# Heterogeneous Consumers and the Data Value Chain: An Agent-based Approach

Zhongli Wang[1]

Bielefeld University and the Catholic University of the Sacred Heart

January 9, 2024

1. Ph.D. student of the program EPOC. Many thanks for guidance of Prof. Herbert Dawid, Prof. Domenico Delli Gatti and Prof. Christiane Fuchs.




This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 956107, "Economic Policy in Complex Environments (EPOC)".

# Research Background: Why do we focus on the data value chain?

- Huge growth of investments in AI companies. (UNCTAD 2021)

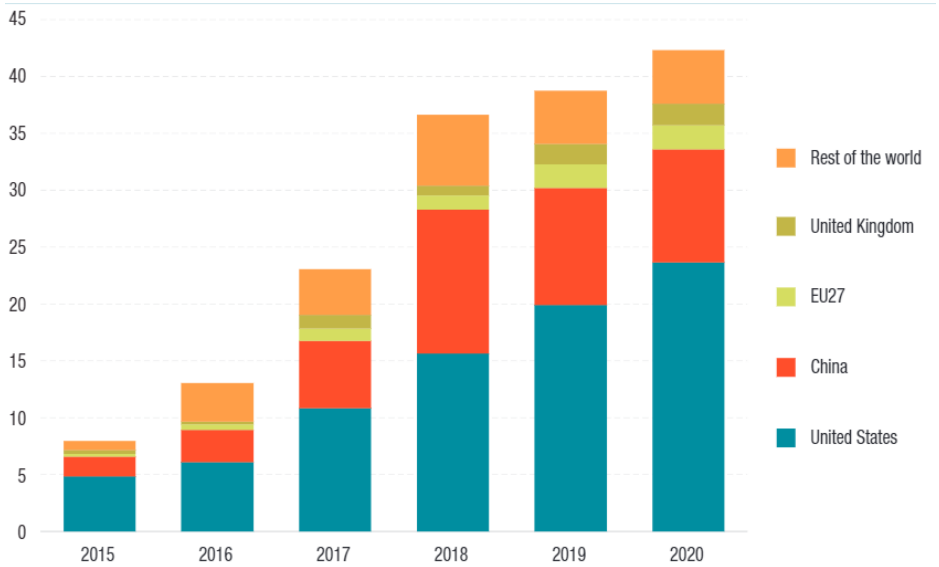



Figure 1: Private investment in AI companies, by economy, 2015–2020 (Billions of dollars). *Source: UNCTAD calculations, based on the publicly available database of NetBase Quid – 2021 AI Index Report (Zhang et al., 2021), available at https://aiindex.stanford.edu/report/ (accessed April 2021).*

- A value chain divides a firm into the discrete activities in designing, producing, marketing, and distributing its product. (Porter 1985)

# Research Background: Why do we focus on the data value chain?

- From raw data to software
- Example of visualization of a data value chain (UNCTAD 2021)


Figure 2: Stages of global data value chain

- Example of visualization of a software value chain (Pussep et al. 2011)

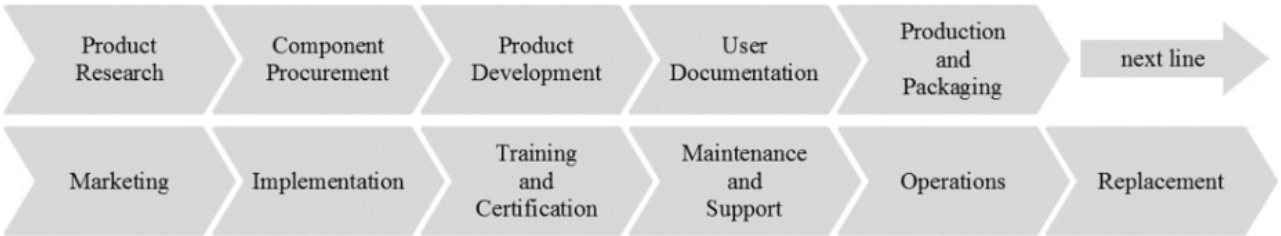



Figure 3: Eleven activities in a unified software value chain

# Research Background: Why do we focus on the data value chain?

- Two phenomena in the software market
  - Not all applications can be purchased and used in a certain geographical region
  - There are some applications which serve for the same purpose differ in prices

# Research Background: Why do we focus on the data value chain?

- One possible factor: the local data restriction
- Policy debate on regulations: (a) Facilitation of free data flows, or (b) domestic localization of data storage to achieve various national objectives? (UNCTAD 2021)
- To enhance the local economic development, governments of some regions restrict the cross-border data flows. For instance, data localisation policies require data pertaining to citizens of a country to be processed and/ or stored within its jurisdiction; moreover, inflexible mandates can completely restrict the flow of data outside the country.(Kathuria et al. 2019)
- Different attitudes towards local data privacy policy influence consumer behavior. (Larisa Serzo 2020)

# Research Background: Why do we focus on the data value chain?

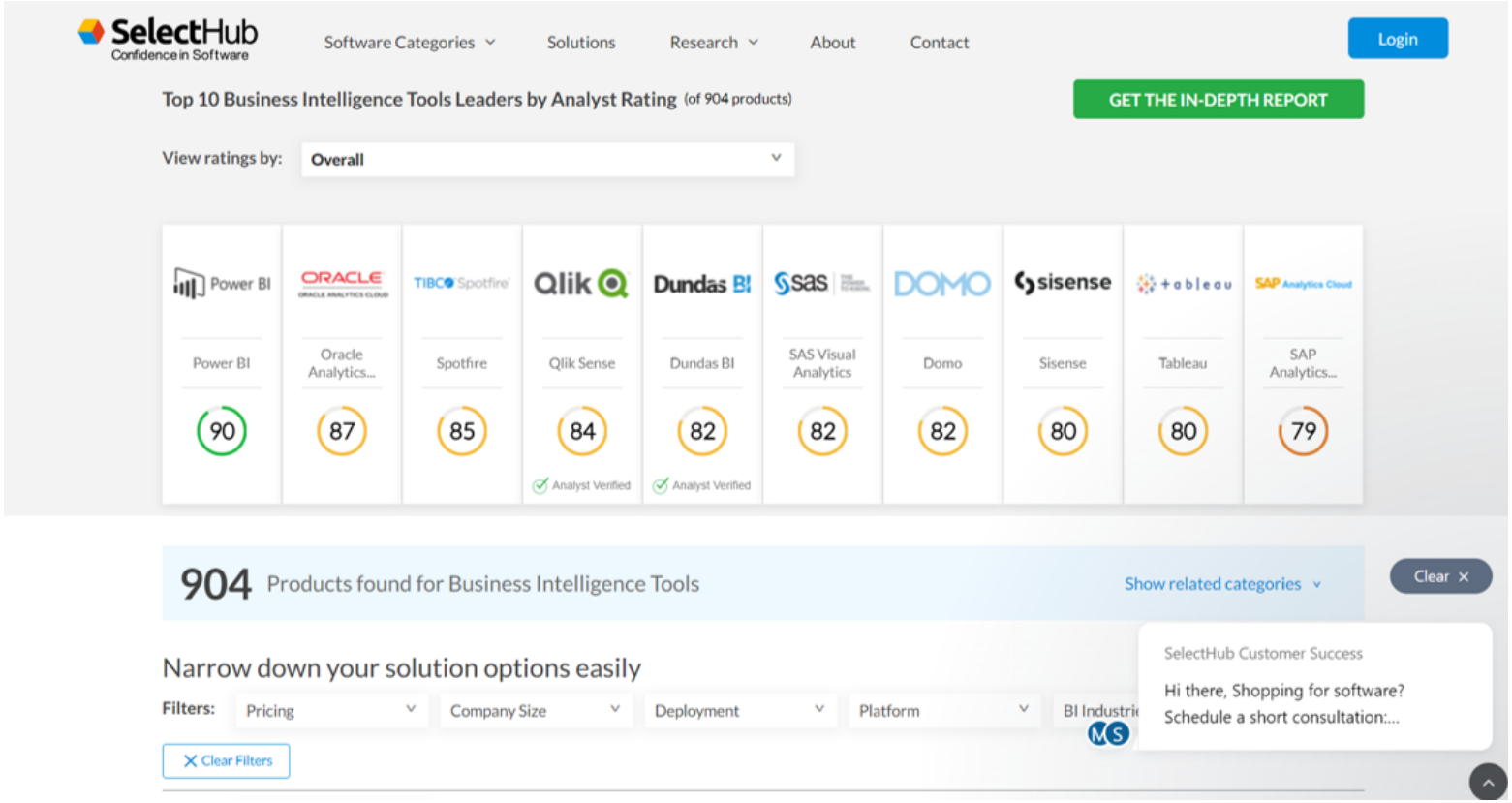



Figure 4: Top 10 Business Intelligence Tools Leaders by Analyst Rating of 904 products. Retrieved 15.11.2023. from https://www.selecthub.com/c/business-intelligence-tools/.

# Research Background: Why do we focus on the data value chain?

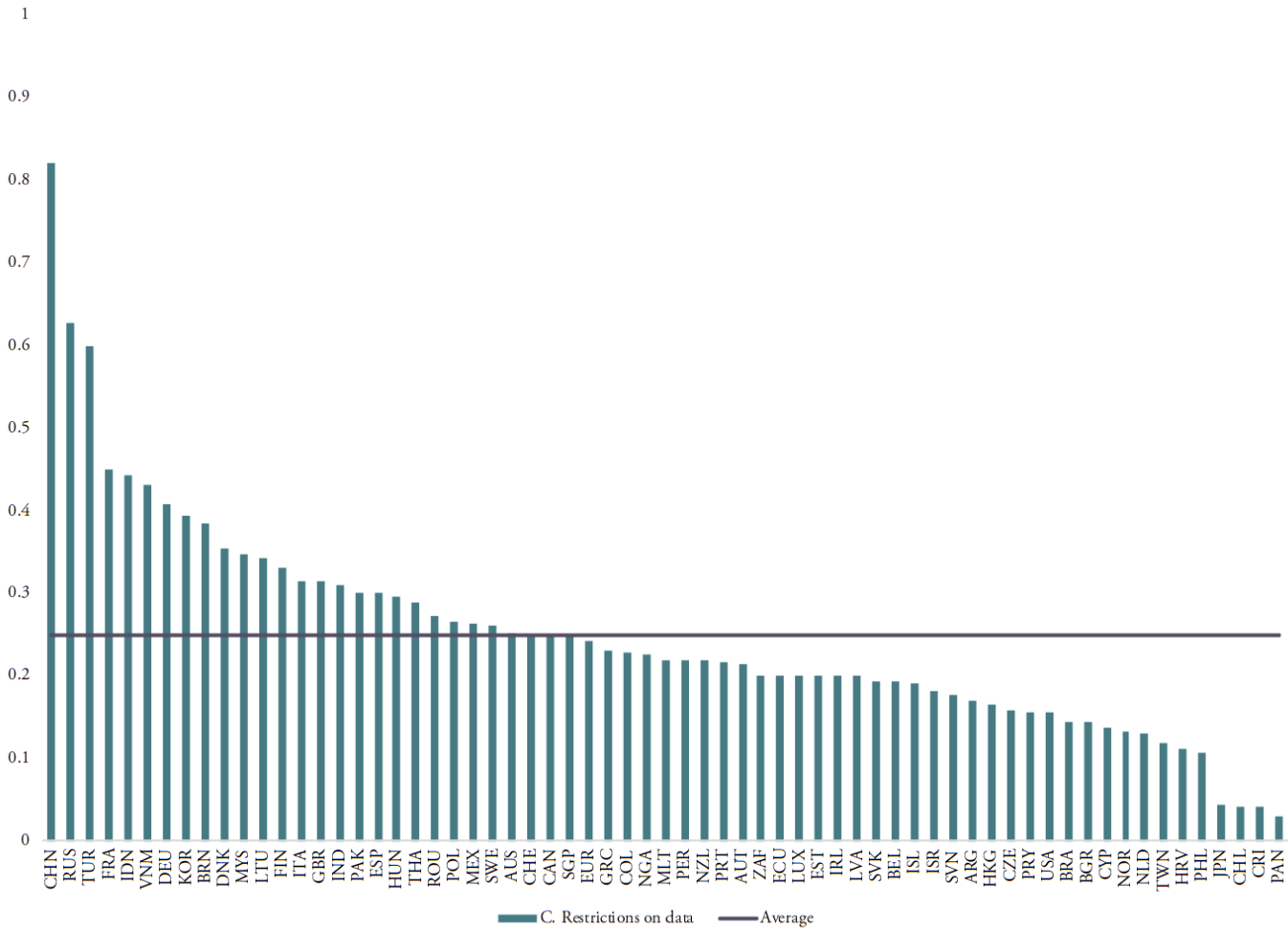



Figure 5: Restrictions on data of the Digital Trade Restrictiveness Index. (Ferracane, Lee-Makiyama, and Van der Marel (2018))

# Research Questions: What do we expect to know?

- How do the local data restrictions make a difference in the spatial distribution of economic activities in one period and in multiple periods?
- What is the optimal policy regarding the data restrictions considering the global welfare and the local economic profit?
- What are the strategies for the firms to make profits across periods?

# Strands of literature and methodology: What have been done in previous research?

| | |
|---|---|
| Global value chain | Rose and Liao (2005), Koks and Thissen (2016), Carvalho et al. (2020), Guan et al. (2020), Dawid, Harting, and Hepp (2021), Burfisher (2020), Kano, Tsang, and Yeung (2020) |
| Digitalization | Goldfarb and Tucker (2017), Connell et al. (2018), Schlüter and Hetterscheid (2017), Dawid and Muehlheusser (2022) |
| Industrial Organization and spatial models | Shy (1996), Hotelling (1929), Gabszewicz and Thisse (1982), Graitson (1982), Tabuchi and Thisse (1995), Larralde, Jensen, and Edwards (2006), Larralde, Stehlé, and Jensen (2009), Guo, Lai, and Zeng (2015), Peng et al. (2020) |
| Cross-border data flow | Aaronson (2018), GSM Association (2018), Nguyen and Paczos (2020), Azmeh, Foster, and Rabuh (2021), Casalini and López-González (2019), Bauer, Ferracane, and Marel (2016), Castro and Mcquinn (2015), Pussep et al. (2011) |

# Strands of literature and methodology: What have been done in previous research?

| | |
|---|---|
| Structure of data value chain | Pussep et al. (2011), UNCTAD (2021) |
| Modeling | Ranking utility (Gabszewicz and Thisse (1979)), the OECD ICIO (Inter-Country Input-Output) Model (De Backer and Miroudot (2013)), digital platforms (Asian Development Bank et al. (2021)), outsourcing (Chalaby (2019)), STRI (Services Trade Restrictiveness Index) and gravity models (Frontier Economics Ltd. (2021)) |
| Heuristics | Tversky and Kahneman (1974), Dawid and Delli Gatti (2018), Nagle and Müller (2018), Dawid et al. (2019) |

# Model: How do we design the model?

- Properties of a data value chain
  - Focusing on *Price* instead of *Quantity*: One software on one device
  - Specific structure of development and outsourcing: An Owner and a Developer are connected with a contract
  - Free of transportation costs

# Model: How do we design the model?

- Model Architecture

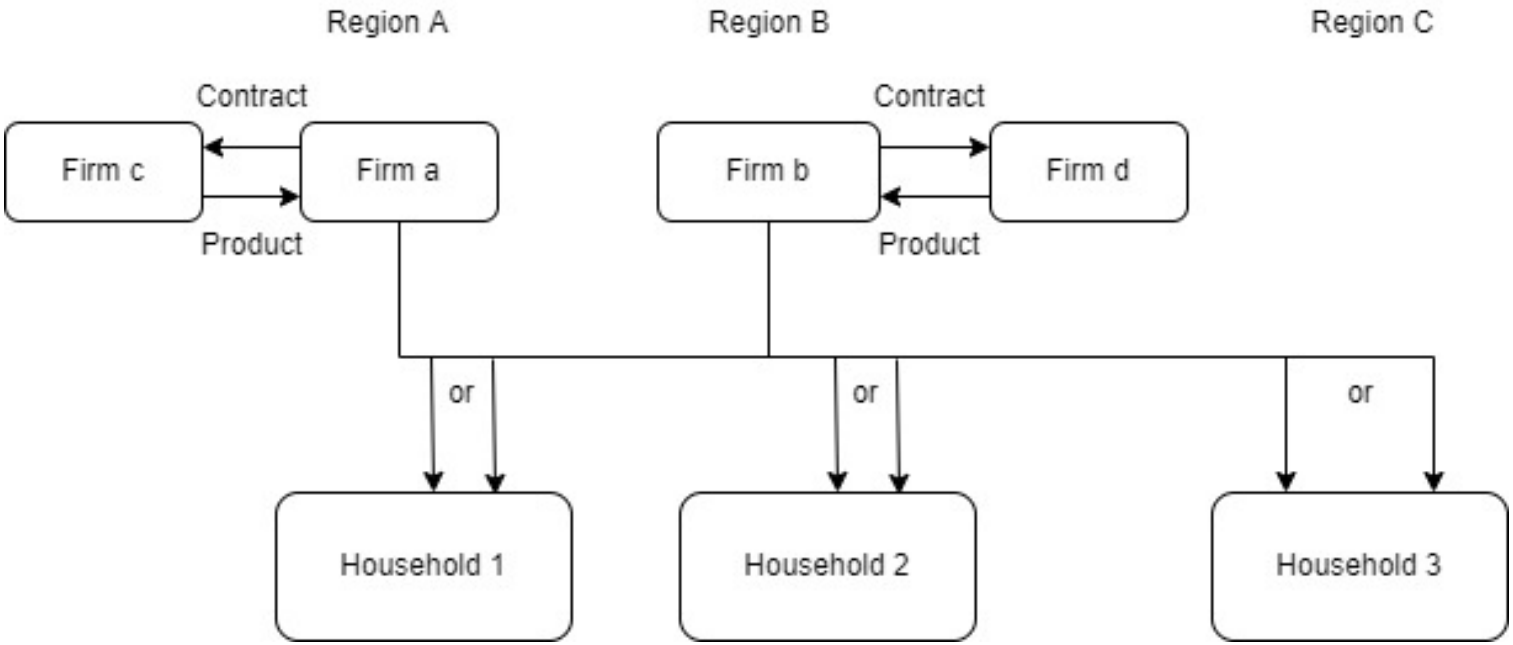



Figure 6: Product selection. Note: Consumers can be in any region and we focus on profits of Firm *a* and Firm *b*.

# Model: How do we design the model?

We assume that the following events takes place in one period:

1. Owners sign contracts with Developers in which they agree on the payment for Developers.
2. During the production of the data product, Developers design the algorithms and devote their intellectual labor to the completion of the product.
3. The Owners pay the Developers for the data products and the Regulators for the cost of data restriction (all of them in the same region). The Owners sell the products in the market (via online platforms) across the borders.
4. Consumers compare the products of different Owners and consider (i) whether to purchase the product. If they decide to purchase a data product, they consider (ii) which product to purchase.
5. Transactions are carried out. The profits of each Owner can be calculated at the end of this period (when all Consumers have made their decisions).

# Model: How do we design the model?

- Model design
- The utility of Consumer $j$ ($j = 1, ..., n$) towards software of Owner $i$:

$$u_{j,i} = \bar{u}_i + x_j\theta_i + y_j c_{res,i} - p_i$$

where $\bar{u}_i$ is a constant (which can be interpreted as the Consumer i's impression towards the Owner company) and $\theta_i$ is the quality of the software. We assume that both Owners have the same reputation and Consumers don't have a preference for the Owners, i.e. $\bar{u}_a = \bar{u}_b$.

# Model: How do we design the model?

- The maximization problem of Owner $a$:

$$\max_{p_i, c_{0,i}} \Pi_i = m_i p_i - c_{total,i}$$

$$s.t.\ \theta_i = \bar{\theta}_i + f(c_{0,i})$$

$$c_{total,i} = f(c_{0,i}) + m_i c_{res,i}$$

$$m_i = \sum_{j=1}^{n} 1_{u_{j,i} > u_{j,-i}} + \sum_{j=1}^{n} \beta 1_{u_{j,i} = u_{j,-i}}$$

$$\beta \sim Bin(1, 0.5)$$

- The global welfare:

$$W = \sum_{j=1}^{n} u_{j,i} 1_{u_{j,i} \geq u_{j,-i}} + \sum_{j=1}^{n} u_{j,-i} 1_{u_{j,i} < u_{j,-i}} - c_{total,a} - c_{total,b}$$

# Static model design

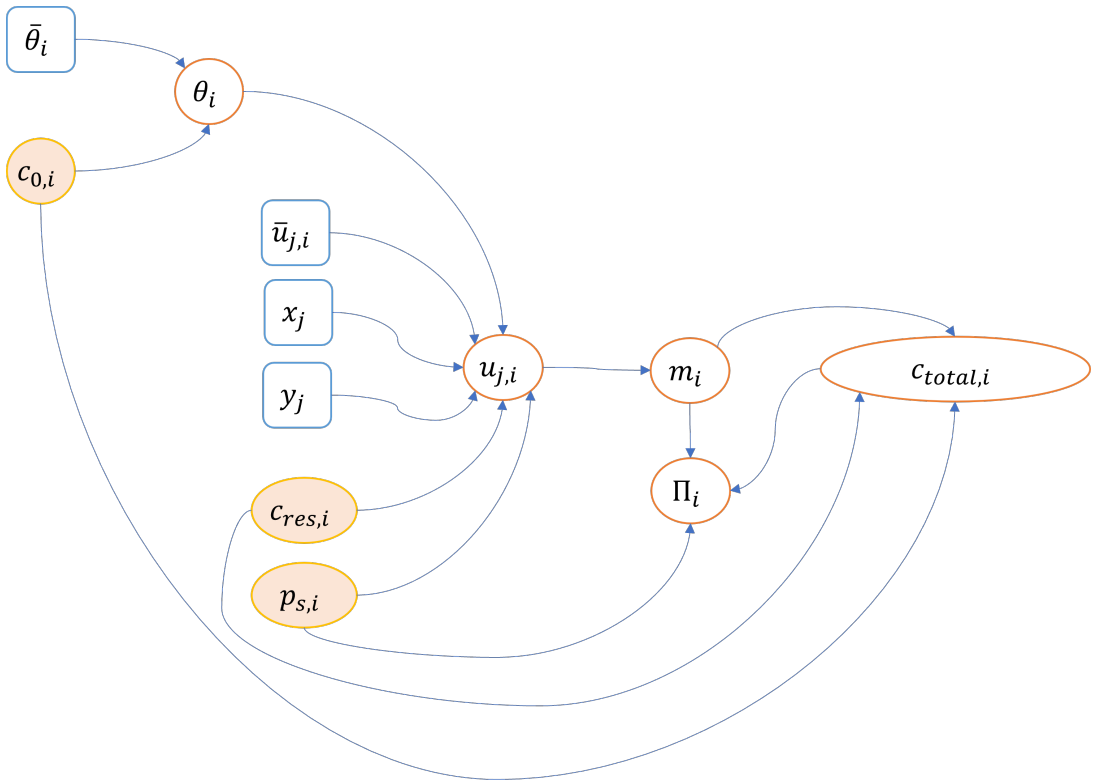



Figure 7: Main variables covered in the model

# Static model design

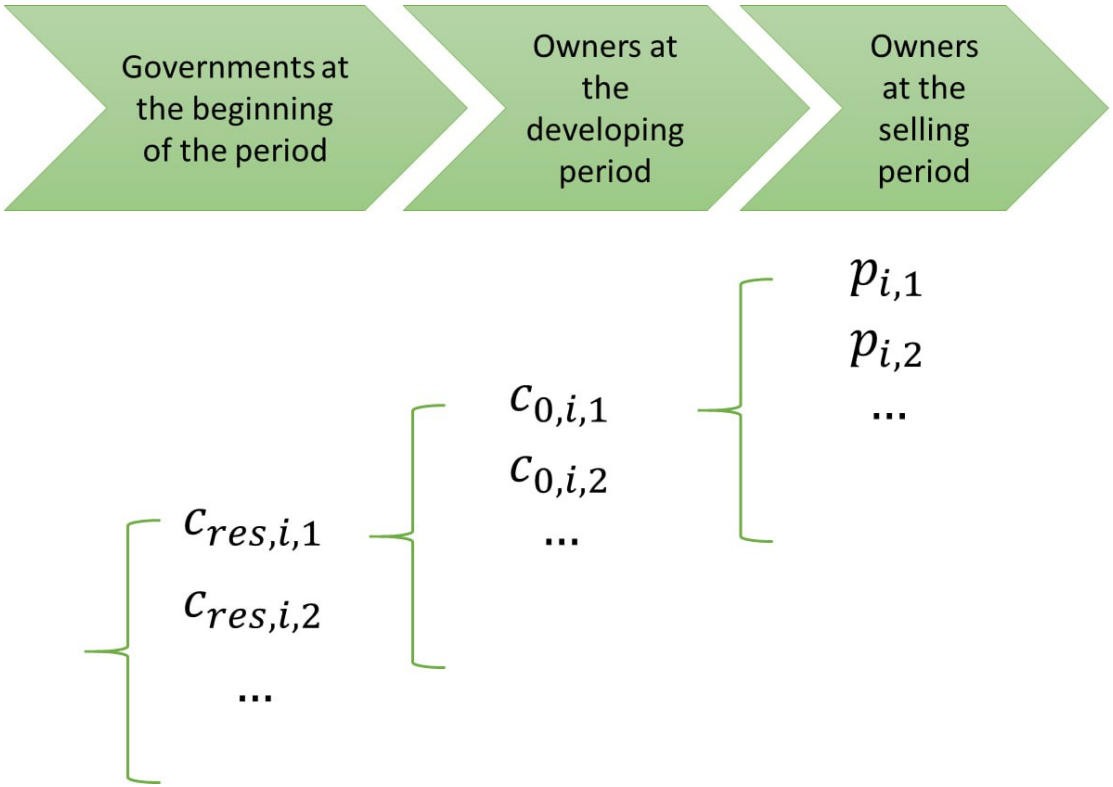



Figure 8: Visualization of the game

# Dynamics with heuristics

Steps of the market analysis (Nagle and Müller (2018) , Dawid et al. (2019)):

1. Within one period, Owners draw representative samples of Consumers and ask them if they will buy the their own products under a variety of prices.
2. Consumers in the sample respond to the surveys.
3. Owners collect the data, analyze the surveys, calculate the expectation of the prices for the next period and set the prices for the next period.

# Dynamics with heuristics

Heuristics with investment updates:

1. At the beginning of the $\tau$-th period ($\tau > 1$), we assume that both Owners increase their investment by $\Delta$ units.
2. At the beginning of the $2\tau$-th period, each Owner calculates its average profit over the last five months. We define the amount of each Owner as $\overline{\Pi_{i,2\tau}}$ ($i \in \{a, b\}$). Owner $i$ compares $\overline{\Pi_{i,\tau}}$ and $\overline{\Pi_{i,2\tau}}$. If $\frac{\overline{\Pi_{i,2\tau}} - \overline{\Pi_{i,\tau}}}{|\overline{\Pi_{i,\tau}}|} > \vartheta$ ($\vartheta$ is the updating threshold), Owner $i$ continues the same direction of investment by $\Delta$ units; If $\frac{\overline{\Pi_{i,2\tau}} - \overline{\Pi_{i,\tau}}}{|\overline{\Pi_{i,\tau}}|} < -\vartheta$ Owner $i$ adjusts investment reversely by $\Delta$ units; Otherwise, Owner $i$ does not change its investment.
3. At the beginning of the $k\tau$-th period ($k > 2$), each Owner compares $\overline{\Pi_{i,k\tau}}$ and $\overline{\Pi_{i,(k-1)\tau}}$ and adjusts the investment in the same way. We note that if the investment holds constant continuously twice, at a probability of 0.5, the Owner randomly increases or decreases the investment by $\psi$ units at the beginning of the next $\tau$-th period ($\psi$ is drawn from $\{-\psi_0, \psi_0\}(\psi_0 > 0)$ with probability 50%).

# Analytical process and simulations: What does the economy present to us?

Suppose regulations and product functionalities differ.

$$\bar{u}_a + x_j\theta_a + y_j c_{res,a} - p_a \geq \bar{u}_b + x_j\theta_b + y_j c_{res,b} - p_b$$

$$\Leftrightarrow$$

$$x_j\theta_a + y_j c_{res,a} - p_a \geq x_j\theta_b + y_j c_{res,b} - p_b$$

$$\Leftrightarrow$$

$$x_j(\theta_a - \theta_b) + y_j(c_{res,a} - c_{res,b}) \geq p_a - p_b$$

$$\Leftrightarrow$$

$$x_j\Delta\theta + y_j\Delta c_{res,R} \geq \Delta p_s$$

# Analytical process and simulations: What does the economy present to us?

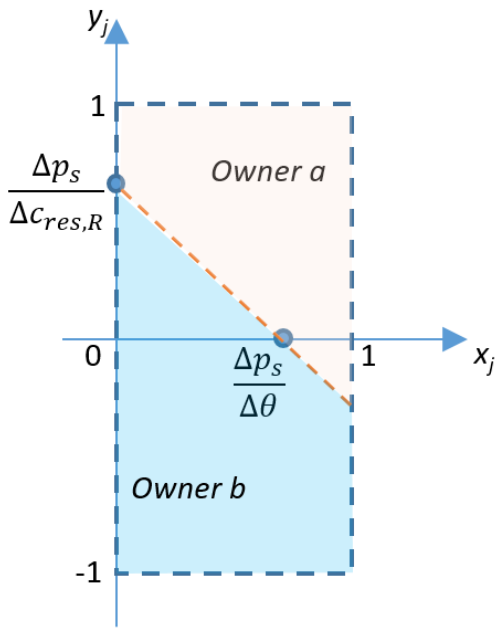



Figure 9: Critical points and the types of Consumers

This is coordinated with our assumption: Products that are better in functionality, less regulated and less expensive are preferred by Consumers.

# Multi-stage game

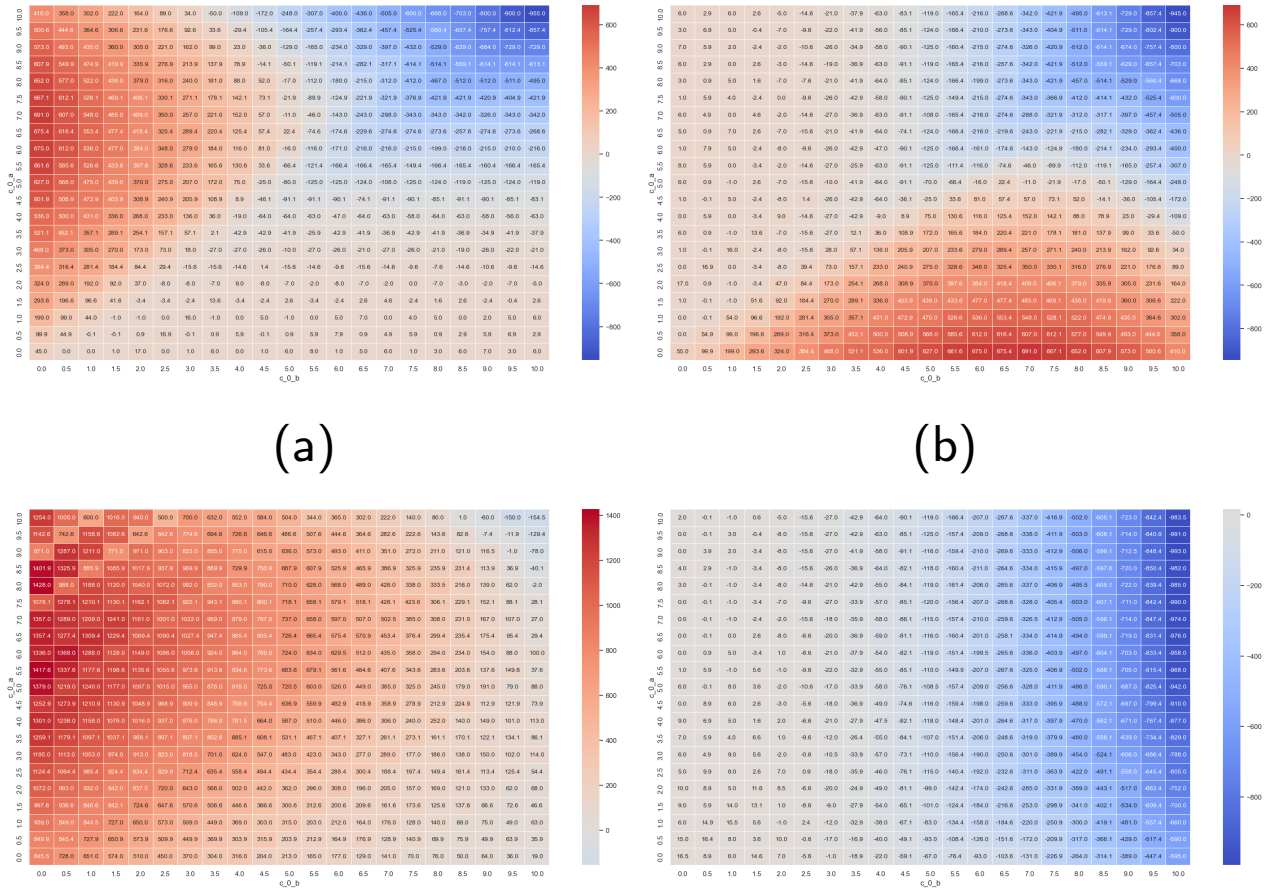
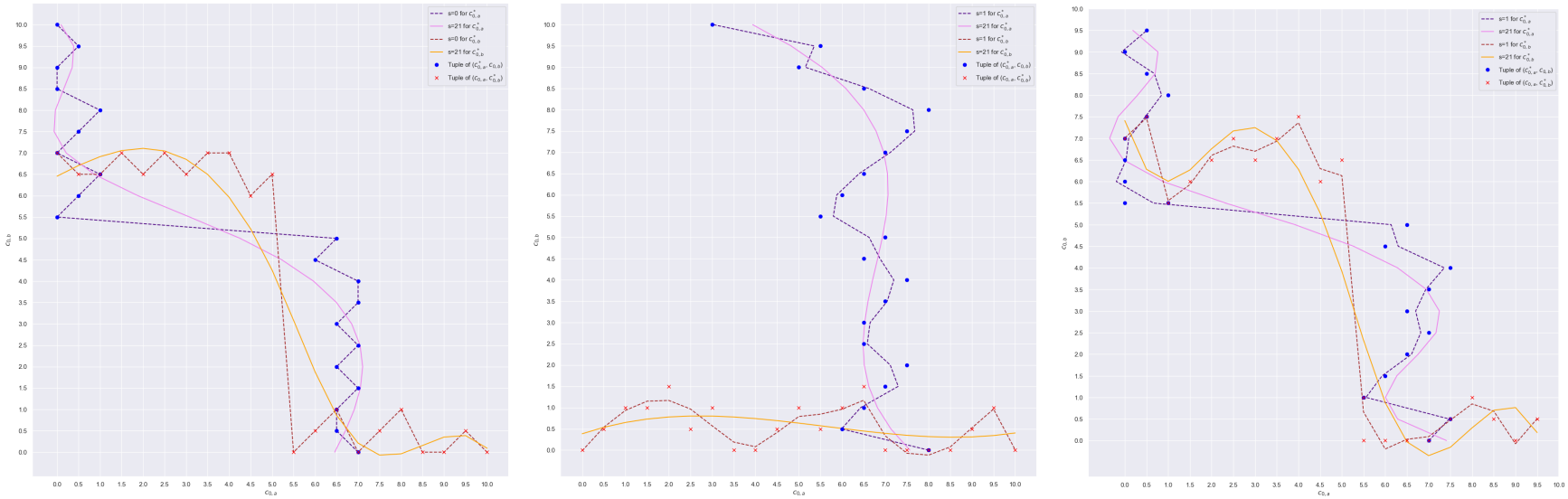



Figure 10: The profits of Owners based on equilibrium prices of Owner $a$ and Owner $b$ under different investment combinations. Grids in Sub-figure (a) - (d) show $\Pi_a^*$ and $\Pi_b^*$ in which $(c_{res,a}, c_{res,b})$ equals (0,0) and (0,10) $(c_{total,i} = c_{0,i}^{3} + m_i c_{res,i})$.

# Multi-stage game



(a) (b) (c)

Figure 11: The best responses of Owners under different investment combinations. Sub-figure (a) - (c) show $p_a^*$ and $p_b^*$ in which $(c_{res,a}, c_{res,b})$ equals (0,0), (0,10) and (10,10), respectively ($c_{total,i} = c_{\mathbf{0},i}^{\mathbf{3}} + m_i c_{res,i}$).

# Results of the multi-stage game

Each local government should have a low data restriction level considering the profit of the local Owner.

- Both Owners are in the regions with a uniformly high restriction level $\rightarrow$ In equilibrium, the products will be sold at higher prices, but the profits are not generally higher
- One Owner is in a region where the data restriction level is higher than the other one $\rightarrow$ In equilibrium, the Owner will earn much less profit than the opponent.
- The static game does not show a symmetric equilibrium despite a symmetric setting (under the same data restriction level). The Nash Equilibrium of asymmetric investment shows that a firm would invest if its opponent does not invest, i.e. the decisions upon the investments of both Owners do not converge.

# Dynamics with heuristics

| Scenario | $c_{0,a}$ | $c_{0,b}$ | $c_{res,R_a}$ | $c_{res,R_b}$ |
|---|---|---|---|---|
| 1' | 6 | 6 | 0 | 0 |
| 2' | 6 | 0 | 0 | 0 |
| 3' | 6 | 6 | 0 | 10 |
| 4' | 6 | 0 | 0 | 10 |
| 5' | 6 | 6 | 10 | 10 |
| 6' | 6 | 0 | 10 | 10 |

Table 1: Initial setup of the scenarios

# Explanation of the parameters

- $\psi_0 > 0$ guarantees the adjustments of the investments. A larger $\psi_0$ indicates that Owners take the random updates to a greater extent.
- $\Delta$ can be regarded as the learning rate. A larger $\Delta$ indicates that Owners take stronger measures regarding their investments.
- $\vartheta$ is the threshold of updates. A larger $\vartheta$ indicates that Owners only update their investments with a higher threshold.
- $\tau$ represents at which time unit an Owner updates its investment and shows the update frequency of the investment, i.e. an Owner updates its investment every $\tau$ months. A larger $\tau$ means slower updates in investment.

# Dynamics with heuristics with investment updates

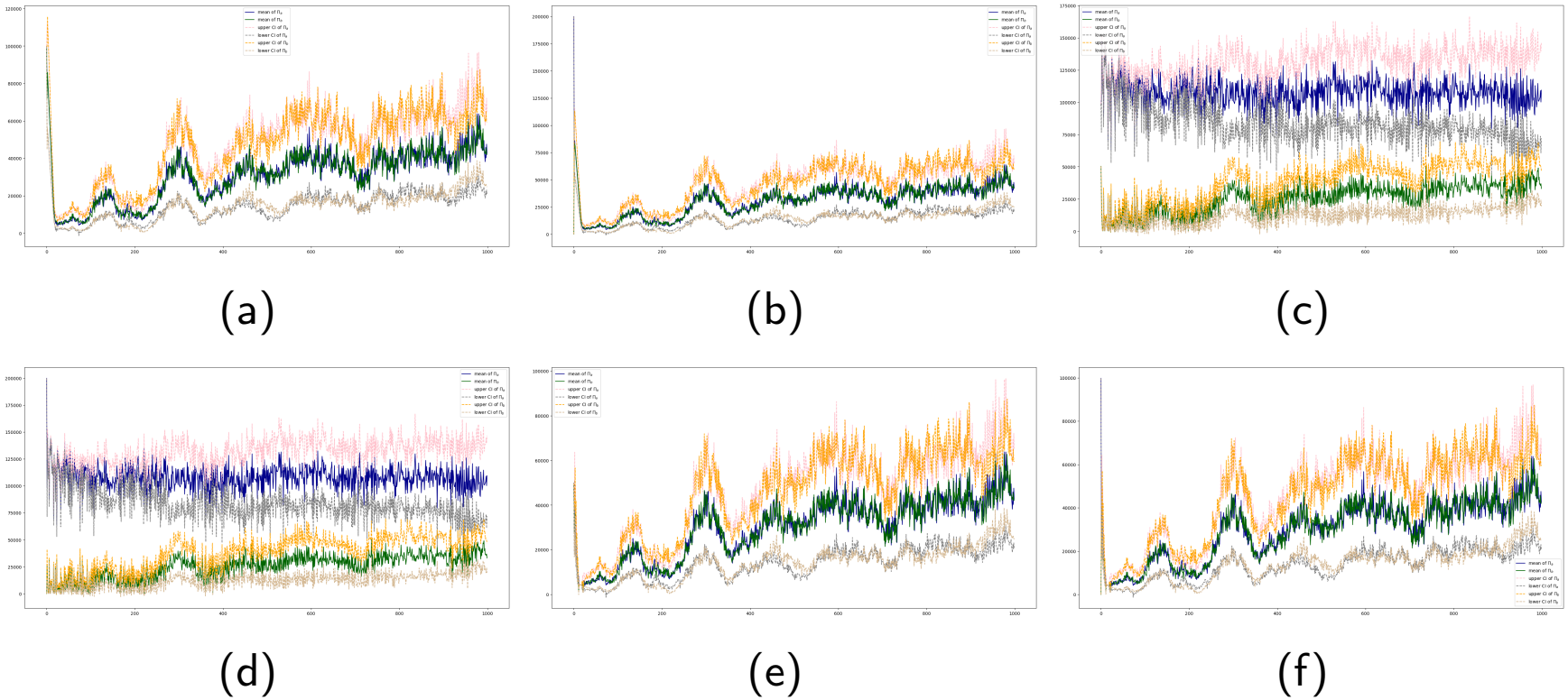



Figure 12: Statistics of $\Pi_a$ and $\Pi_b$ according to 30 time series. Tastes of Consumers follow a bivariate normal distribution (Dist. 1). Sub-figures (a) - (f) represent Scenarios 1' - 6'. $\Delta = 3$, $\psi_{\mathbf{0},a} = \psi_{\mathbf{0},b} = 1$, $\vartheta = 0.01$, $\tau = 6$.

# Dynamics with heuristics with investment updates

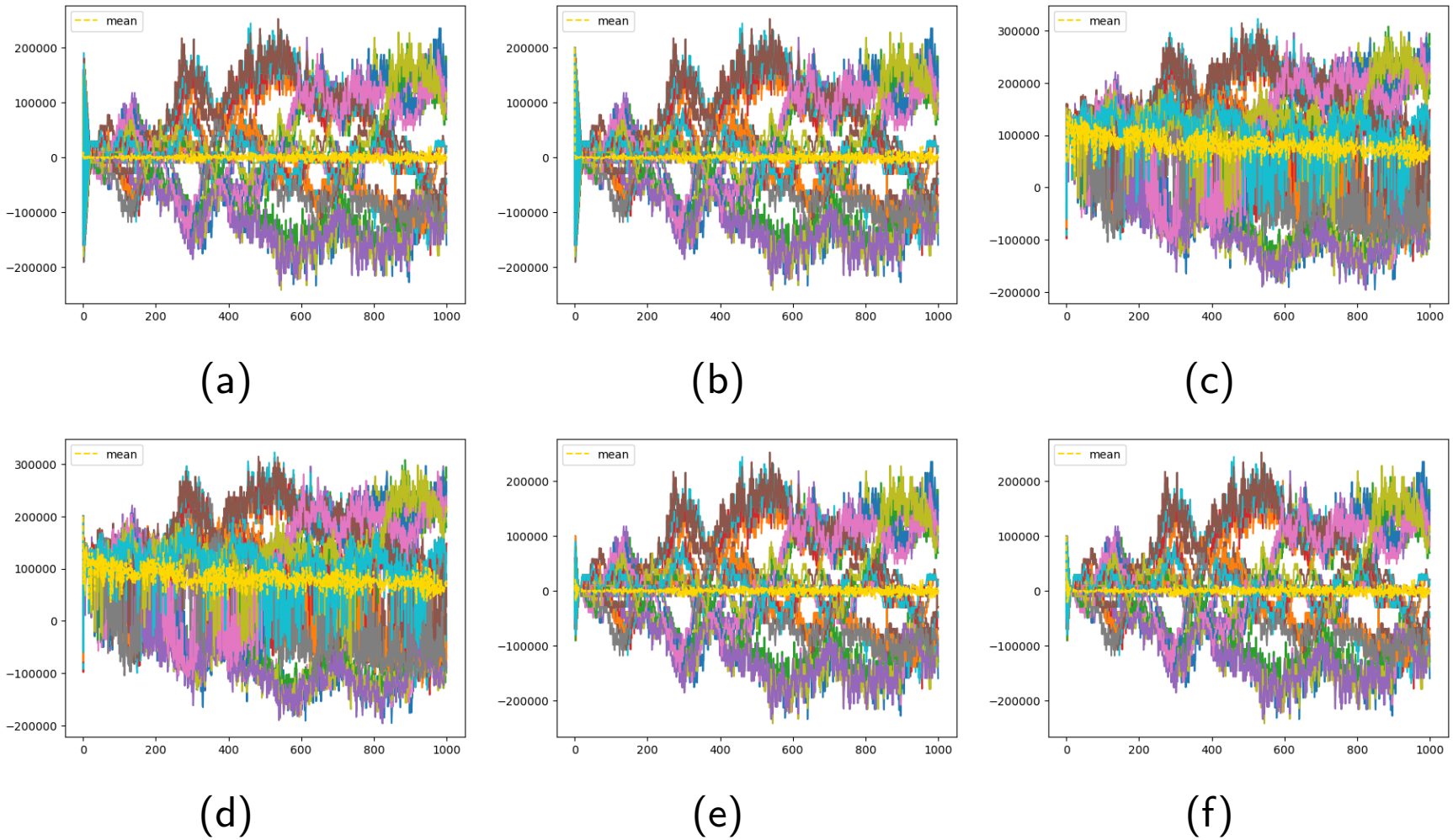



Figure 13: Difference of $\Pi_a$ and $\Pi_b$ according to 30 time series. Tastes of Consumers follow a bivariate normal distribution (Dist. 1). Sub-figures (a) - (f) represent Scenarios 1' - 6'. $\Delta = 3$, $\psi_{\mathbf{0},a} = \psi_{\mathbf{0},b} = 1$, $\vartheta = 0.01$, $\tau = 6$.

# Conclusion: What does the approach indicate?

- Static case
  - Increasing the data restriction level is not likely to bring more profit to the local Owners despite the rise in the selling price. Instead, an Owner may experience a loss under higher a data restriction level.
- Dynamic case
  - The prices may not reach a convergence under the settings of heuristics and both Owners may be better off after a large number of periods even if they are under different levels of data restrictions.
  - Difference in data restriction levels brings vertical difference to products, and through market competitions, both Owners could earn a profit (the Owner under a lower data restriction level earns more).
  - Cooperation across regions can focus on dividing the market based on a certain level of difference in local data restriction levels to attract target consumers.

# Reference I

Aaronson, Susan Ariel. 2018. *Data Is Different: Why the World Needs a New Approach to Governing Cross-border Data Flows.* Technical report 197. Waterloo: Centre for International Governance Innovation, November.

Asian Development Bank, Research Institute for Global Value Chains at the University of International Business, the World Trade Organization, the Institute of Developing Economies – Japan External Trade Organization, and the China Development Research Foundation. 2021. *Global Value Chain Development Report 2021: Beyond Production.* Technical report. November.

Azmeh, Shamel, Christopher Foster, and Ahmad Abd Rabuh. 2021. "The Rise of the Data Economy and Policy Strategies for Digital Development." *Digital Pathways Paper Series* (Oxford) (February).

Bauer, Matthias, Martina F Ferracane, and Erik van der Marel. 2016. "Tracing the Economic Impact of Regulations on the Free Flow of Data and Data Localization." *Centre for International Governance Innovation and Chatham House,* no. 30 (May).

Burfisher, Mary E. 2020. *Introduction to Computable General Equilibrium Models.* 3rd ed. Cambridge University Press, January. https://doi.org/10.1017/9781108780063.

# Reference II

Carvalho, Vasco M, Makoto Nirei, Yukiko U Saito, Alireza Tahbaz-Salehi, Daron Acemoglu, Pol Antràs, David Baqaee, et al. 2020. *Supply Chain Disruptions: Evidence from the Great East Japan Earthquake.* Technical report.

Casalini, Francesca, and Javier López-González. 2019. "Trade and Cross-Border Data Flows." *OECD Trade Policy Papers,* no. 220 (January).

Castro, Daniel, and Alan Mcquinn. 2015. "Cross-Border Data Flows Enable Growth in All Industries." *Information Technology & Innovation Foundation,* no. February, 1–23.

Chalaby, Jean K. 2019. "Outsourcing in the UK television industry: A global value chain analysis." *Communication Theory* 29 (4): 169–190. https://doi.org/10.1093/ct/qty024.

Connell, Julia, Renu Agarwal, Sushil, and Sanjay Dhir. 2018. *Global Value Chains, Flexibility and Sustainability.* Edited by Julia Connell, Renu Agarwal, Sushil, and Sanjay Dhir. Springer Nature Singapore Pre Ltd.

Dawid, Herbert, and Domenico Delli Gatti. 2018. "Agent-Based Macroeconomics." In *Handbook of Computational Economics,* 4:63–156. Elsevier B.V., January. https://doi.org/10.1016/bs.hescom.2018.02.006.

# Reference III

Dawid, Herbert, Philipp Harting, and Jasper Hepp. 2021. "Introduction Digitization along the Global Value Chain in an Agent-Based Model."

Dawid, Herbert, Philipp Harting, Sander van der Hoog, and Michael Neugart. 2019. "Macroeconomics with heterogeneous agent models: fostering transparency, reproducibility and replication." *Journal of Evolutionary Economics* 29, no. 1 (March): 467–538. https://doi.org/10.1007/s00191-018-0594-0.

Dawid, Herbert, and Gerd Muehlheusser. 2022. "Smart products: Liability, investments in product safety, and the timing of market introduction." *Journal of Economic Dynamics and Control* 134 (January). https://doi.org/10.1016/j.jedc.2021.104288.

De Backer, K, and S Miroudot. 2013. "Mapping Global Value Chains." *OECD Trade Policy Papers*, no. 159 (December). https://doi.org/10.1787/5k3v1trgnbr4-en.

Ferracane, Martina F., Hosuk Lee-Makiyama, and Erik Van der Marel. 2018. "Digital Trade Restrictiveness Index."

Frontier Economics Ltd. 2021. *The Value of Cross-border Data Flows to Europe: Risks and Opportunities - Report prepared for DIGITALEUROPE.* Technical report. June.

# Reference IV

Gabszewicz, J Jaskold, and J.-F. Thisse. 1979. "Price Competition, Quality and Income Disparities." *Journal of Economic Theory* 20 (3): 340–359.

———. 1982. "Product Differentiation with Income Disparities: An Illustrative Model." *The Journal of Industrial Economics* 31 (1/2): 115–129.

Goldfarb, Avi, and Catherine Tucker. 2017. "Digital Economics."

Graitson, Dominique. 1982. "Spatial Competition à la Hotelling: A Selective Survey." *The Journal of Industrial Economics* 31 (1/2): 11–25.

GSM Association. 2018. *The Data Value Chain.* Technical report. June.

Guan, Dabo, Daoping Wang, Stephane Hallegatte, Steven J. Davis, Jingwen Huo, Shuping Li, Yangchun Bai, et al. 2020. "Global supply-chain effects of COVID-19 control measures." *Nature Human Behaviour* 4, no. 6 (June): 577–587. https://doi.org/10.1038/s41562-020-0896-8.

Guo, Wen Chung, Fu Chuan Lai, and Dao Zhi Zeng. 2015. "A Hotelling model with production." *Mathematical Social Sciences* 73 (January): 40–49. https://doi.org/10.1016/j.mathsocsci.2014.11.005.

Hotelling, Harold. 1929. "Stability in Competition." *The Economic Journal* 39 (153): 41–57.

# Reference V

Kano, Liena, Eric W.K. Tsang, and Henry Wai chung Yeung. 2020. "Global value chains: A review of the multi-disciplinary literature." *Journal of International Business Studies* 51 (June): 577–622. https://doi.org/10.1057/s41267-020-00304-2.

Kathuria, Rajat, Mansi Kedia, Gangesh Varma, and Kaushambi Bagchi. 2019. *Economic Implications of Cross-Border Data Flows.* Technical report. November.

Koks, Elco E., and Mark Thissen. 2016. "A Multiregional Impact Assessment Model for disaster analysis." *Economic Systems Research* 28, no. 4 (October): 429–449. https://doi.org/10.1080/09535314.2016.1232701.

Larisa Serzo, Aiken O. 2020. *Cross-border Data Regulation for Digital Platforms: Data Privacy and Security.* Technical report. Quezon City: Philippine Institute for Development Studies, December.

Larralde, Hernán, Pablo Jensen, and Margaret Edwards. 2006. "Two dimensional Hotelling model : analytical results and numerical simulations." November.

Larralde, Hernán, Juliette Stehlé, and Pablo Jensen. 2009. "Analytical solution of a multi-dimensional Hotelling model with quadratic transportation costs." *Regional Science and Urban Economics* 39 (3): 343–349. https://doi.org/10.1016/j.regsciurbeco.2009.01.001.

# Reference VI

Nagle, Thomas T., and Georg Müller. 2018. *The Strategy and Tactics of Pricing: A Guide to Growing More Profitability.* 6th ed. Taylor & Francis.

Nguyen, David, and Marta Paczos. 2020. "Measuring the Economic Value of Data and Cross-Border Data Flows: A Business Perspective." *OECD Digital Economy Papers,* no. 297 (August).

Peng, Yu, Qian Lu, Xue Wu, Yueru Zhao, and Yue Xiao. 2020. "Dynamics of Hotelling triopoly model with bounded rationality." *Applied Mathematics and Computation* 373 (May). https://doi.org/10.1016/j.amc.2019.125027.

Porter, Michael E. 1985. *Competitive advantage: Creating and Sustaining Superior Performance.* New York: The Free Press.

Pussep, Anton, Markus Schief, Thomas Widjaja, Peter Buxmann, and Christian M. Wolf. 2011. "The Software Value Chain as an Analytical Framework for the Software Industry and its Exemplary Application for Vertical Integration Measurement." In *17th Americas Conference on Information Systems 2011, AMCIS 2011,* vol. 5. January.

Rose, Adam, and Shu Yi Liao. 2005. "Modeling regional economic resilience to disasters: A computable general equilibrium analysis of water service disruptions." *Journal of Regional Science* 45 (1): 75–112. https://doi.org/10.1111/j.0022-4146.2005.00365.x.

# Reference VII

Schlüter, Florian, and Endric Hetterscheid. 2017. "Supply Chain Process Oriented Technology-Framework for Industry 4.0," https://doi.org/10.15480/882.1467.

Shy, Oz. 1996. *Industrial Organization: Theory and Applications.* The MIT Press.

Tabuchi, Takatoshi, and Jacques-François Thisse. 1995. "Asymmetric equilibria in spatial competition." *International Journal of Industrial Organization* 13 (2): 213–227.

Tversky, Amos, and Daniel Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science* 185, no. 4157 (September): 1124–1131.

UNCTAD. 2021. *Digital economy report 2021. Cross-border data flows and development: For whom the data flow.* 238. September.