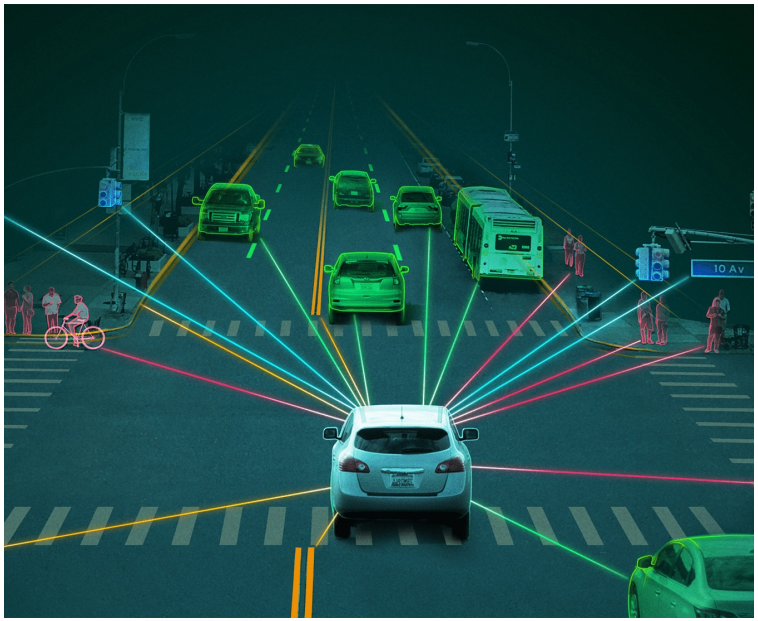
Data Sharing and Market Protection

Linsheng Zhuang (speaker)
with
Jussi Keppo and **Zhi Chen**

linsheng.z@u.nus.edu
NUS Institute of Operations Research and Analytics

Data-Driven New Products

The development of data-driven new products requires huge amount of training data



Firms invest heavily in data collection in different markets

"[Tesla Autopilot] does not work quite as well in China as it does in the U.S. because most of our engineering is in the U.S."

– Elon Musk

Over 1,400 self-driving vehicles are now in testing by 80+ companies across the US

Darrell Etherington @etherington / 11:54 PM GMT+8 • June 11, 2019

Comment

Image Credits: Ford

Audi, Mercedes and BMW conduct trials despite having to partner with Beijing groups

Data Protection Regulation

Restrict the type and granularity of data that can be transferred beyond the borders

China clamps down on auto data collection by Tesla and others

New rules would ban overseas transfer of road and landscape information

Consequence of Data Protection Regulation

"Normally, we would like to make effective use of data we collect around the world in order to accelerate the development of autonomous driving technology, but that will become difficult. We have no choice but to utilize data collected in China for only China."

AUTOMOBILES

China data rules to squeeze overseas development of self-driving tech

Information generated in the country will mostly stay within the country

Tesla vehicles at the Zhongnanhai leadership compound during a meeting between Tesla CEO Elon Musk and Chinese Premier Li Keqiang in Beijing in 2019. Tesla has been suspected of suspected of transferring vehicle data from China to the U.S © Reuters

SHUNSUKE TABETA, Nikkei staff writer
August 22, 2021 01:26 JST

Research Questions

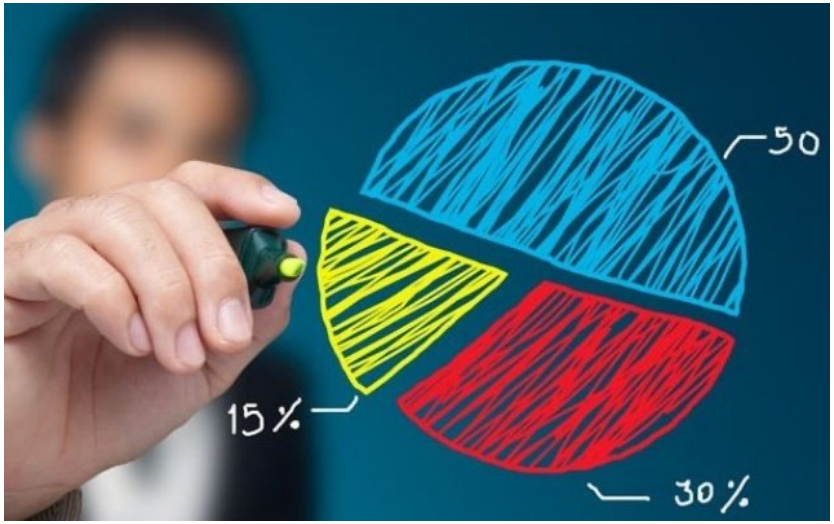
For multinational firms operating in multiple markets

- What are the effects of these data protection regulations on their data collection investments?

- From the government's perspective, how restrictive should be the regulation in order to balance between innovation and national security?

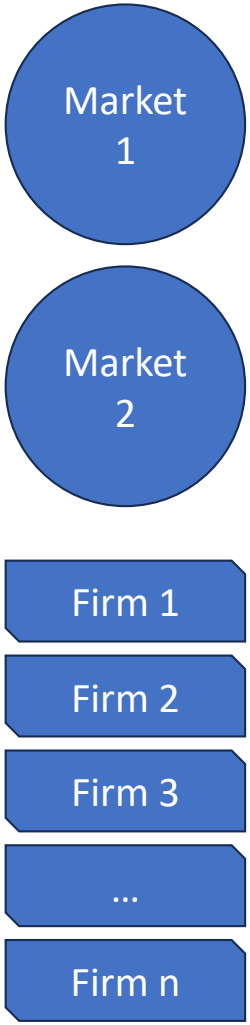
Modelling Framework: Contests/Tournaments

Multiple firms expend efforts at their own cost to compete for an award.



Literature Review

- Design of Innovation Contests
  - Prize distribution (Moldovanu and Sela 2001, Terwiesch and Xu 2008, Ales et al. 2017)
  - Number of participants (Taylor 1995, Che and Gale 2003, Korpeoglu and Cho 2017)
  - Contest duration (Korpeoglu et al. 2021)
  - Problem specifications (Erat and Krishnan 2012, Jiang et al. 2021b)
  - ……
- Economics of Data
  - Consumer's privacy (Fainmesser et al. 2021, Acemoglu et al. 2022)
  - The usage of new (AI) technologies (Jones and Tonetti 2020, Farboodi and Veldkamp 2021, Acemoglu et al. 2022)
  - Smart city and AV industry (Hasija et al. 2020, Mak 2020, Baron et al. 2021)
  - ……



Market sizes M_1, M_2 is publicly known.

The government of each market announces the permissible level of data sharing $\theta_j \in [0,1], j = 1,2,$ with the objective that strikes a balance between innovation and cost of data sharing (e.g., national security).

Firms enter each market and collect data $d_{i,1}, d_{i,2}, i = 1,2, \dots, n$ at a cost of $c(d)$, which is strongly convex.

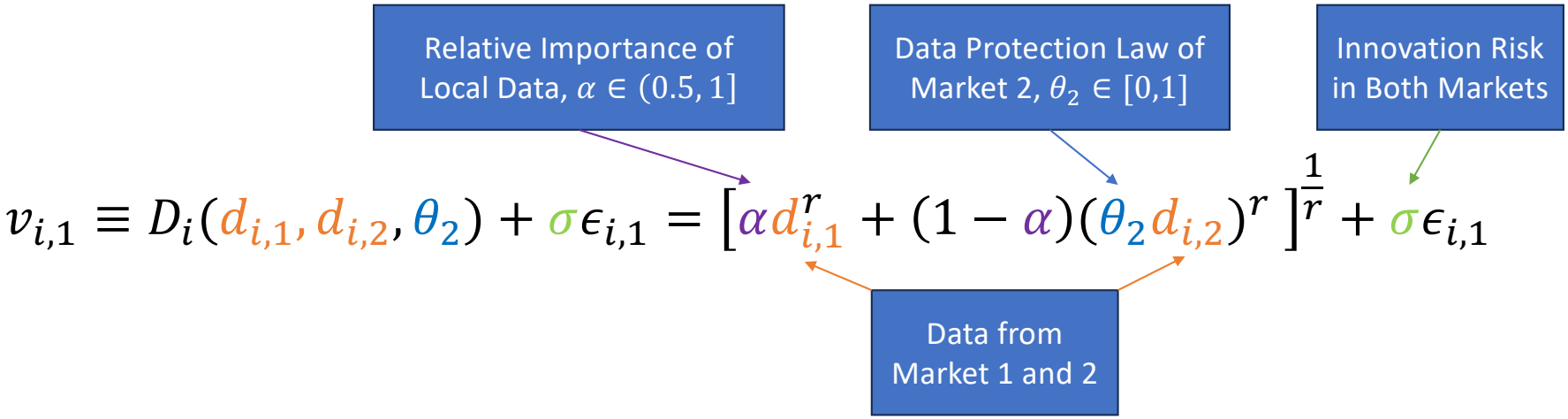
The performance of product is modelled by a CES function plus the innovation shock

$$v_{i,1} = \left[\alpha d_{i,1}^r + (1-\alpha)\left(\theta_2 d_{i,2}\right)^r\right]^{1/r} + \sigma\epsilon_{i,1}$$

In each market, there is a contest where the winner takes all market share.

The Product Performance

- The training performance of product is a CES function



- $r > 0$ measures the relationship of two datasets
  - If $r \to 0$, $E(v_{i,1}) \to (d_{i,1})^{\alpha}(\theta_2 d_{i,2})^{1-\alpha}$, which is Cobb-Douglas function.
  - If $r \to \infty$, $E(v_{i,1}) \to \max\{d_{i,1}, \theta_2 d_{i,2}\}$.
  - We say the datasets are complement if $r \in (0,1)$, and substitute if $r > 1$.

The Equilibrium

$E[v_1^\star]$ is the expected performance of market 1 product

$\psi(\cdot)$ is the privacy cost

- Government's Choices:

$$B_1(\theta_2) = \max_{\theta_1 \in [0,1]} E[v_1^\star] - \psi(\theta_1 d_1^\star)$$

$$B_2(\theta_1) = \max_{\theta_2 \in [0,1]} E[v_2^\star] - \psi(\theta_2 d_2^\star)$$

- Representative firm i's Choices:

$$\max_{d_{i,1} \geq 0, d_{i,2} \geq 0} M_1 \Pr\left(v_{i,1} > v_{k,1}, \forall k \neq i\right) - c\left(d_{i,1}\right) + M_2 \Pr\left(v_{i,2} > v_{k,2}, \forall k \neq i\right) - c(d_{i,2})$$

- First solve the symmetric subgame Nash equilibrium (d_1^*, d_2^*), then the equilibrium policies $(\theta_1^\star, \theta_2^\star)$.

Research Questions

For multinational firms operating in multiple markets

- **What are the effects of these data protection regulations on their data collection investments?**

- From the government's perspective, how restrictive should be the regulation be in order to balance between innovation and national security?

Effect of Data Protection Regulations

- **Result 1**: The equilibrium amount of data collection d_1^* always rises in θ_1, the permissible level of data sharing in the local market.

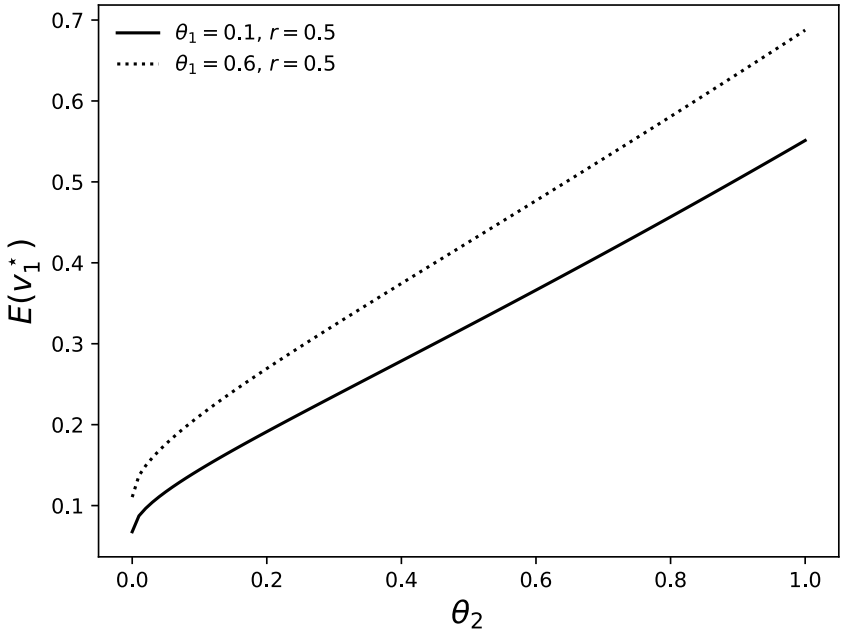
- **Result 2**: The equilibrium amount of data collection d_1^* rises in θ_2 if the data substitution rate $r < 1$ and falls in θ_2 if $r > 1$.

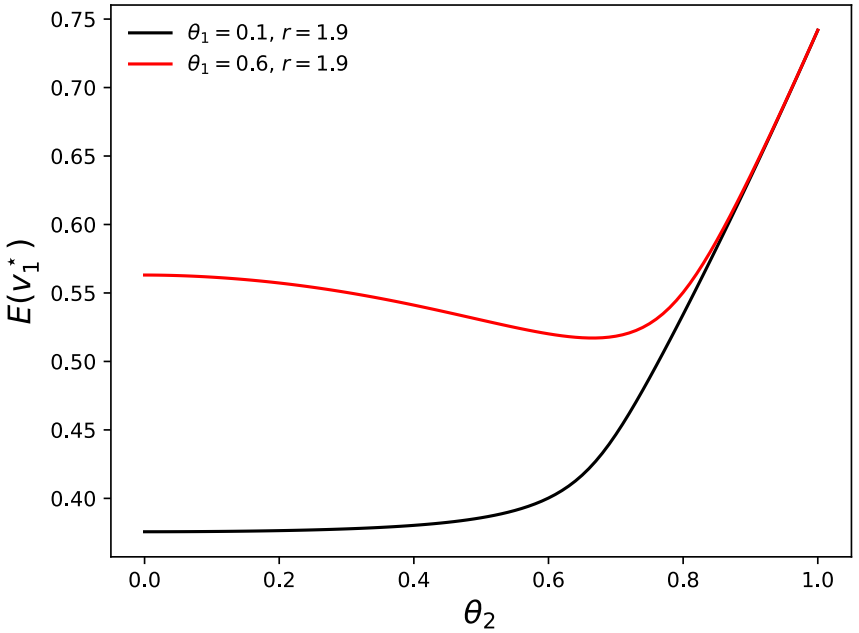
Effect of Data Protection Regulations

- **Result 3**: The expected product performance $E(v_1)$ in equilibrium always rises in θ_1, permissible level of data sharing in the local market.

- **Result 4 (numerical)**: The expected product performance $E(v_1)$ in equilibrium
  1. First falls then rises in θ_2 if r is large, θ_1 is large and M_1 and M_2 are similar.
  2. Rises in θ_2 in other cases.

Effect of Data Protection Regulations





Effect of Data Protection Regulations

- Why do we observe the falling part in the red line?
  - If $r \to \infty$, the CES function approaches a max function

$$E(v_1^\star) \to \max\{d_1^\star, \theta_2 d_2^\star\}$$

  - If θ_1 is large, then $d_1^\star$ is large and active in the max function when θ_2 is small. Remember that data sets are redundant, $d_1^\star$ falls in θ_2.
  - As θ_2 grows, the term $\theta_2 d_2^\star$ rises and becomes active.

Research Questions

For multinational firms operating in multiple markets

- What are the effects of these data protection regulations on their data collection investments?

- **From the government's perspective, how restrictive should be the regulation in order to balance between innovation and national security?**

Government's Choice

- Payoff of the two markets are

$$B_1(\theta_2) = \max_{\theta_1 \in [0,1]} E(v_1^\star) - \psi(\theta_1 d_1^\star)$$
$$B_2(\theta_1) = \max_{\theta_2 \in [0,1]} E(v_2^\star) - \psi(\theta_2 d_2^\star)$$

Government's Choice: the Equilibrium

- **Result 5 (numerical)**: The equilibrium policy $\theta_1^\star$ always falls in M_1, the local market size.

Conclusions

- The more open the markets is, the higher the firms' data collection investment in the market, and the higher performance the AI products in the market.
- The effect of a country's data protection regulation on another country's AI investment depends on the relationship of the datasets from the two markets.
- Bigger market tends to be more protective.

Thank you!