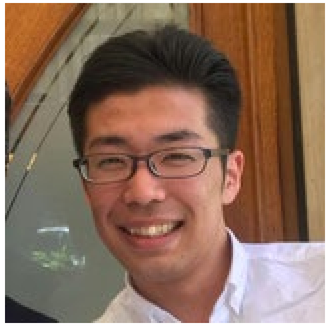
DDAP 13@YITP, invited talk
9:30-10:20 (35+15 mins), 4th July 2024

Statistical physics of the long-memory order flow in financial market microstructure

Kiyoshi Kanazawa, Kyoto University (Department of Physics)

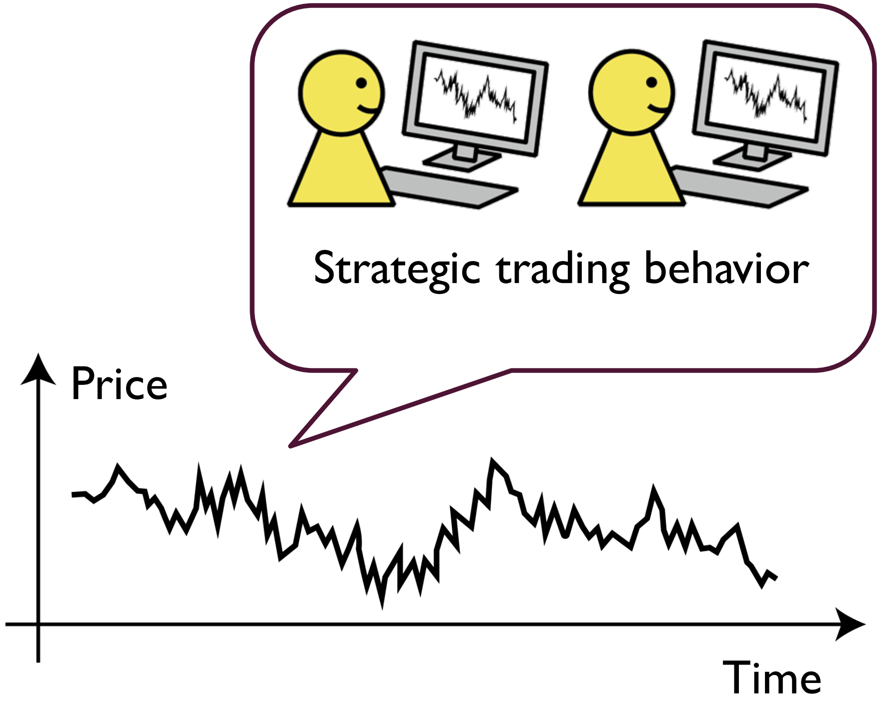
- Y. Sato and KK, *Phys. Rev. Lett.* **131**, 197401 (2023); highlighted in *Physics (Viewpoints)* by APS
- Y. Sato and KK, *Phys. Rev. Res.* **5**, 043131 (2023)
- Y. Sato and KK, *J. Stat. Phys.* **191**, 58 (2024)

Long term workshop in the 2nd week:
"Stochastic thermodynamics for general non-Markovian processes"
as collaboration with A. Dechant



Collaborator
Yuki Sato (D2)

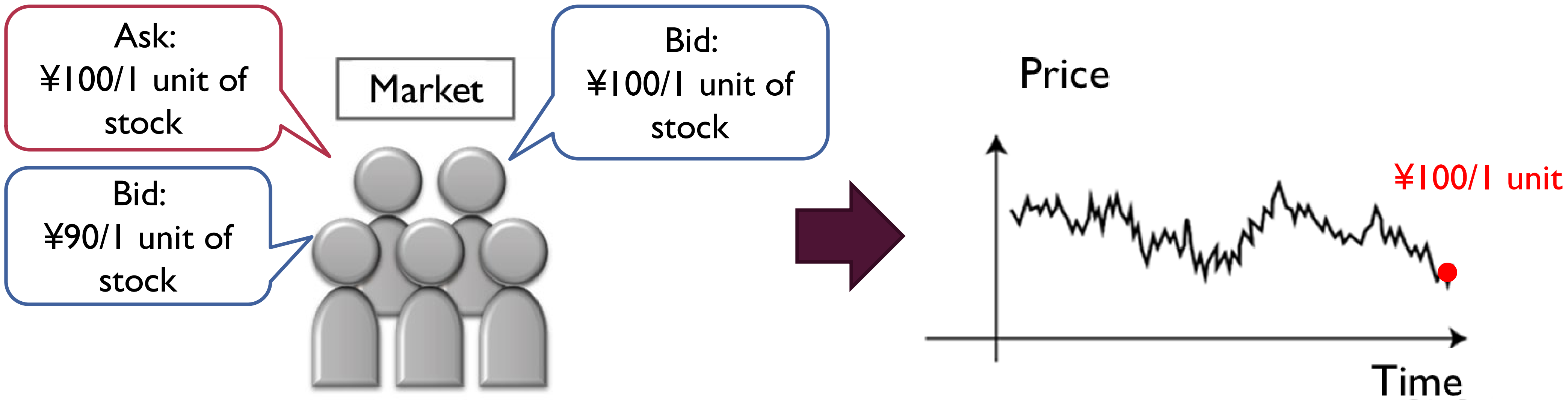
Econophysics of market microstructure: developing financial microscopic theories on the level of individual traders



- Econophysics: interdisciplinary statphys for economics
  - ✓ Finance
  - ✓ Network science
  - ✓ Social Network Science
- Market microstructure of finance: modeling at the level of individual traders regarding order submission
- This talk focuses on the data analytical examinations of econophysics theories

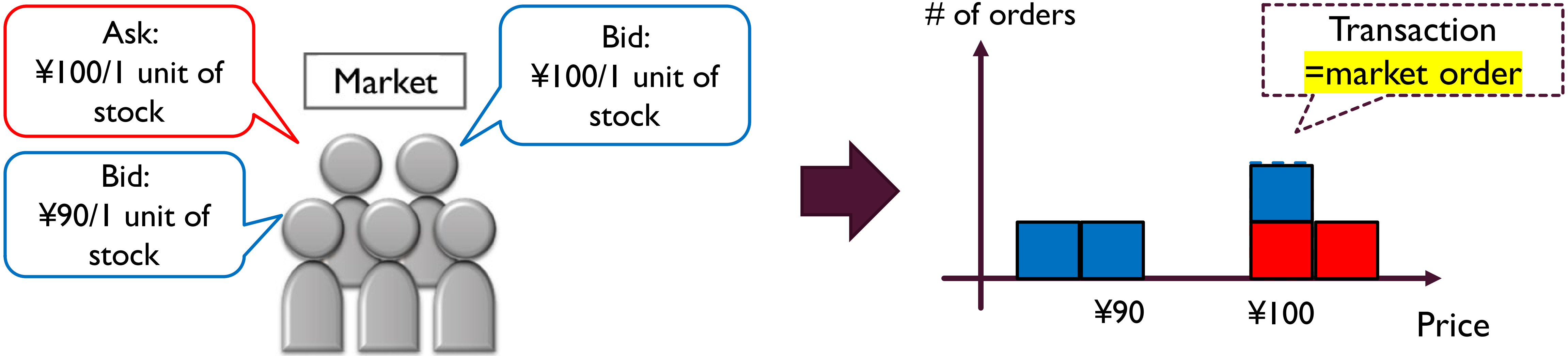
Goal: developing trader-level models to study macroscopic behavior of financial markets via statistical physics

Review: trading rule of financial markets (continuous double auction)



1. In advance, traders submit their bid and ask limit orders
2. Transaction = matching btw bid and ask prices
3. Transacted price recorded as time series

Visualization of the order flow: order book dynamics and price formation

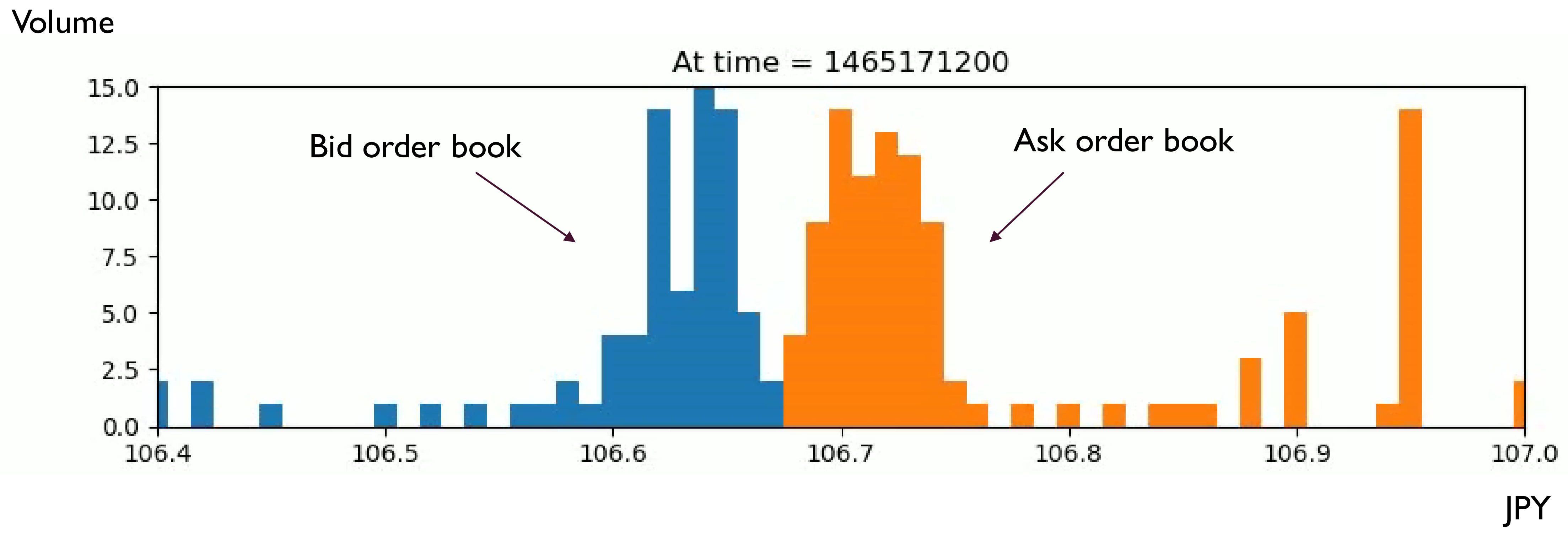


- Limit orders: order flow of bid and ask is displayed as the order book
  i. Red block = ask limit order, waiting for transactions
  ii. Blue block = bid limit order, waiting for transactions
- Market orders = orders triggering immediate transactions

Today's focus

  ✓ Transaction = Matching btw bid and ask "blocks"

Movie of order-book dynamics (forex, USDJPY)



2 topics in this talk

Focus: microscopic data analyses of *market orders*

- Part 1: an econophysics theory of the long memory of market orders
  - ✓ Statphys theory of the market-order autocorrelation
  - ✓ Precise verification of the microscopic statphys theory by a data analysis
- Part 2: nonlinear response of the market price to large market orders
  - ✓ Hypothetical universal scaling relation regarding the market response
  - ✓ Precise verification of this universality hypothesis

Part 1:
Long-range correlation in the market-order flow

Focus of the 1st part: the origin of the persistence of buy-sell market order signs

※Long range correlation = LRC, market order = MO

Lag $= \tau$

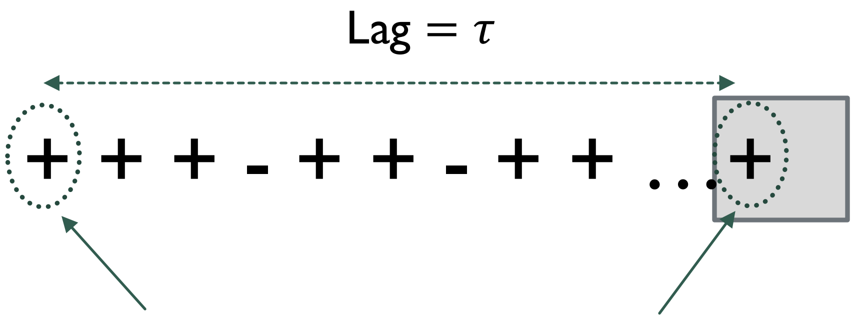
\+ + + - + + - + + … ?

Let's predict the future sign!

- Buy MO $\Rightarrow \epsilon_t = +$
- Sell MO $\Rightarrow \epsilon_t = -$

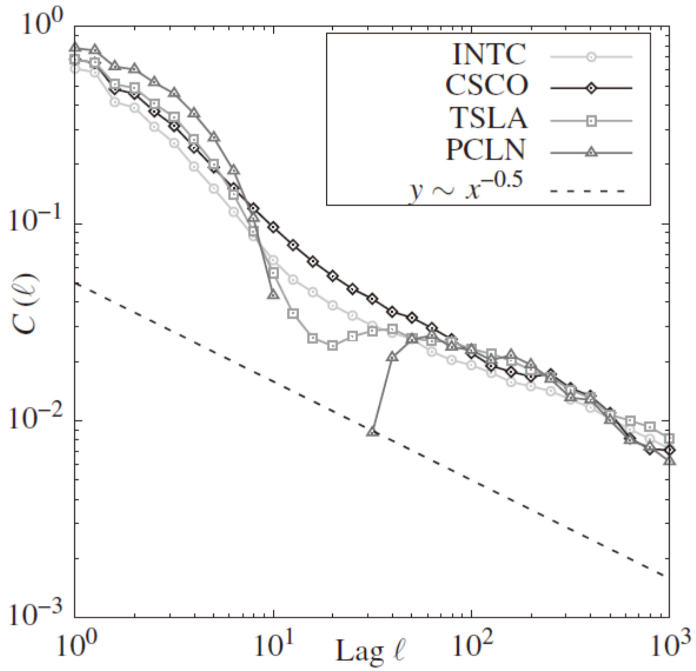
Focus of the 1st part: the origin of the persistence of buy-sell market order signs

※Long range correlation = LRC, market order = MO



1 buy MO (+)
⇒ buy MO (+) likely to appear

- Buy MO ⇒ $\epsilon_t = +$
- Sell MO ⇒ $\epsilon_t = -$



Autocorrelation function (ACF)

$$C(\tau) := \mathrm{E}[\epsilon_t \epsilon_{t+\tau}] \propto \tau^{-\gamma}$$

$$0 < \gamma < 1$$

Ref.: Bouchaud, *Trades, Quotes and Prices* (Cambridge Univ. Press, 2018)

Focus of the 1st part: the origin of the persistence of buy-sell market order signs

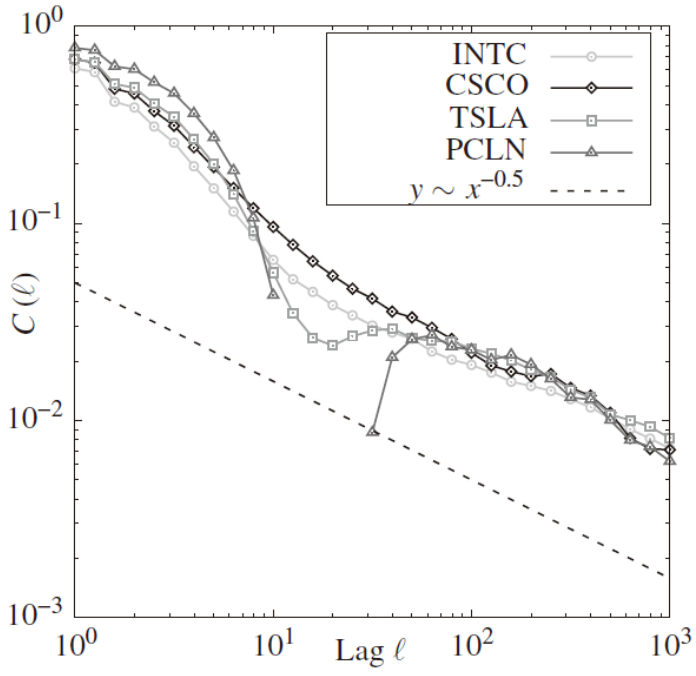
※Long range correlation = LRC, market order = MO

Lag = τ

+ + + - + + - + + … +

1 buy MO (+)
⇒ buy MO (+) likely to appear

- Buy MO ⇒ $\epsilon_t = +$
- Sell MO ⇒ $\epsilon_t = -$



Autocorrelation function (ACF)

$$C(\tau) := \mathrm{E}[\epsilon_t \epsilon_{t+\tau}] \propto \tau^{-\gamma}$$

$$0 < \gamma < 1$$

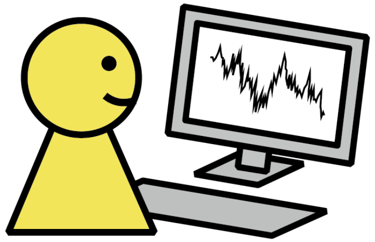
※ no profit over spread

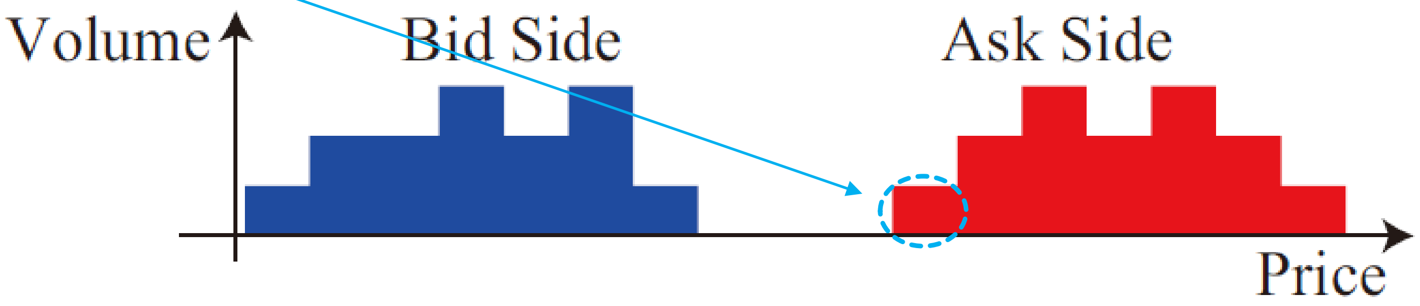
- Empirical law: order signs are very persistent (LRC)
- Example: 1 buy order ⇒ order sign is predictable for a few hours/days

Field interest: what is the microscopic origin of this phenomenon?

Previous study 1: hypothesis on the microscopic origin of the LRC = order splitting hypothesis (based on practical restriction)

Decision by the trader: 1000 units buys
But, *only 10 units* on the best ask…





Order splitting by the identical trader (Trader A): 1000 units = 10 units × 100 times

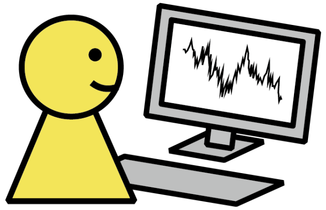
\+ + + − + + − + + …+ +

- Practice: traders split large ***metaorders*** into small child orders
- Interpretation: shortage of liquidity on the order book…

- No way to avoid splitting
- Predictable order signs

NOTE: a schematic of buy-sell order signs on the level of a *single trader*

100 successive buys (+) ⇒ we call **Run $L = 100$**



Order signs of
a specific trader A

$+ \; + \; + \; + \; + \; + \; + \; + \; + \; + \; \ldots$

Order splitting

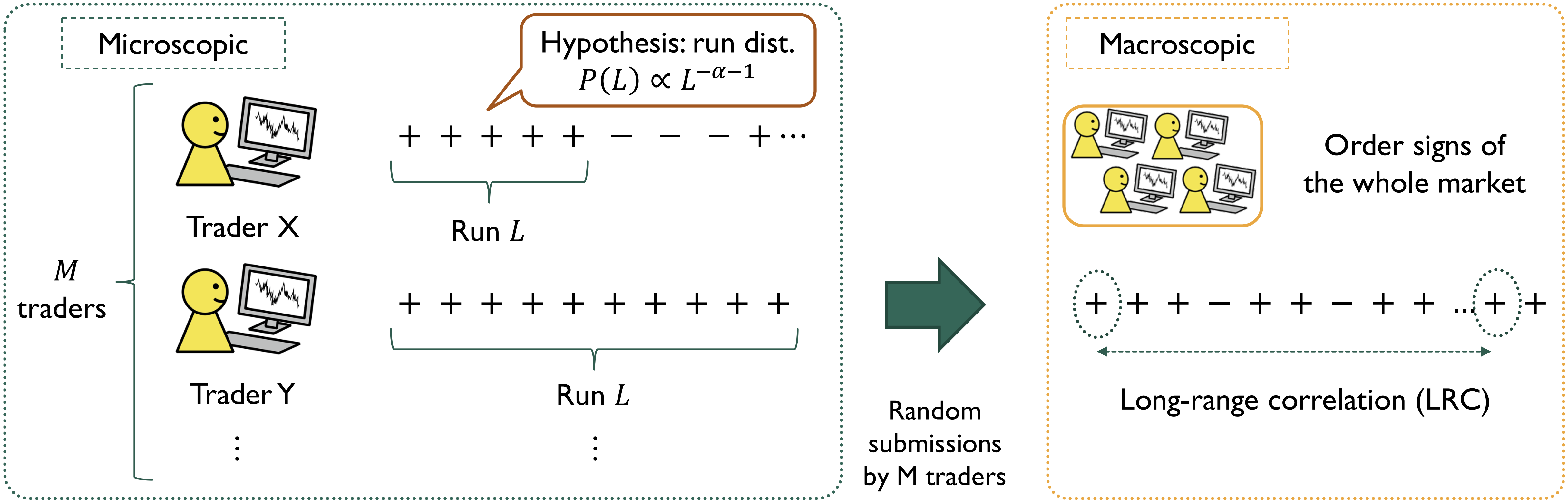
Orders of the identical trader
(that's why the signs are the same)

Order signs of the
whole market

$+ \; + \; + \; - \; + \; + \; - \; + \; + \; \ldots + \; +$

Ref.: F. Lillo, S. Mike, and J. D. Farmer, *Phys. Rev.* E **71**, 066122 (2005).

Previous study 2: microscopic model of the LRC = Lillo-Mike-Farmer (LMF) model



- M traders randomly split their orders
- Run-length dist. obeys a power law $P(L) \propto L^{-\alpha-1}$

The LRC appears in the whole market

Ref.: F. Lillo, S. Mike, and J. D. Farmer, Phys. Rev. E **71**, 066122 (2005).

Previous study 3: quantitative prediction by LMF model
Exponent γ in the ACF is related to the exponent α in the run-length PDF

Microscopic parameter α

- Assumption: power-law dist. for the run-length L

$$P(L) \propto L^{-\alpha-1}, \qquad 1 < \alpha < 2$$

Order sign of a specific trader A

$+ + + + + - - - + \cdots$

Run L

Macroscopic parameter γ

- Power-law decay of the ACF

$$C(\tau) := \mathrm{E}[\epsilon_t \epsilon_{t+\tau}] \propto \tau^{-\gamma}$$

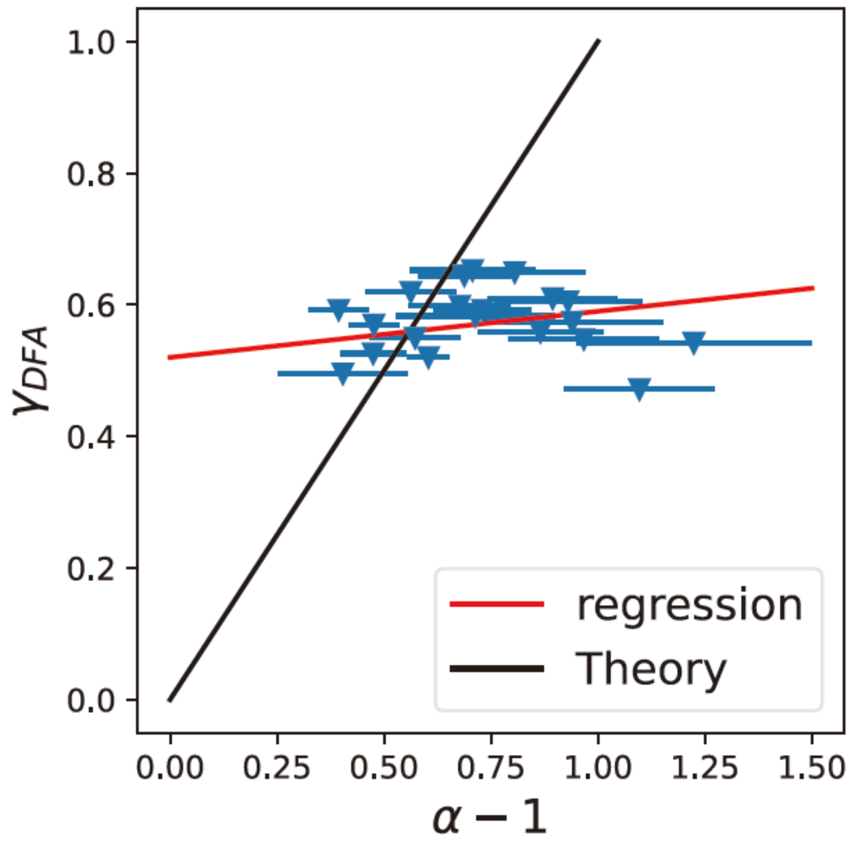
Order sign of the whole market

$+ + + - + + - + + \ldots + +$

LRC

LMF prediction: $\boldsymbol{\gamma = \alpha - 1}$ (predicted from micro to macro)

Previous study 4: empirical verification in the original article in LMF 2005



NOTE: The data extracted from LMF 2005 to put the regression line

- No appropriate data available at that time
- They used the off-book market data of London Stock Exchange as "imperfect proxy"
- Their observation:
  - ✓ Positive (*qualitative* level): the theoretical line passes through the centre
  - ✓ Negative (*quantitative* level): the regression line shows no correlation…

The prediction confirmed at the qualitative level, but *not at the quantitative level* perhaps due to the problem of the data size or statistical analysis…

Ref. on the *trading desk*: Goshima et al. 2019

Dataset: special order-book dataset on Tokyo Stock Exchange (TSE)

- Dataset: provided by TSE (with all IDs hashed)
  1. Order ID included; lifecycle of all the orders can be tracked
  2. Virtual server ID included
  3. For all the stocks during 2011-2020
- Virtual server ID
  - ✓ A unit of trading accounts on TSE
  - ✓ Technically, not a trader ID since a trader may have several virtual servers
- By appropriate aggregation of virtual server IDs, called *trading desk*, it is virtually possible to analyse individual traders' behavior

In our study, trader IDs are allocated by appropriate aggregation of virtual server IDs

Goal of this talk: (i) identification of splitting traders by strategy clustering, (ii) confirmation of the quantitative prediction of the LMF model

(i) Identification of splitting traders (strategy clustering)

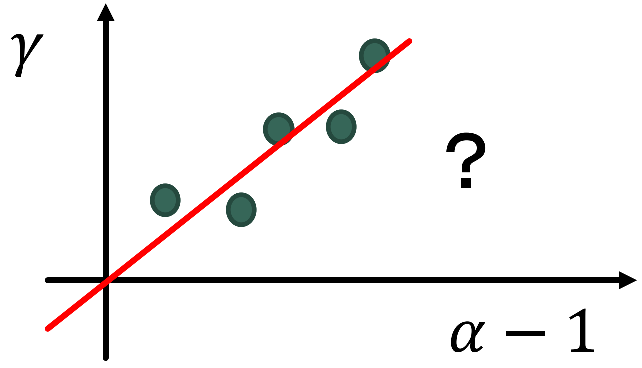
- Q: Are the splitting traders truly present in the TSE markets?
- Q: How to identify the cluster of the splitting traders?
- Q: Measurement of the metaorder length distribution. Does the PDF really obey the power law $P(L) \propto L^{-\alpha-1}$?

(ii) Validation of the LMF prediction (scatterplot)

- Quantitative prediction by the LMF:

$$\gamma = \alpha - 1$$

- Validation: scatterplot btw α and γ



Results in Part 1:
strategy clustering to identify splitting traders and verification of the LMF theory

Strategy clustering of individual traders in terms of MOs
Random traders (RTs) vs. splitting traders (STs)

Statistical test for order-sign sequences to classify traders into RTs or STs

Trader X : $+++++++++-------------++++++++++\cdots$

Run $L = 9$ Run $L = 13$

Random traders (RTs)

- Null hypothesis: order-sign sequence is random
- E.g., $+ + - + - + - + + - -$
- Random ⇔ Run-length dist. is exponential

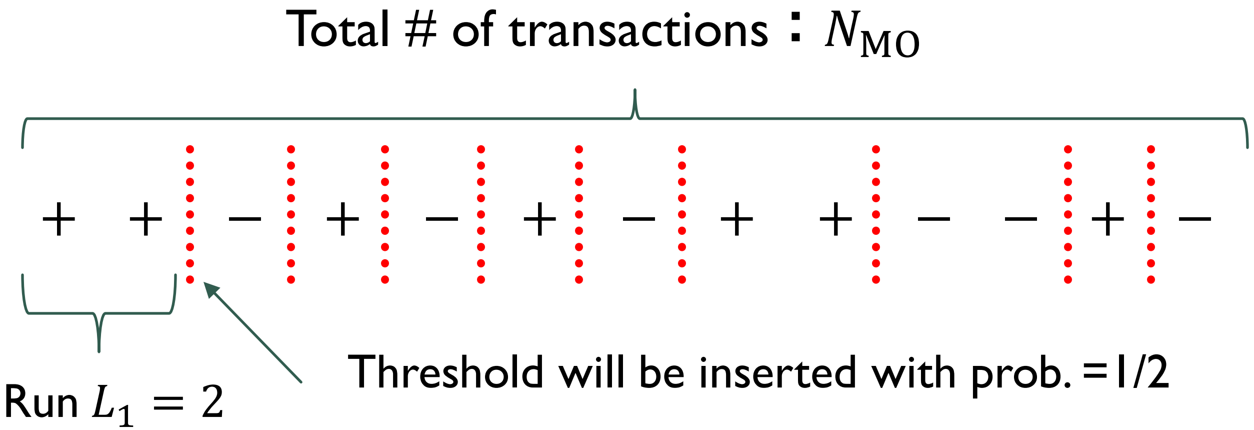
$$P(L) = \frac{1}{2^L}$$

Splitting traders (STs)

- Alternative hypothesis: order-sign sequence is not random
- E.g., $+ + + + + + + - - - - - - + \ldots$
- Non-random ⇒ Run-length dist. has a fatter tail

Statistical test (binomial) with $p = 0.01$ to classify traders into RTs or STs at the level of individual traders

Test statistic for the strategy clustering



Run: $\{L_k\}_{k=1,\ldots N_{\mathrm{run}}}$
Total # of runs: N_{run}

For this case,
$N_{\mathrm{MO}} = 13, N_{\mathrm{run}} = 10$

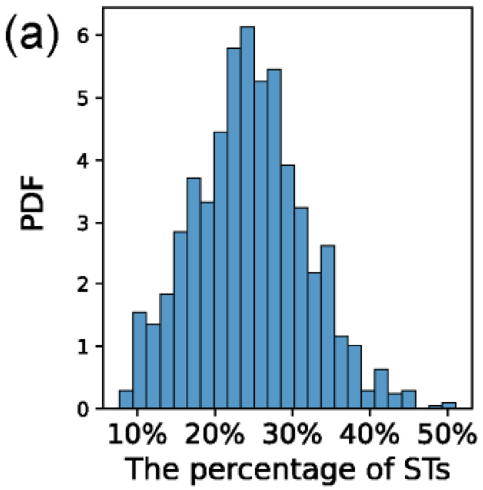
- Null hypothesis: purely random (symmetric Bernoulli process with $p = 1/2$)
- Distribution of the total # of the runs N_{run}:

$$P(N_{\mathrm{run}}) = \frac{1}{2^{N_{\mathrm{MO}}-1}} \binom{N_{\mathrm{MO}} - 1}{N_{\mathrm{run}}}$$

One-sided test regarding N_{run} for each trader with the significance level $\theta = 0.01$;
If the null hypothesis is rejected, the trader is a *splitting trader*; otherwise a *random trader*

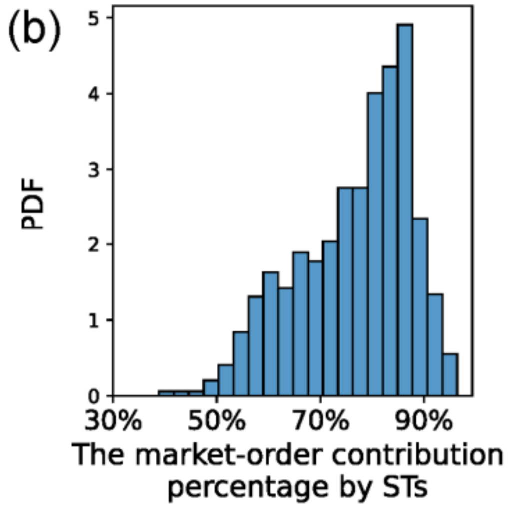
Result 1: direct confirmation of the presence of the splitting traders (STs)

(a) Percentage of the splitting traders



- Typically, 25% of the traders are splitting traders
- 1 datapoint = 1stock for a year

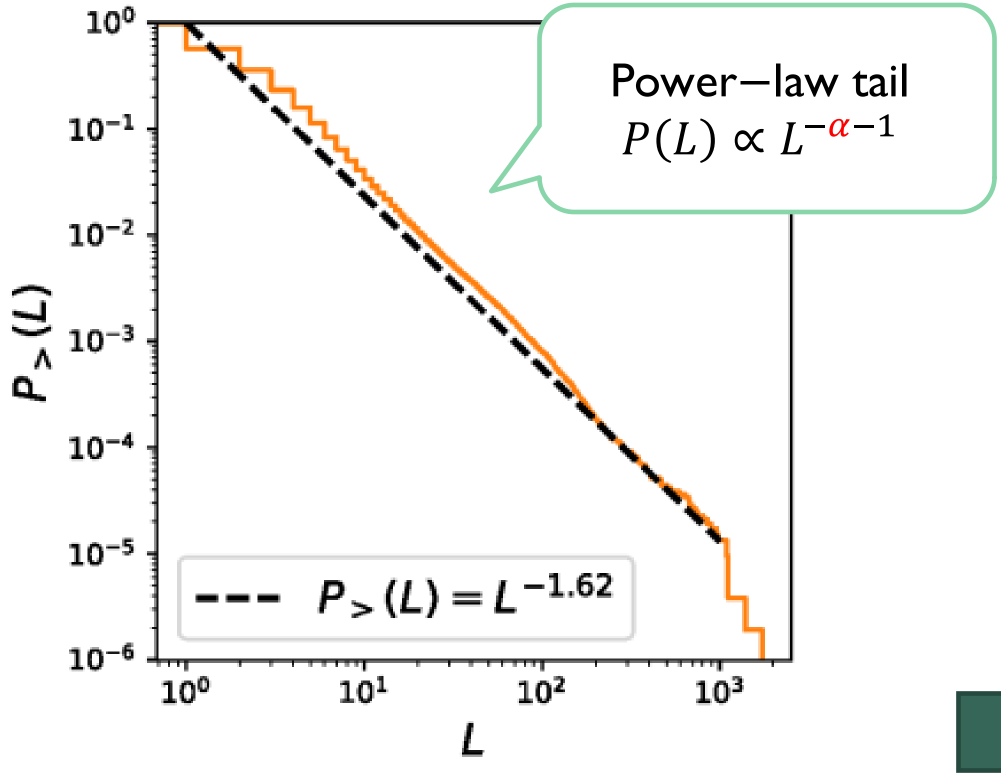
(b) Contribution by the splitting traders



- Typically, 80% of the market orders are submitted by the splitting traders

Splitting traders are actually present, and exhibiting the major contribution to the market orders

Result 2: Run-length dist. for splitting traders $P(L)$ to measure the microscopic power-law exponent α



Note: Toyota (7203) 2020

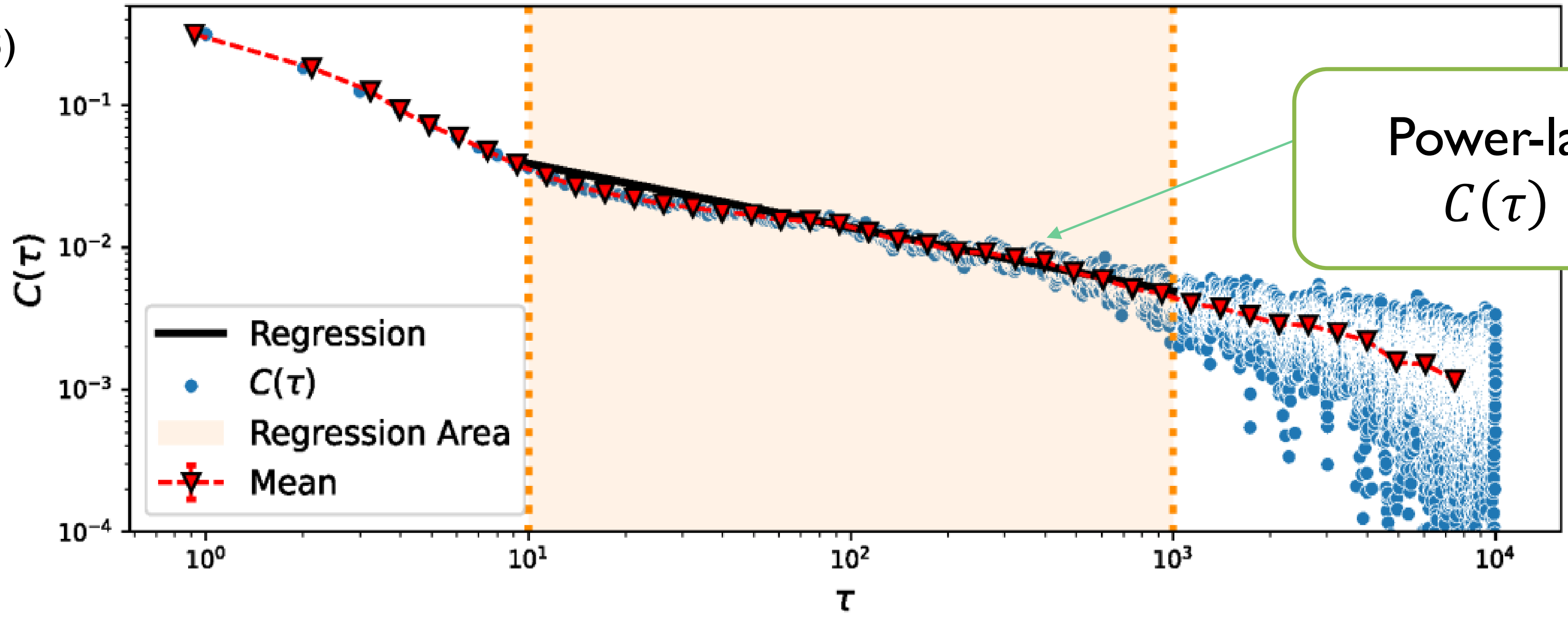
- Run-length dist. for splitting traders ⇒power-law dist. as expected
- The microscopic assumption in the LMF model was precisely verified
- Clauset's algorithm is used to estimate α
- Every stock every year

Microscopic parameter α is measured
⇒ Next, we measure the macroscopic parameter γ

Ref: J. Alstott, E. Bullmore, D. Plenz D, PLoS ONE **9**, e85777 (2014)

Result 2: measurement of the macroscopic parameter γ, the power-law decay of the ACF

TOYOTA(7203)
in 2020

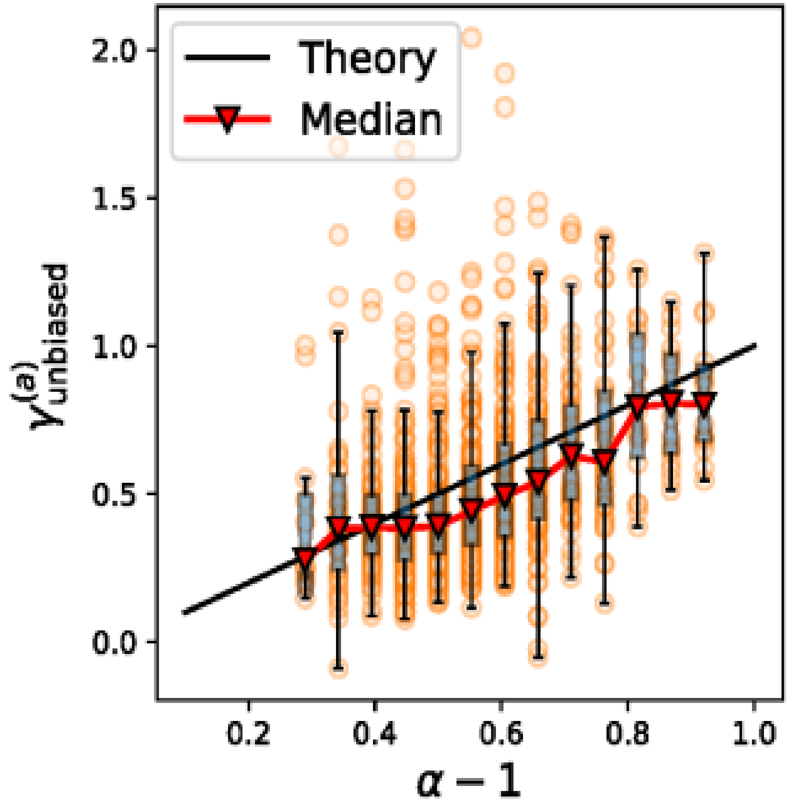


Power-law decay
$C(\tau) \propto \tau^{-\gamma}$

- Measurement of the macroscopic parameter γ
  ⇒Log-log ACF plot confirms its power-law decay

Nonlinear least squares + removal of statistical finite-size bias (in appendix)

Result 3: confirmation of the LMF model from our microscopic data (scatterplot between γ and α)



- Scatterplot for $(\alpha, \gamma_{\text{unbiased}})$ based on the unbiased estimator γ_{unbiased}
- Agreement between the theory and real data:

$$\gamma = \alpha - 1$$

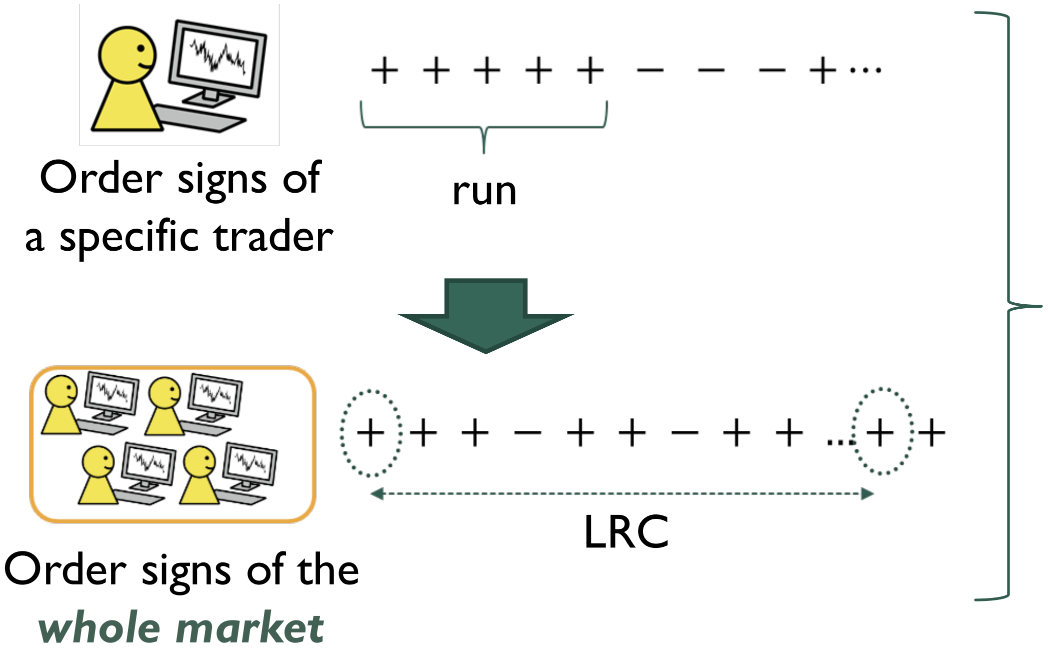
Method

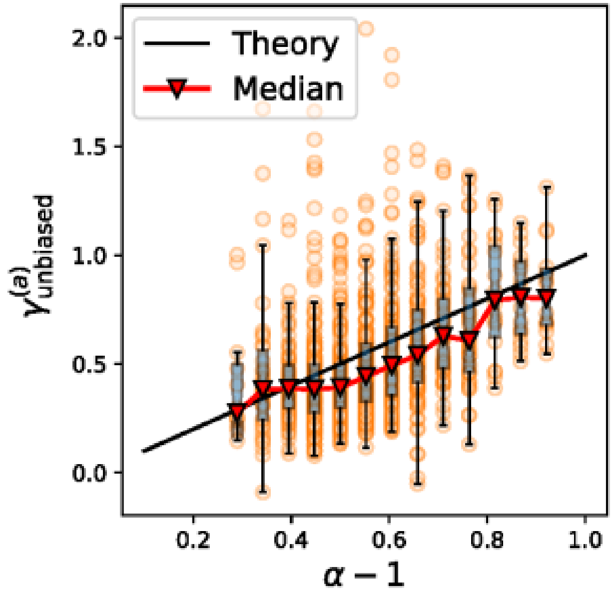
1. 1 datapoint = yearly 2012-2020 for 1 stock
2. Filter 1: $N > 0.5$ million transactions
3. Filter 2: $\alpha - 1 < 1$

Confirmation of the LMF prediction $\gamma = \alpha - 1$ even at the *quantitative* level

- Y. Sato and KK, Phys. Rev. Lett. **131**, 197401 (2023)
- Y. Sato and KK, Phys. Rev. Res. **5**, 043131 (2023)
- Y. Sato and KK, J. Stat. Phys. **191**, 58 (2024)

Conclusion 1: the quantitative prediction of the LMF model is *quantitatively* confirmed





Data analysis: $\gamma = \alpha - 1$

- Strategy clustering to identify splitting traders
- The LMF prediction is confirmed $\gamma_{\text{unbiased}} \approx \alpha - 1$
- Long-standing problem is solved supporting the order-splitting hypothesis

PHYSICAL REVIEW LETTERS

Highlights | Recent | Accepted | Collections | Authors | Referees | Search | Press | About

Featured in Physics | Open Access

Inferring Microscopic Financial Information from the Long Memory in Market-Order Flow: A Quantitative Test of the Lillo-Mike-Farmer Model

Yuki Sato and Kiyoshi Kanazawa
Phys. Rev. Lett. **131**, 197401 – Published 8 November 2023

Physics See Viewpoint: Decoding the Dynamics of Supply and Demand

- Highlighted in *Physics* by APS as a *Viewpoint* article
- Short introductory article available written by Prof. Lillo

Physics NEWS AND COMMENTARY

Decoding the Dynamics of Supply and Demand

November 8, 2023

An analysis of data from the Tokyo Stock Exchange provides the first quantitative evidence for the Lillo-Mike-Farmer model—a long-standing theory in economics.

Viewpoint on:

Yuki Sato and Kiyoshi Kanazawa
Phys. Rev. Lett. **131**, 197401 (2023)

Yuki Sato and Kiyoshi Kanazawa
Phys. Rev. Research **5**, 043131 (2023)

Physics ABOUT BROWSE PRESS COLLECTIONS Search articles

VIEWPOINT

Decoding the Dynamics of Supply and Demand

Fabrizio Lillo
Department of Mathematics, University of Bologna, Bologna, Italy
Faculty of Sciences, Scuola Normale Superiore, Pisa, Italy

November 8, 2023 • *Physics* 16, 192

An analysis of data from the Tokyo Stock Exchange provides the first quantitative evidence for the Lillo-Mike-Farmer model—a long-standing theory in economics.