) Romanosky, S. (2016). Examining the costs and causes of cyber incidents. Journal of Cybersecurity, 2(2), 121-135.
33) Romanosky, S., Ablon, L., Kuehn, A., and Jones, T. (2019). Content analysis of cyber insurance policies: how do carriers price cyber risk? Journal of Cybersecurity, 5(1), tyz002.
34) Schmidtke, H. R. (2018). Is the internet spatial?. Journal of Reliable Intelligent Environments, 4(3), 123-129.
35) Subroto, A. and Apriyana, A. (2019). Cyber risk prediction through social media big data analytics and statistical machine learning. Journal of Big Data, 6(1), 50.
36) Tranos, E. (2013). The geography of the internet: Cities, regions and internet infrastructure in Europe. Cheltenham: Edward Elgar.
37) Wang, S. S. (2019). Integrated framework for information security investment and cyber insurance. Pacific-Basin Finance Journal, 57, 101173.
38) Wheatley, S., Maillart, T., and Sornette, D. (2016). The extreme risk of personal data breaches and the erosion of privacy. The European Physical Journal B, 89(7), 1-12.
39) Xie, X., Lee, C., and Eling, M. (2020). Cyber insurance offering and performance: An analysis of the US cyber insurance market. The Geneva Papers on Risk and Insurance-Issues and Practice, 45, 690-736.
40) Xu, M., and Hua, L. (2019). Cybersecurity insurance: Modeling and pricing. North American Actuarial Journal, 23(2), 220-249.
41) Xu, M., Hua, L., and Xu, S. (2017). A vine copula model for predicting the effectiveness of cyber defense early warning. Technometrics, 59(4), 508-520.
42) Yang, Z., Liu, Y., Campbell, M., Ten, C. W., Rho, Y., Wang, L., and Wei, W. (2020). Premium calculation for insurance businesses based on cyber risks in IP-based power substations. IEEE Access, 8, 78890-78900.
43) Zeller, G., and Scherer, M. (2022). A comprehensive model for cyber risk based on marked point processes and its application to insurance. European Actuarial Journal, 12(1), 33-85.