How should data markets be regulated? Navigating the fine line between insider trading and market efficiency.

Lina Lukyantseva

Motivation

- SEC v. Huang (2016): Data analyst for Capital One downloaded and analyzed data on retail purchases made with Capital One credit cards. He used this information to predict revenues of retailers, and then traded retailers' stocks in advance of the public release of quarterly sales announcements by these companies.

- What are the consequences?
  - **More efficient outcome** – information was reflected in the stocks' prices earlier which led to better investors' decisions on the allocation of their capital.
  - **Unfair outcome** – Huang received high profits at the expense of other traders who did not have access to the information.

- Huang was found liable for illegal insider trading.

Motivation

- LSE Alternative Investment Conference 2017
- "Hedge funds see a gold rush in data mining", *Financial times* (2017):

> One LSE student asked how all this data could help Point72 if everyone had access to the same information. The answer was exclusivity agreements, Mr Granade said: "The great thing about this area is you can arrange deals where you are the only ones who get it."

American vs European Regulation

- E.U. (Market Abuse Regulation):
  It is illegal to trade based on material non-public information.

- US (SEC.gov):
  It is illegal to trade based on material non-public information, in breach of a fiduciary duty or other relationship of trust and confidence.

Research question

- Why we might want to regulate data markets?
- What are the consequences of different policies?

Model: Sectors of the Economy

financial market

- N hedge funds
- liquidity traders
- market makers

data market

- N hedge funds
- data seller (he)
- regulator (she)

production economy

Model: Data

- One risky asset with payoff $\tilde{v} \sim N(\overline{v}, \sigma^2)$
- DS can provide a dataset which contains an unbiased estimate of $\tilde{v}$ at a fixed cost $R \geq 0$: $\tilde{v} + \eta$
  - $\eta \sim N(0, c_\eta \sigma^2)$, c_η is chosen by the DS
  - c_η is an objective measure of the quality of the data
- If fund i gets the dataset, it observes $\tilde{v} + \eta + \epsilon_i$
  - Interpretation error $\epsilon_i \sim N(0, c_i \sigma^2)$
  - Funds that are more quantitative in nature and have better data science teams will have lower c_i
- All parameters are commonly known, all parameters besides c_η are exogenous

Model: Financial Market

A modification of Kyle (1985).

Trading happens in 2 steps:

- Step 1: hedge funds and liquidity traders simultaneously place market orders
  - liquidity traders trade for exogenous reasons, $y_l \sim N(0, \sigma_l^2)$
  - Hedge fund i chooses y_i to maximize the expected profit
  - if fund i has the dataset, it solves

$$\max_{y_i} \mathbb{E}\left(y_i \cdot (\tilde{v} - P) \mid \tilde{v} + \eta + \epsilon_i\right)$$

  - if fund i does not have the dataset, it solves

$$\max_{y_i} \mathbb{E}\left(y_i \cdot (\tilde{v} - P)\right)$$

- Step 2: market makers observe the aggregate demand $\sum_{i=1}^{N} y_i + y_l$ and compete for the opportunity to fulfill it. This drives the price to $P = \mathbb{E}\left(\tilde{v} \mid \sum_{i=1}^{N} y_i + y_l\right)$

Model: Regulator's Objective

- Price Informativeness (PI) = $Cov(P, \tilde{v})$
- Output Y is an increasing function of PI
- $u_{reg} = \pi_{DS} + \sum_{i=1}^{N} \pi_i + \pi_{mm} + w\pi_I + Y(PI)$, $w \geq 1$
  - If $w = 1$, $u_{reg} = -R + Y(PI)$
- $u_{reg} = -R + (w-1)\pi_I + Y(PI)$
  - the regulator cares about fairness and price informativeness

Model: Timing of the Game

- Step 1: The regulator chooses a "policy";
- Step 2: The profile of the analyst error variances of the funds $\{c_1\sigma^2, ..., c_N\sigma^2\}$ is realized (c_i is drawn i.i.d. from some distribution F on $[0, \infty)$);
- Step 3: The data seller decides whether to collect a dataset at a fixed commonly known cost $R \geq 0$.
  - If he does not, then the game proceeds to step 5;
  - If he does, he also chooses the data error variance $c_\eta \geq 0$. Then the game proceeds to step 4;
- Step 4: Some funds purchase the data from the data seller (in accordance with the policy);
- Step 5: The play in the financial market happens. The profits and the utility of the regulator are realized.

Outline

1 Equilibrium in the financial market

2 Price informativeness in equilibrium

3 Regulation regimes:

- Benchmark: No regulation (American regulation)
- Policy 1: Fixed price (European regulation)
- Policy 2: Lower bound on the quantity
- Policy 3: Auction with a lower bound on the quantity
- Policy 4: Auction with contingent payments and a lower bound on the quantity

Equilibrium in the Financial Market

Proposition

Given a set of informed hedge funds $\mathbb{I}$, the following profile of market orders $y = \{y_1, ..., y_N\}$ and market price P constitute an equilibrium:

(1) $P = \overline{v} + \lambda(\sum_{i=1}^{N} y_i + y_I)$;

(2) If fund i is uninformed (i.e. $i \notin \mathbb{I}$), then $y_i = 0$;

(3) If fund i is informed (i.e. $i \in \mathbb{I}$), then $y_i = \frac{\alpha_i}{\lambda}(\tilde{v} + \eta + \epsilon_i - \overline{v})$;

where

$$\alpha_i = \frac{1}{(1 + c_\eta + 2c_i)\left(1 + \sum_{j \in \mathbb{I}} \frac{1}{1 + \frac{2c_j}{1 + c_\eta}}\right)}$$

and

$$\lambda = \frac{\sigma}{\sigma_I}\sqrt{\sum_{i \in \mathbb{I}} \alpha_i^2 (1 + c_\eta + c_i)}.$$

What Determines Price Informativeness?

- $PI = Cov(\tilde{v}, P)$
- $PI(\mathbb{I}, c_\eta)$, where $\mathbb{I}$ is the set of the informed funds

- Given the data error variance c_η, for any $\mathbb{X}, \mathbb{Y} \subseteq \mathbb{F}$ if $\mathbb{X} \subset \mathbb{Y}$, then $PI(\mathbb{X}, c_\eta) < PI(\mathbb{Y}, c_\eta)$

- Given the data error variance c_η, for any $\mathbb{X} \subset \mathbb{F}$ and any $y, z \in \mathbb{F} \setminus \mathbb{X}$, $c_y < c_z \iff PI(\mathbb{X} \cup \{y\}, c_\eta) > PI(\mathbb{X} \cup \{z\}, c_\eta)$

- Given the set of informed funds $\mathbb{I}$, $PI(\mathbb{I}, c_\eta)$ is decreasing in c_η

Roughly speaking, the more funds receive the data, the more competent they are and the higher the quality of the data, the more informative the market outcome.

Benchmark: No regulation. Setup

- The data seller is able to commit to sell to an exclusive set of funds
  - Step 1: data seller chooses c_η which becomes commonly known
  - Step 2: data seller simultaneously and publicly makes Take-It-or-Leave-It offers to some set $\mathbb{I}$ of the funds
  - Step 3: Each fund that received an offer decides whether to accept it or not. Those funds who accepted the offer receive the dataset at their respective prices.

In equilibrium, the data seller chooses $c_\eta^*, \mathbb{I}^*$ that solve

$$\max_{(\mathbb{I}, c_\eta)} \sum_{i \in \mathbb{I}} \pi_i(\mathbb{I}, c_\eta)$$

and makes TILI offers to all funds $i \in \mathbb{I}^*$ at price $P_i = \pi_i(\mathbb{I}^*, c_\eta^*)$. The funds accept their offers.

Benchmark: No regulation. Optimal data seller's action

Proposition

In the absence of regulation the data seller always chooses the highest quality of the data, i.e. $c_\eta^* = 0$.

Proposition

In the absence of regulation the data seller makes offers to a subset of the most competent funds.

- If all the funds are equally good at interpreting the data, i.e. for all $i \in \mathbb{F}$ $c_i = c$, then it is optimal to sell to $n = 1 + 2c$ or one of the two nearest integers if c_i is not an integer funds.
- When funds are less precise, it is harder for the market makers to make a correct inference about $\tilde{v}$ and so adding an additional buyer creates less competition effect on the existing buyers.

Benchmark: No regulation. Optimal data seller's action

Proposition

In the absence of regulation the data seller always chooses the highest quality of the data, i.e. $c^*_\eta = 0$.

Proposition

In the absence of regulation the data seller makes offers to a subset of the most competent funds.

- If all the funds are equally good at interpreting the data, i.e. for all $i \in \mathbb{F}$ $c_i = c$, then it is optimal to sell to $n = 1 + 2c$ [1] funds.
- When funds are less precise, it is harder for the market makers to make a correct inference about $\tilde{v}$ and so adding an additional buyer creates less competition effect on the existing buyers.

[1] or one of the two nearest integers if c_i is not an integer

Benchmark: No regulation. Welfare implications

Proposition

In the absence of regulation the equilibrium outcome is the worst possible outcome from the liquidity traders' perspective.

Proof:

$$\sum_{i=1}^{N} \pi_i + \pi_{mm} + \pi_l = 0$$

$$\pi_{mm} = 0 \text{ in equilibrium}$$

$$\pi_{DS}^{\text{no reg}} = \sum_{i \in \mathbb{I}} \pi_i = \sum_{i=1}^{N} \pi_i = -\pi_l$$

Conclusion: Conditional on the data quality remaining the same, selling to additional buyers would increase both price informativeness (strictly) and welfare of the liquidity traders (weakly).

Policy 1: Fixed Price

- Step 1: The data seller chooses c_η and P
- Step 2: All hedge funds simultaneously decide whether to buy the dataset at price P

Proposition

If in the absence of regulation it is optimal for the data seller to sell to only one fund, then under the fixed price policy it is still optimal and feasible to sell to only one fund.

Proof

Policy 2: Lower Bound on the Quantity

- The regulator sets a lower bound on the number of the buyers $\underline{n}$ but the data seller can choose who to sell to and at what prices.
- The policy can be useful Example

Proposition

Let $\mathbb{I}^*$ be the set of informed funds in the absence of regulation, $|\mathbb{I}^*| = k$. Let $\underline{n} > k$ be the lower bound on the quantity, and suppose that there exists a set $\mathbb{X}$ of size $\underline{n} - k$ of "extremely incompetent" funds (i.e. $c_i \to \infty$ for all $i \in \mathbb{X}$) that are not in the current set of buyers $\mathbb{I}^*$.

Then

$$\pi_{DS}(\mathbb{I}^* \cup \mathbb{X}) \to \pi_{DS}(\mathbb{I}^*)$$

$$PI(\mathbb{I}^* \cup \mathbb{X}) \to PI(\mathbb{I}^*)$$

$$\pi_I(\mathbb{I}^* \cup \mathbb{X}) \to \pi_I(\mathbb{I}^*)$$

Policy 3: Auction with a Lower Bound on the Quantity

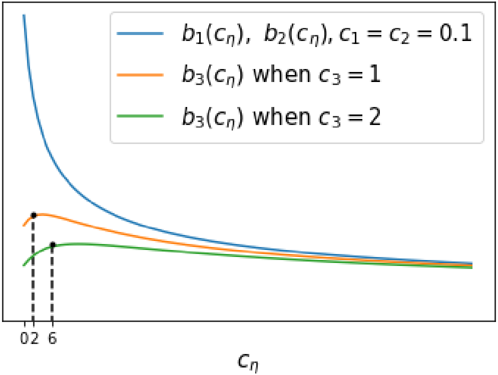
- The regulator sets $\underline{n}$, then the data seller chooses c_η and $K \geq \underline{n}$ to sell K datasets through a $K+1$ price auction
- Let's enumerate the funds so that $c_1 \leq ... \leq c_N$
- Equilibrium bidding strategy is
  - if $i \leq K$, $b_i = \pi_i(\{1, ..., K\}, c_\eta)$
  - if $i > K$, $b_i = \pi_i(\{1, ..., K-1, i\}, c_\eta)$

Policy 3: Auction with a Lower Bound on the Quantity

Proposition

It can be optimal for the data seller to not choose the highest quality of the data.

Example: Consider a data market with 3 funds: $c_1 = c_2 = 0.1$ and $c_3 > 0.1$. Let $\underline{n} = 2$. The optimal level of noise c_η^* is increasing in c_3 and PI is decreasing in c_3. As $c_3 \to \infty$, $c_\eta^* \to \infty$ and $PI \to 0$.



Policy 4: Auction with Contingent Payments and a Lower Bound on the Quantity

Based on Hansen, 1985

- Step 1: The regulator sets the lower bound on the quantity $\underline{n}$ and a small entry fee f
- Step 2: The data seller chooses c_η and a number $K \geq \underline{n}$ of datasets to sell
- Step 3: Hedge funds submit their bids as percentage of the future profit
- Step 4: Funds with top K bids win and pay $K + 1$ bid's percentage of their profits
- Step 5: Regulator collects the entry fees from the winners and returns the entry fees to the losers

Policy 4: Auction with Contingent Payments and a Lower Bound on the Quantity

Based on Hansen, 1985

- Step 1: The regulator sets the lower bound on the quantity $\underline{n}$ and a small entry fee f
- Step 2: The data seller chooses c_η and a number $K \geq \underline{n}$ of datasets to sell
- Step 3: Hedge funds submit their bids as percentage of the future profit
- Step 4: Funds with top K bids win and pay $K + 1$ bid's percentage of their profits
- Step 5: Regulator collects the entry fees from the winners and returns the entry fees to the losers

Proposition

It is optimal for the data seller to choose the highest data quality, and K most precise funds receive the dataset in equilibrium.

Conclusion

- Trading on data sold through data markets can have similar consequences as insider trading;
- When considering different regulations one should be mindful about how they affect the equilibrium set of data buyers as well as data seller's incentives to provide high quality data;
- Auction with contingent payments and a lower bound on the quantity increases both liquidity traders' welfare and price informativeness in expectation.