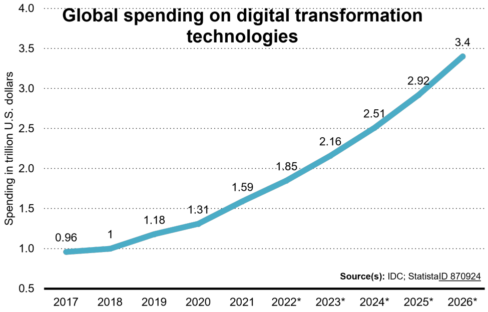
Data Risk, Firm Growth, and Innovation

Orlando Gomes (ISCAL),
Roxana Mihet (HEC Lausanne and SFI),
Kumar Rishabh (HEC Lausanne and Uni Basel)

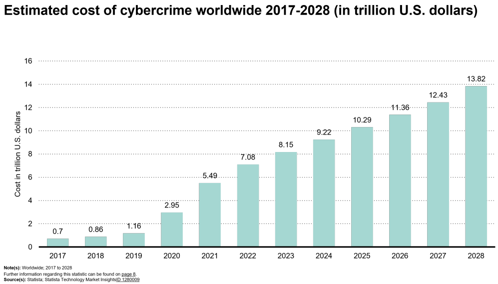
EEA-ESEM 2024

Rotterdam

Emergence of the data economy seems inevitable, but it comes with risks



Firms are spending exponentially more on data (AI) technologies. Lots of opportunities for customization and efficiency.



Severity of attacks is on a rise; Cost surpasses the GDP of all but U.S. and China.

How do firms react in face of increasing cyber risk?

Question (1): How do firms change their financial, growth and innovation strategies in face of increasing cyber risk?

- Divert resources from innovation into protection
- Reduce growth, profitability, innovation
- Risk might impact AI-intensive firms the most

Question (2): Can cyber riss spur growth & innovation, especially in high-tech sectors?

- Forces innovation in data security
- Could spurs broader tech advances
- Could benefit high-tech firms most, as they're at the digital forefront—possibly transforming data security challenges into innovation drivers

1-Click to success: The data security innovations behind Amazon's e-commerce dominance

- Amazon's 1-click ordering system revolutionized e-commerce
- Amazon's patent that underpins its 1-click ordering is its most cited patent—once Apple licensed it for iTunes
- This innovation is built on Amazon's earlier breakthrough patents in secure transmission of credit card information over unsecured network like internet
- These CS patents are Amazon's 7^{th} and 9^{th} most cited patents ever

Study design

- We study these questions both empirically and theoretically.
- **Empirically:** In the context of the US public firms...
  - Study the firm profitability, growth and innovation response to data risk
  - Develop a method to identify AI-intensive firms
  - To make *causal* statements, we use a quasi-experimental difference-in-difference analysis to study the impact of data breach notification laws on innovation
- **Theoretically:** Build a growth model...
  - Firms are subject to data risk (their data can be destroyed by cyber criminals)
  - AI-intensive firms invest in in-house data security
  - Non-AI firms buy external data security from AI firms
  - In-house data security augments product quality, external data security does not

Data and methodology

US Firm-level data: 2000-2022

1. Data breach risk: NLP method on firms' 10Ks from Florakis et al. (RFS, 2023)
2. Innovation: Extended patents from KPSS (2017)
3. Data security innovation: Data security patents based on USPTO classification
4. AI-intensity of firms: We develop ourselves
5. In-house data security protection: We develop ourselves

Explained variables of interest:
Innovation output: citation-weighted counts of patents *filed* by a firm in a year
Financial vars: size (log assets), profitability (ROA)

Methodology:
Poisson regressions, with Fixed effects and lagged cyber risk score
sDiD using the state-level adoption of Data Breach Notification Laws

Results: Higher data risk correlates to more innovation, growth and profits

	Citation-weighted Patent Count		Knowledge and R&D		Financial Vars	
	(1) Overall	(2) Non-CS	(3) Knowledge	(4) R&D	(5) Log assets	(6) ROA
L. Data-risk score	0.243**	0.226*	0.0612	0.122*	0.159**	0.065***
	(0.134)	(0.131)	(0.0563)	(0.0683)	(0.060)	(0.019)
L. Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	12900	14122	15111	21358	20238	20234

One standard-deviation increase in data risk leads to about 7% increase in patents filed; The effect is also observed in the non-data security patents; it leads to a 3% increase in R&D, 3.7% in firm size, and 1% in ROA.

Do AI-intensive firms respond differently to data risk?

Identifying AI-intensive firms

	Citation-weighted Patent Counts			R&D	Financial Vars	
	(1) Overall	(2) Product	(3) Process	(4) R&D	(5) Log assets	(6) ROA
L. Data-risk score$\times(AI = 0)$	0.216	0.132	-0.101	0.0783	0.0798	0.0189
	(0.164)	(0.144)	(0.165)	(0.0888)	(0.0509)	(0.0174)
L. Data-risk score$\times$(AI $=$ 1)	**0.384****	**0.347****	0.161	**0.198***	**0.249*****	**0.0811*****
	(0.174)	**(0.148)**	(0.165)	**(0.0816)**	**(0.070)**	**(0.027)**
L. Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	13375	11497	10786	21358	20238	20234

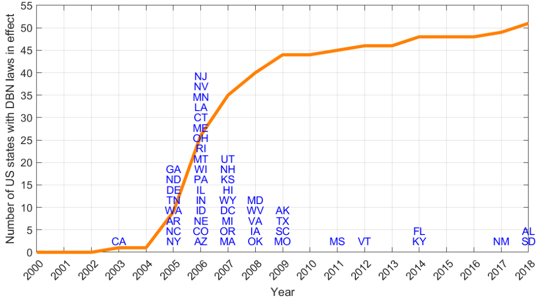
AI-intensive firms drive the results with just 40% of the observations

Addressing endogeneity

- **Limitations of simple regression with lagged data risk:**
  - Lagged variables may not fully account for dynamic endogeneity—where past, present, and future values of data risk and innovation influence each other.
- **Why we need exogenous variation:**
  - To establish a causal relationship by leveraging variation in data risk that is independent of the firm's innovation activities and other confounding factors.
  - An exogenous variation (instrument) provides a clean source of variation in data risk that can be used to isolate its impact on innovation, addressing endogeneity

Data Breach Notification Laws in the USA

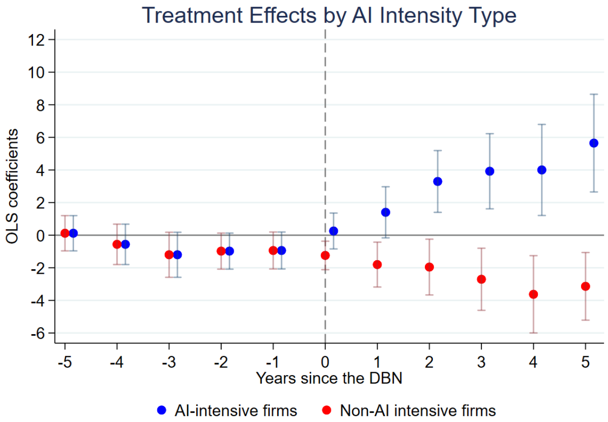
- DBNL mandate firms to notify individuals and state authorities depending on the breach's scale and severity.
- Laws include provisions for penalties for non-compliance, enforcing accountability for data protection.
- All 50 US states have enacted DBNL, in a staggered way. By 2008, over half of the states had adopted DBN law.
- Literature has shown DBN laws led to an increase in firm data risk.



Do data risk and data protection lead to **more overall innovation**?

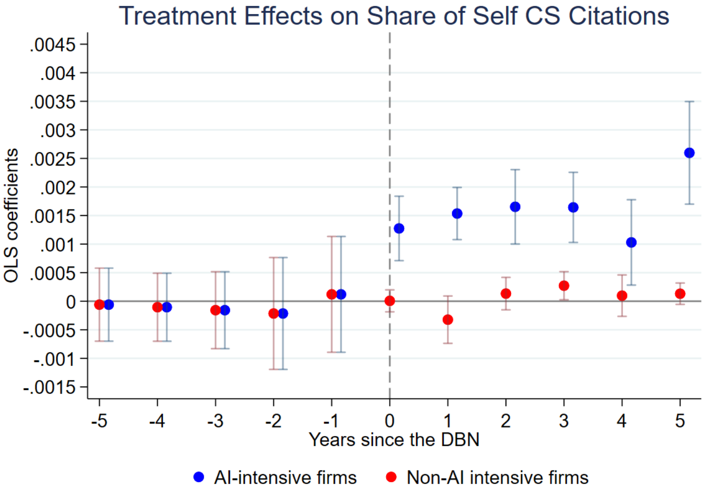
Results on innovation input

Figure: Citation-weighted patent count by data intensity (DI).



Does data risk lead to **more data security innovation**? Identifying CS patents

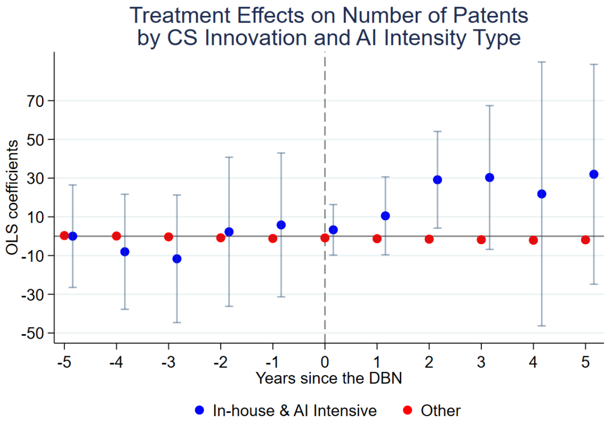
Figure: Share of *self*-data security patent citations by AI intensity (AI).



Do **AI & in-house security** firms respond differently to data risk?

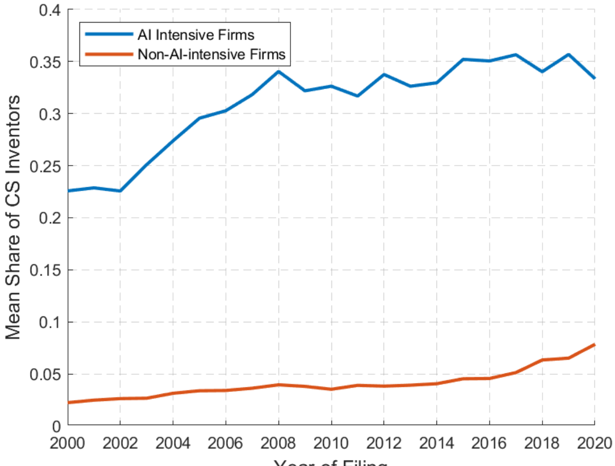
AI-intensive firms

Figure: Citation-weighted patent count, AI intensity interacted with in-house protection



Do **AI** firms have engineers working both on data security and product development? AI-intensive firms

Figure: Inventors common on data security patents and non-data security patents



In which years does data risk create positive externalities? When does data risk have the most **intense effects**?

Figure: Treatment by cohort

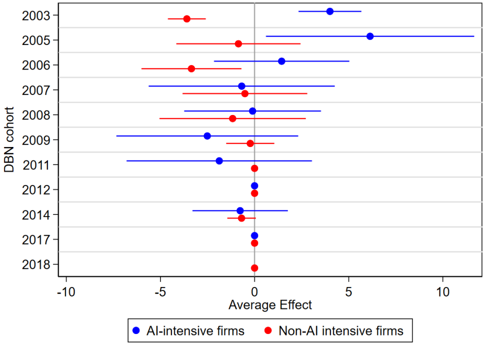
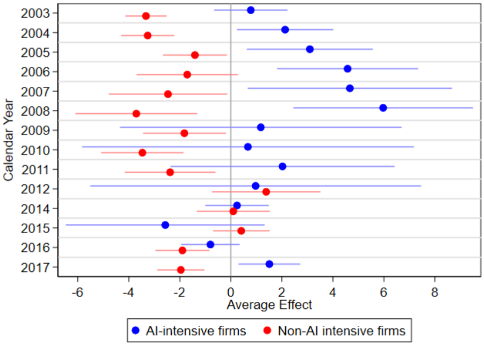


Figure: Treatment by calendar year



How do AI-intensive firms' **financial outcomes** change with data risk?

Figure: Size

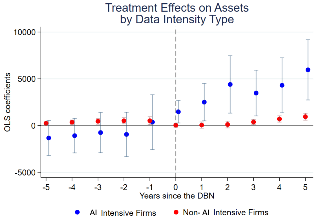


Figure: Profitability

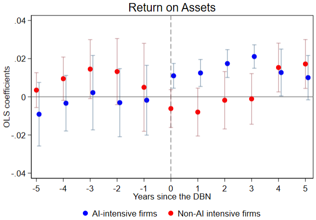
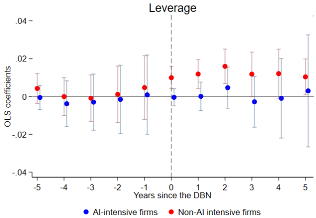


Figure: Leverage



Rationalize findings with a theoretical model

We build **a growth model of the data economy** and perform some comparative statics

- Firms maximize profits
- Data is information extracted from the relation with customers
- Data allows to accumulate knowledge
- Knowledge lowers uncertainty and improves efficiency in production

Cyber risk:

- Threatens data availability and, indirectly, the accumulation of knowledge
- Diverts resources from innovation to damage control

Basic building blocks: heterogeneous firms

Firm heterogeneity:

- Some firms are high-capability and develop security in-house [H-type firms]
- Other firms are low-capability and outsource [L-type]

H-firms invest in cyber security in order to:

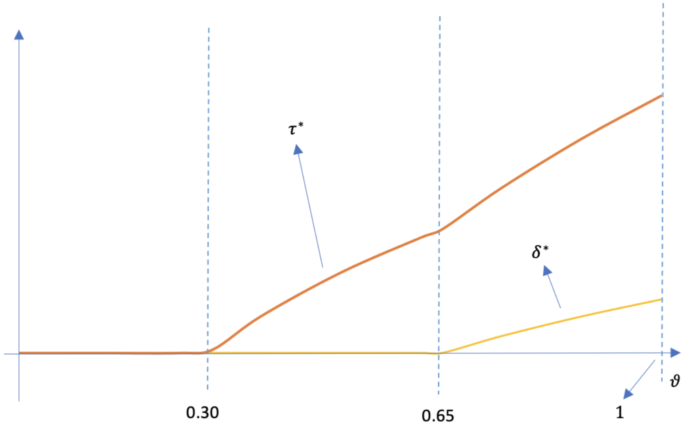
- Lower the impact of cyber risk over the availability of data
- Foster innovation, counteracting the resource diversion effect of cyber risk

L-firms acquire cyber security from H-firms:

- It secures their data and allows them to accumulate knowledge
- But they cannot use the security resources to innovate (they can use the program, but they don't know the code)

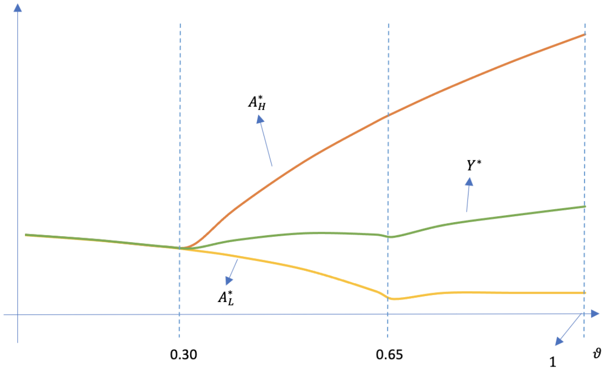
Graphical results (1): Investment in cyber security for different levels of cyber risk

- Two critical thresholds: L-type buy protection only for $\nu > 0.6583$.
- H-type are indifferent between investing in protection or not at a critical threshold level of $\nu = 0.3$.



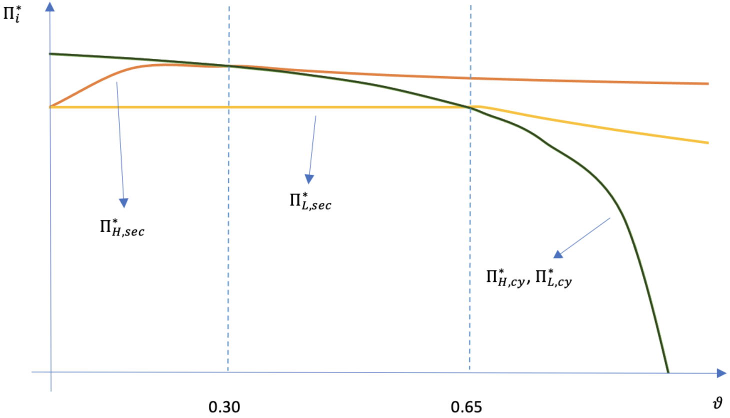
Graphical results (2): Output for different levels of cyber risk

- H-type (orange) use protection to innovate, ↑ quality & quantity of production.
- L-type do not have this positive spillover; they use security only for protection.
- The evolution of Y^* gains momentum when L-type start protecting as well.



Graphical results (3): Profits for different levels of cyber risk

- Without protection, the profits (green) of H-type equal profits of L-type's.
- With protection, profits of H-type (orange) always higher than L-type's (yellow).
- As data risk increases, the profits of H-type decrease by less than L-type's.



Conclusion: Necessity is the Mother of Invention

- For a small subset of AI-intensive firms: innovations in digital protection spill over to overall product and service innovation (firms thrive amid cyber risk)
- For the majority of companies: cyber threats are disruptive, but negative effects are mitigated through security outsourcing
- The way forward: recognize the role of high-capability firms as guardians of cyber security and drivers of innovation & support SMEs accessibility to cyber innovation and cyber protection

Thank You!

Appendix

Data risk and innovation input

back

	R&D Assets	Knowledge Assets
	(1)	(2)
L.Data risk	0.116*	0.0753
	(0.0657)	(0.0541)
Size + other controls	Yes	Yes
Firm FE	Yes	Yes
Year FE	Yes	Yes
N	15038	14921

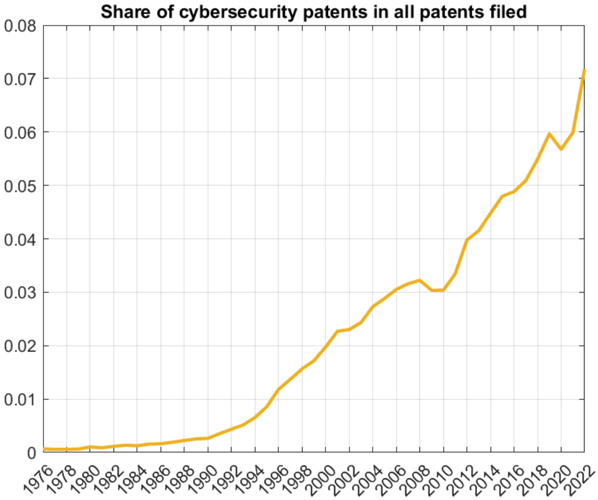
R&D assets ↑ by about 3% following one-SD ↑ in data risk

Data security patents

back CS back overall

Identifying CS patent based on USPTO Cooperative Patent Classification (CPC) codes. Example classification codes:

- G06F 21/ : "Security arrangements for protecting computers, components thereof, programs or data against unauthorised activity"
- H04L 9/00 "arrangements for secret or secure communications; Network security protocols."

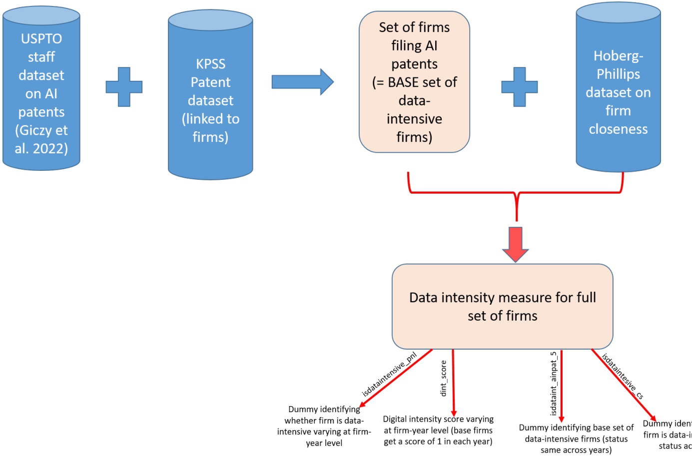


Identifying AI-intensive firms

back

The principle behind constructing set of data intensive firms:

- Firms active in AI innovation must be data intensive
- $\Rightarrow$ Firms filing AI patents are data intensive ("base set of data-intensive firms")
- Firms that describe their business operations in similar words as the base set of data-intensive firms are also data intensive



Identifying in-house CS firms

back

The principle behind constructing set of firms with in-house data security protection:

- Examine backward citations of the public firms' patents (from the USPTO).
- Backward citations refer to the citations a patent makes to preceding patents, which serve as references or foundational works for the current patent.
- We ascertain whether the patent they cite is
  1. a data security patent and
  2. belongs to the firm itself
- Firms that cite their *own* data security patents in any of its patents are classified as in-house data security firms.
- For robustness, we also look at firms that cite their *own* data security patents in any of its non-data security patents are classified as narrower in-house data security firms.