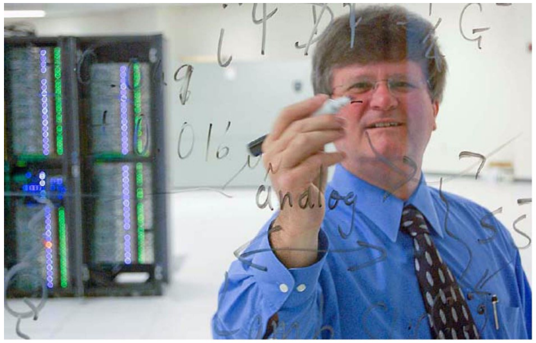
Universal Behaviour in the Stock Market

AYSE KIZILERSU

(University of Adelaide)

with



A.W. Thomas (University of Adelaide)

Markus Kreer (University of Frankfurt)

11-16 December 2022

ADELAIDE

AIP- AUSTRALIAN INSTITUTE OF PHYSICS CONGRESS

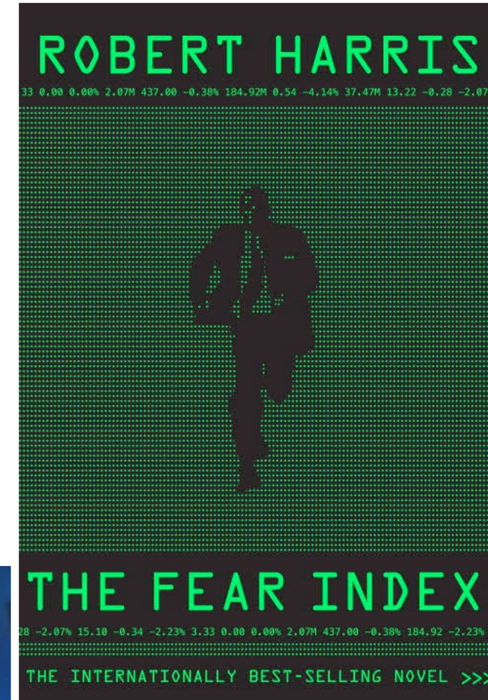
High Frequency Algorithm Trading

Receiving Information

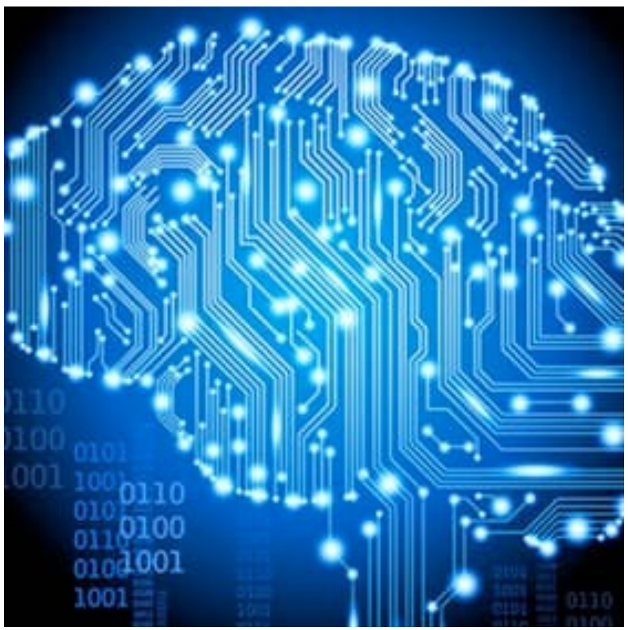
Nano second time scale

Algorithms respond to information

Electronic Order Book



Processing Speed

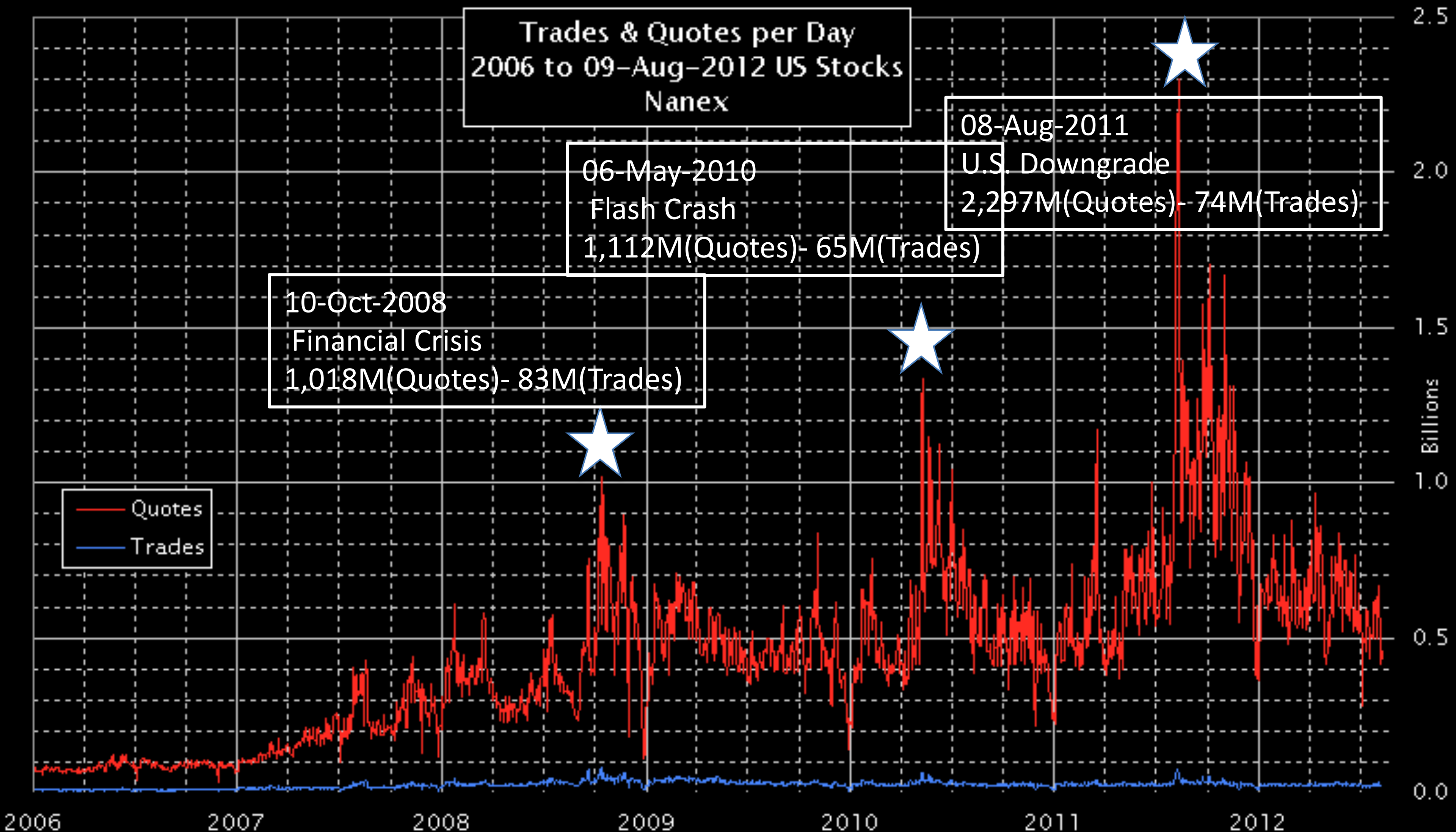


2000

year

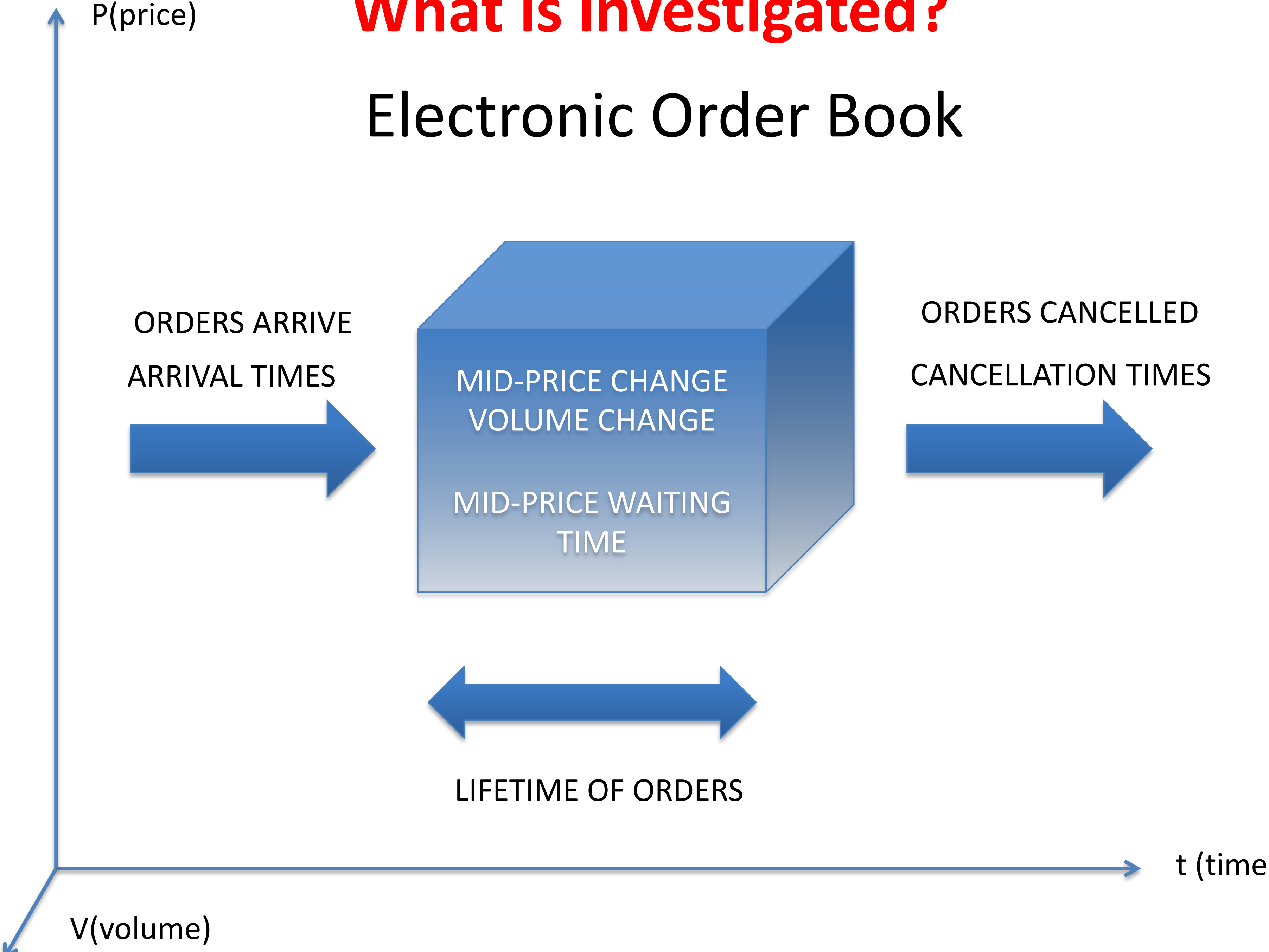
TOP 5 Sites 2022

Rank	Site	System	Cores	Rmax (TFlop/s)
1 USA	**FRONTIER**, Hewlett Packard Enterprise (HPE), the Oak Ridge National Laboratory (ORNL), USA 10^{18} calculations/sec		**8,335,360**	**1.102 exaFLOPS / 1.685 exaFLOPS**
2 JAPAN	**FUGAKU**, FUJITSU, Riken Center for Computational Science in Kobe, Japan		**7,630,848**	**442 PFLOPS**
3 FINLAND	**LUMI**, Hewlett Packard Enterprise (HPE), Finland		**1,110,144**	**375 PFLOPS**
4 USA	**SUMMIT**, IBM, the Oak Ridge National Laboratory (ORNL), USA		**2,414,592**	**200 PFLOPS**
5 USA	**SIERRA**, the Lawrence Livermore National Laboratory, USA		**1,572,480**	**125 PFLOPS**


Trades & Quotes per Day
2006 to 09-Aug-2012 US Stocks
Nanex
10-Oct-2008
Financial Crisis
1,018M(Quotes)- 83M(Trades)
06-May-2010
Flash Crash
1,112M(Quotes)- 65M(Trades)
08-Aug-2011
U.S. Downgrade
2,297M(Quotes)- 74M(Trades)
Quotes
Trades
Billions
0.0
0.5
1.0
1.5
2.0
2.5
2006
2007
2008
2009
2010
2011
2012

What is investigated?

Electronic Order Book



Analysis of tick-by-tick Data from the London Stock Exchange

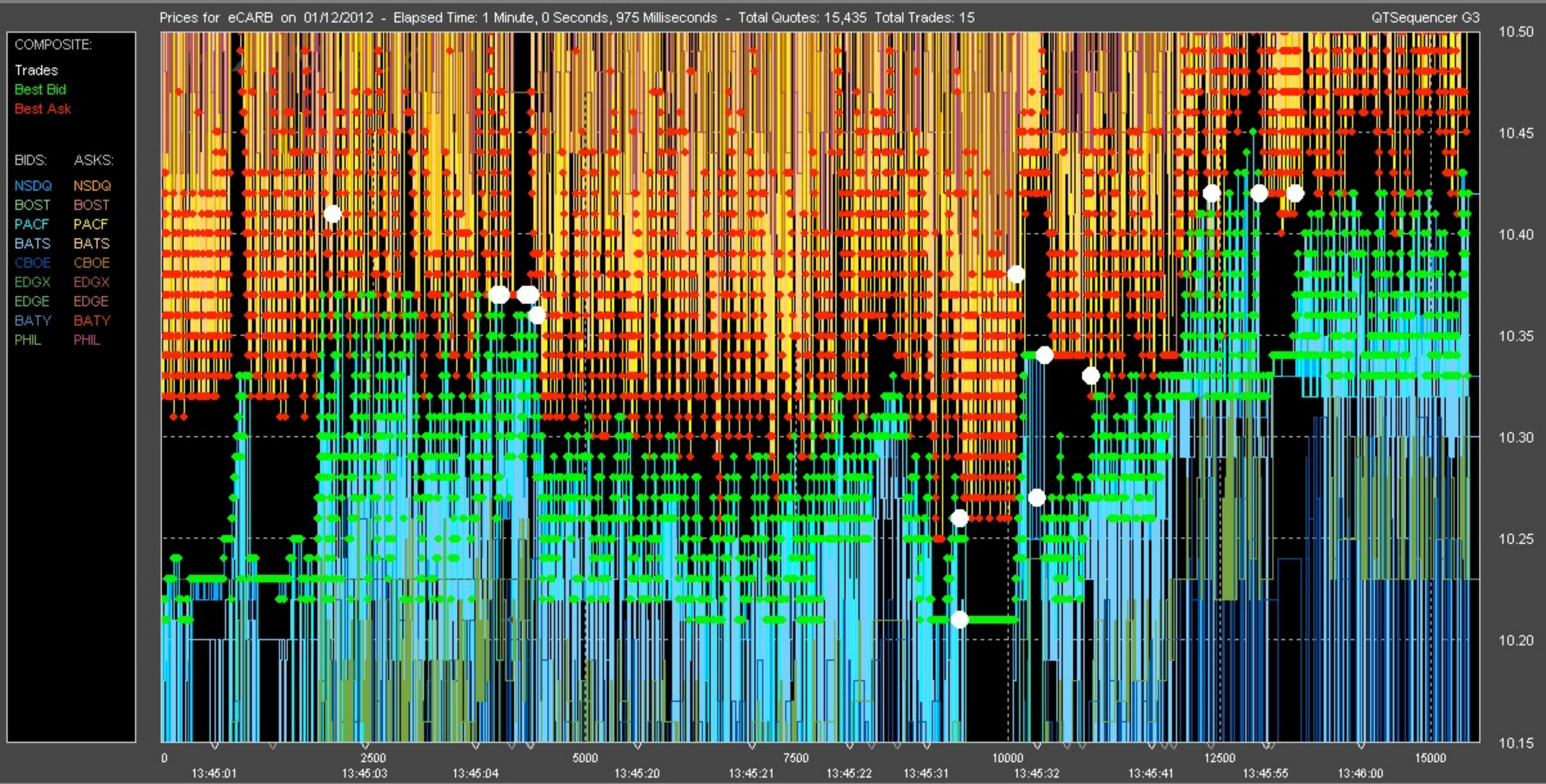
Electronic Order Book Components

Analyzing time components of electronic order book

- Limit order (LO) ARRIVAL time difference
- LO CANCELLATION time difference
- LO LIFETIME difference
- Market Order (MO) ARRIVAL time difference
- MID-PRICE WAITING time difference

CARB stock one day (tick chart)

Asks



Bids

TRUNCATION

- Zero Inflation
- Time Delay
- Rounding Error – discretization of continuous time

Data: Descriptive Statistics

	Ticker	τ_L = 0				τ_L = 11	
		# LO's	max Δt [ms]	average Δt [ms]	# Zero-inflated	# LO's	average Δt [ms]
1. June	RIO	394,663	19,978	78	227,912	101,609	299
Average June	RIO	280,669	36,169	119	159,596	75,896	429
1. June	BARC	240,702	21,555	127	133,880	67,271	453
Average June	BARC	236,964	41,062	141	122,348	69,150	468
1. June	VOD	105,416	40,324	290	56,234	37,577	813
Average June	VOD	102,968	60,768	310	58,433	31,862	992
1. June	RRLN	47,371	62,871	646	27,557	16,001	1,911
Average June	RRLN	36,033	92,582	885	18,977	13,405	2,376
1. June	SSELN	38,110	48,108	803	20,094	15,155	2,018
Average June	SSELN	28,737	106,536	1,117	15,119	11,284	2,843
1. June	ABFLN	40,622	49,643	753	20,970	16,674	1,834
Average June	ABFLN	25,569	115,257	1,326	13,618	9,728	3,447
1. June	YELLLN	30,491	262,126	1,002	17,614	8,127	3,758
Average June	YELLLN	23,212	360,849	1,535	12,533	6,638	5,304

ANALYSIS METHODS

- We wish to describe the data (time differences) for all time scales: small and large.

Candidate Models

- Weibull
- Log-Normal
- Log-Logistic
- Gamma
- Pareto

Statistical tools for parameter estimation

- Maximum likelihood estimators
- Bayesian estimators (Expectation- Maximisation method)

Machine Learning Algorithm : Unsupervised learning/teaching

GoF tests

Frequentist

- Kolmogorov–Smirnov test
- Cramér–von Mises test
- Anderson–Darling test
- Chi Square test
- KUIPIER'S test

Bayesian

- Akaike Information Criterion (AIC)
- Bayesian Information Criterion (BIC)

Model Sellection

Left-truncated Weibull Distribution and Likelihood

PDF :

$$f(\tau|\alpha,\beta,\tau_L) = \frac{\beta}{\alpha}\left(\frac{\tau}{\alpha}\right)^{\beta-1} \exp\left[\left(\frac{\tau_L}{\alpha}\right)^{\beta} - \left(\frac{\tau}{\alpha}\right)^{\beta}\right] \quad \text{for} \quad \tau > \tau_L\,.$$

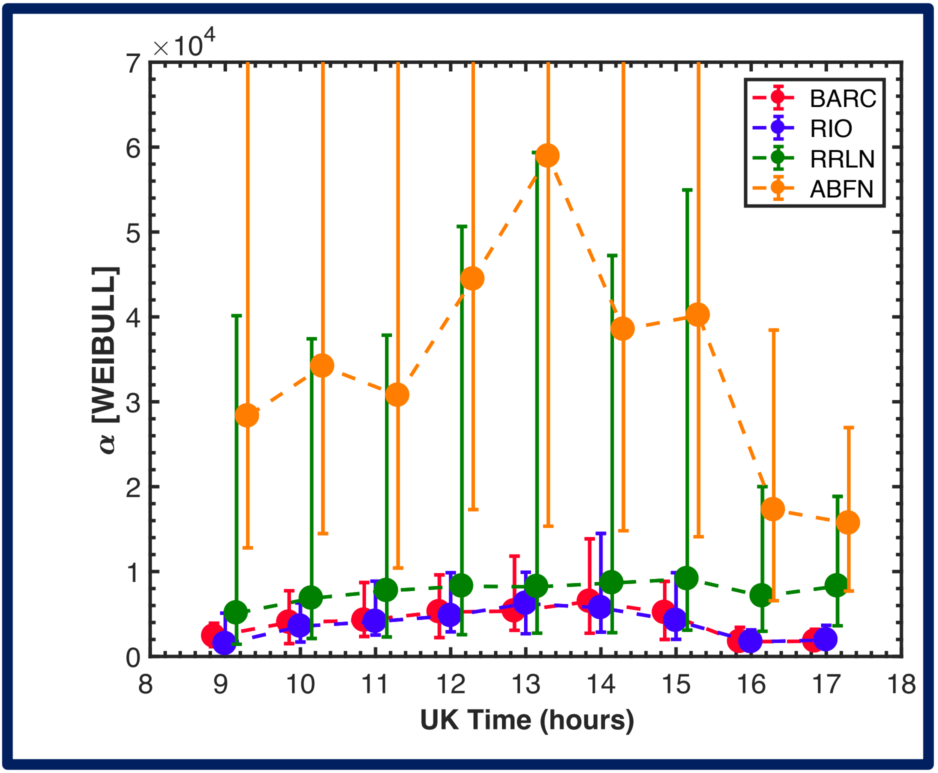
CDF:

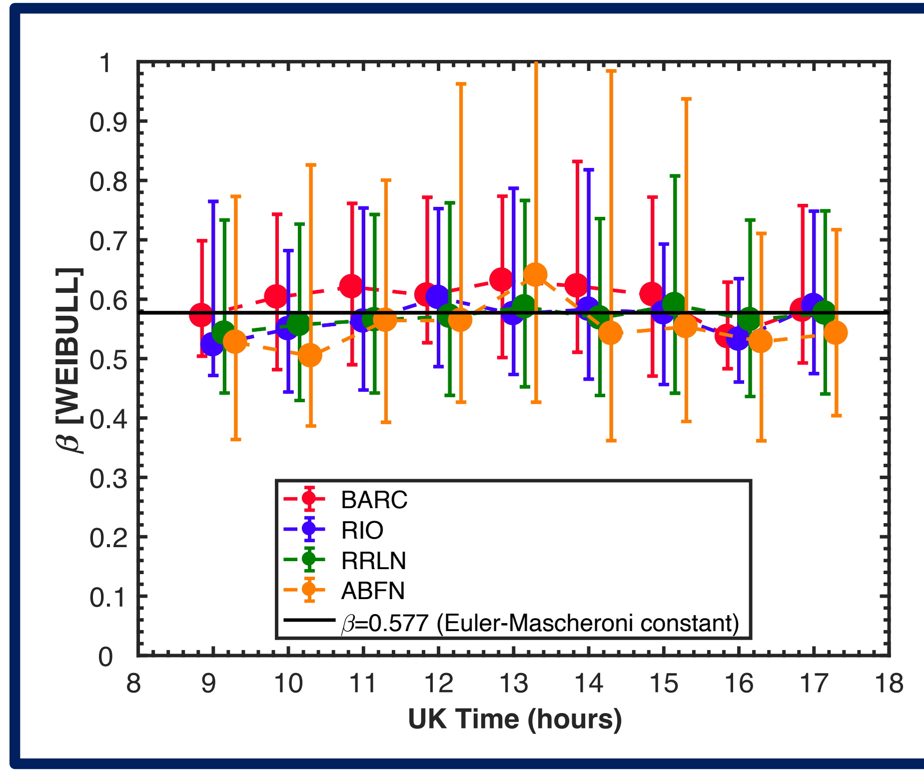
$$F(\tau|\alpha,\beta,\tau_L) = 1 - \exp\left[\left(\frac{\tau_L}{\alpha}\right)^{\beta} - \left(\frac{\tau}{\alpha}\right)^{\beta}\right] \quad \text{for} \quad \tau > \tau_L$$

LOG-LIKELIHOOD

$$\begin{aligned} L_{trunc}(\tau_1,\tau_2,\ldots,\tau_n|\alpha,\beta,\tau_L) &= \prod_{i=1}^{n} \frac{\beta}{\alpha}\left(\frac{\tau_i}{\alpha}\right)^{\beta-1} e^{\left(\frac{\tau_L}{\alpha}\right)^{\beta} - \left(\frac{\tau_i}{\alpha}\right)^{\beta}} \\ &= \log L\left(\boldsymbol{\tau}|\alpha,\beta,0\right) + n\left(\frac{\tau_L}{\alpha}\right)^{\beta} \end{aligned}$$

INTERTRADE WAITING TIMES

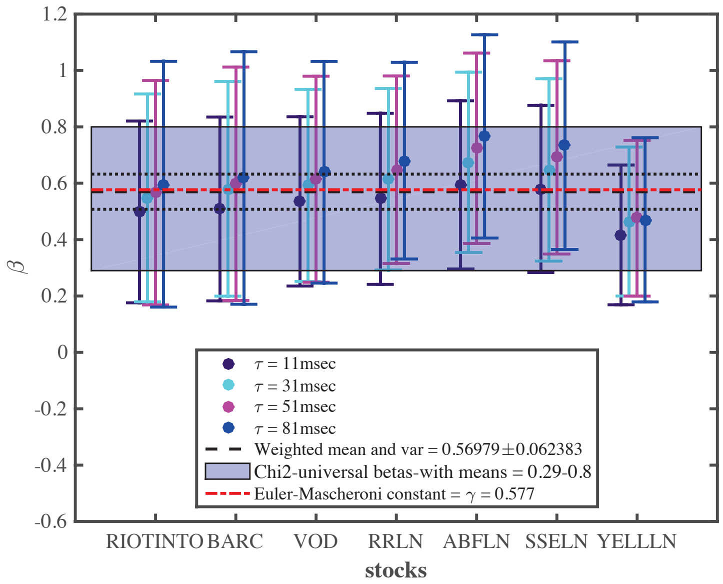




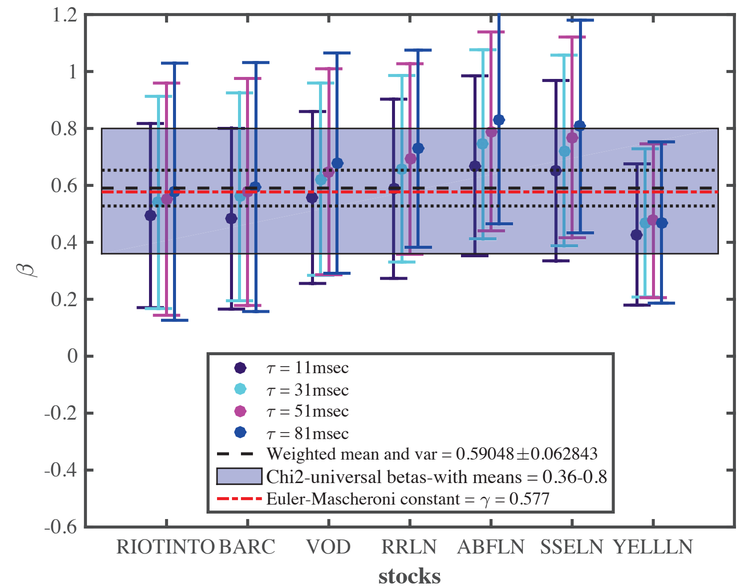
M.Kreer, A. Kizilersu and A.W.Thomas, “Censored expectation maximization algorithm for mixtures: Application to intertrade waiting times”, Physica A 587 (2022) 126456

Weibull beta for Time Differences

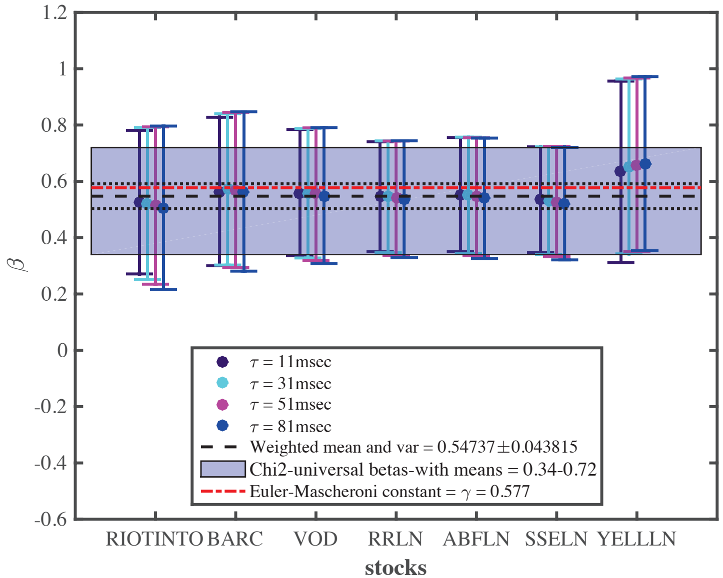
ARRIVAL TIMES



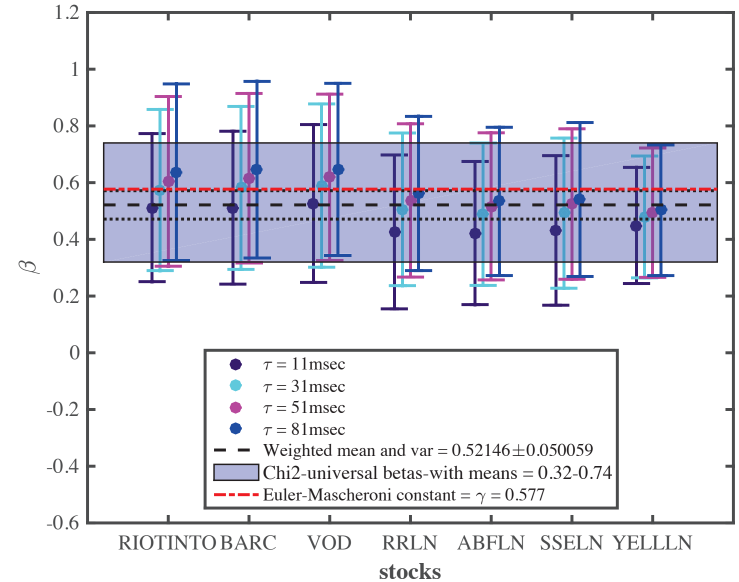
CANCELLATION TIMES



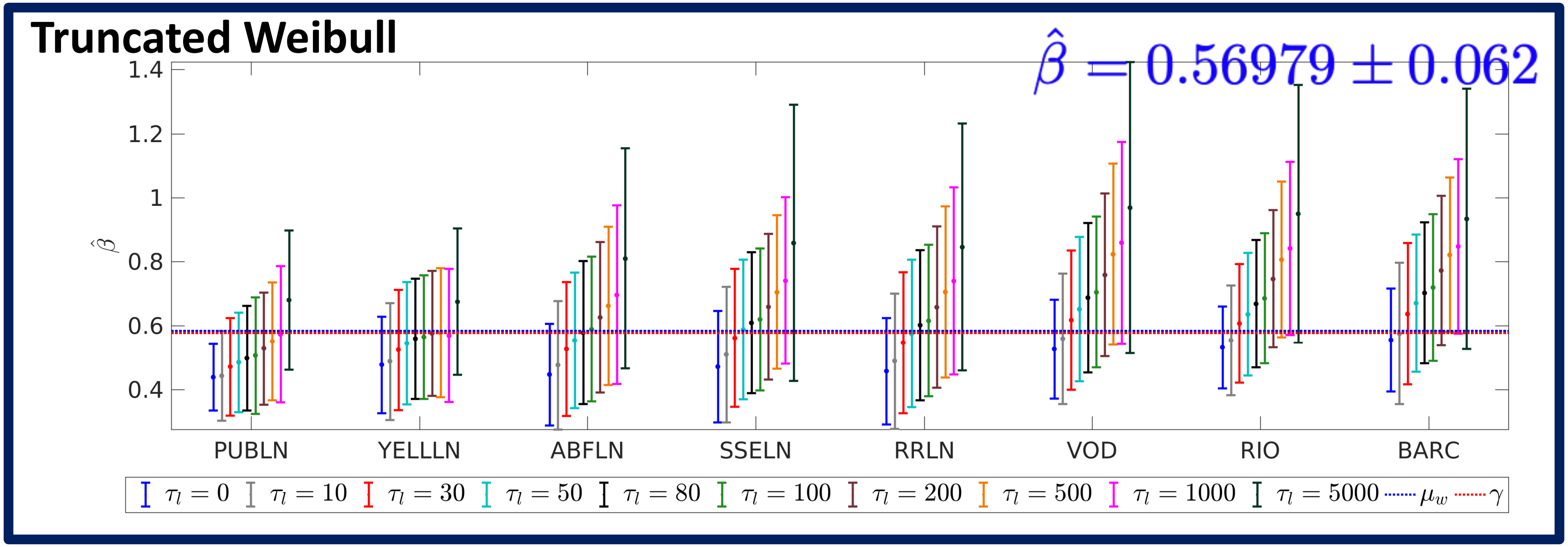
LIFETIMES

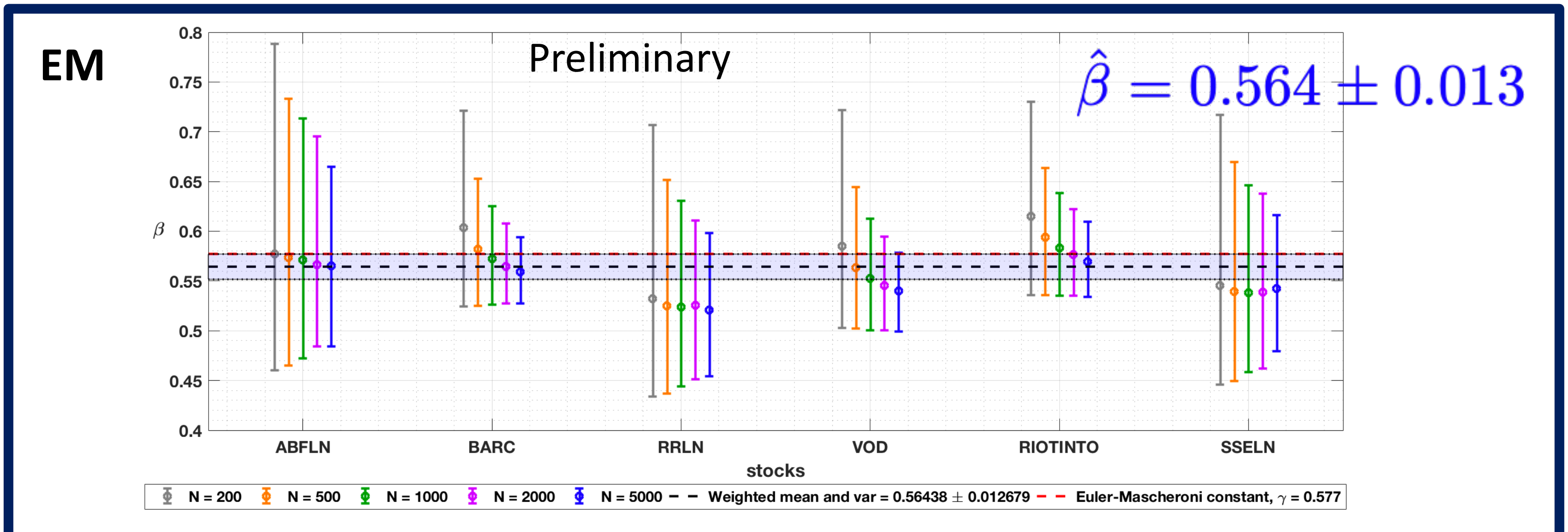


MO ARRIVAL TIMES



Comparison of EM and Truncated Weibull





Shannon/Information Entropy

- Entropy is a measure of uncertainty
  - The greater the entropy the less you can predict the outcome
- If the probability density function is known, the continuous entropy is defined as

$$H \equiv -\int f(x) log f(x) dx$$

- For the Gaussian distribution

$$H = log\left(\sigma\sqrt{2}\pi e\right)$$

the entropy increases monotonically with variance

- For the Weibull distribution

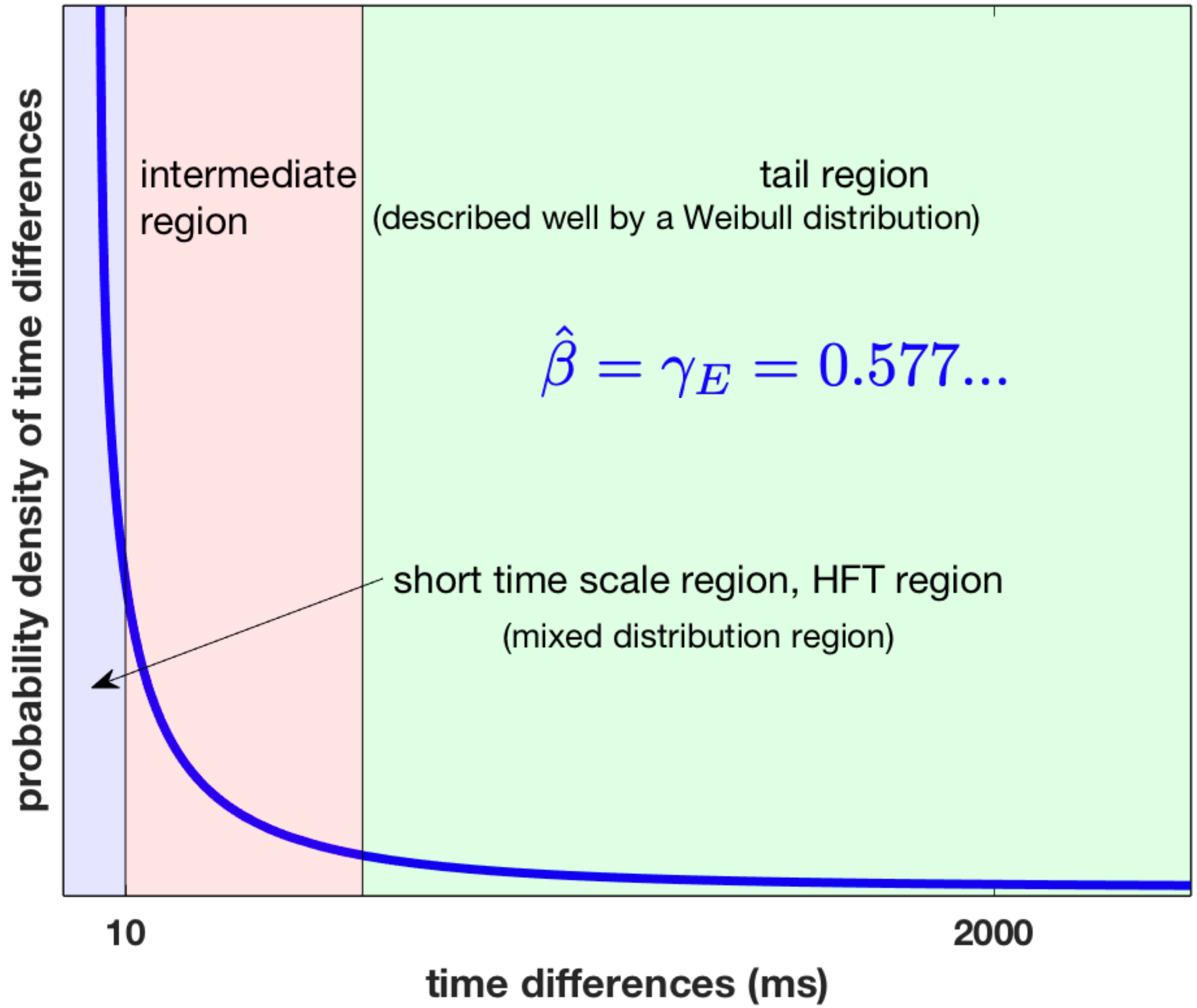
$$H = \frac{(\beta - 1)}{\beta}\gamma_E + log\frac{\alpha}{\beta} + 1$$

where $\gamma_E = 0.577\ldots$ is Euler–Mascheroni constant.

The Weibull distribution has maximum entropy when

$$\boxed{\beta = \gamma_E \quad for\ all \quad \alpha}$$

(30%) Exponential /(15%) Exponential/ (50%) Weibull Mixture



Conclusion Stock Market Time Series Analysis

- Ultra-high frequency manipulation (activity) occurs at < 10msec
- For time > 10 msecs the EOB is described by **left-truncated Weibull distribution**
- The shape parameter of the Weibull distribution is constant (**$\beta = \gamma_E$ = 0.577=Euler-Mascheroni constant) and universal**
- The universal shape parameter corresponds to maximum entropy of the time series distribution

HISTORY !