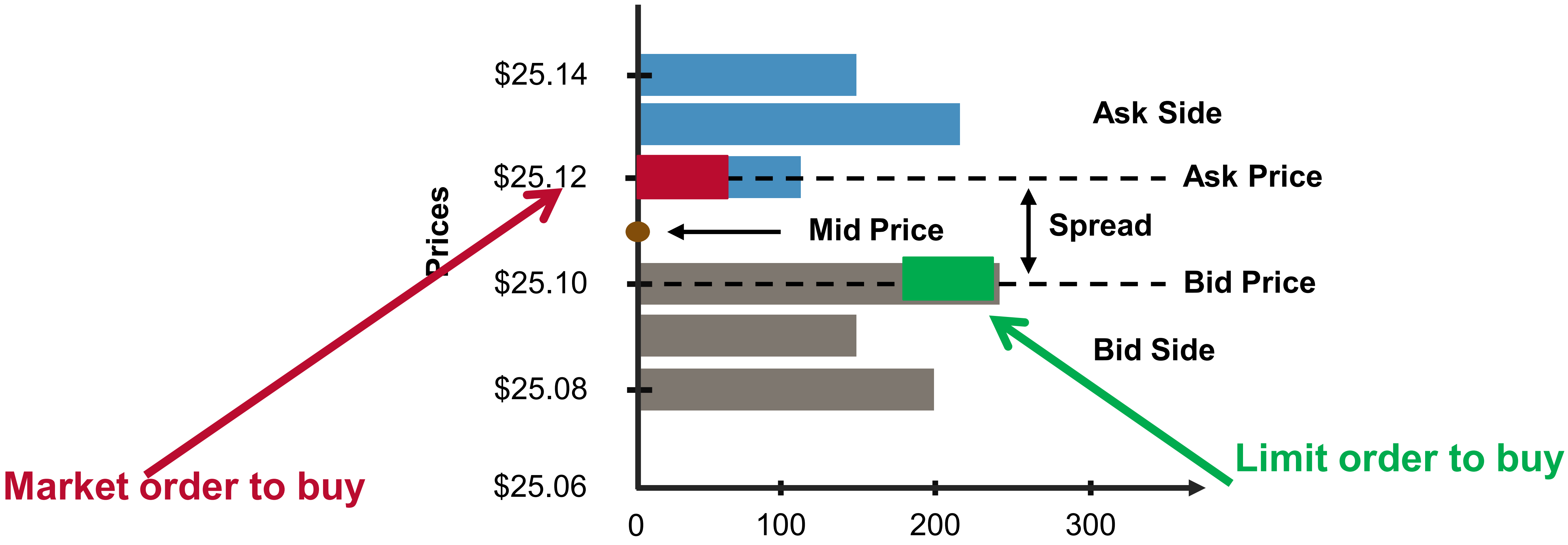
INDENG 290

Lecture 2

Market Impact

- Market impact is the effect that the market participant has on the market due to its trading activity
  - Market orders have high market impact, “peel market away’
- Market impact grows with the size of the order
  - In practice it’s common to “slice” the order to “space out” execution



Transaction cost model

- Client is placing an order size X to the market wants to estimate its average transaction cost (spread + market impact)

How to derive what market impact of the trade will be from historical data

Tick Data Example

	A	B	C	D	E
1	Time	Bid	BidVol	Ask	AskVol
2	2018.03.05 13:49:00.674	1.23041	50	1.23042	25.1
3	2018.03.05 13:49:02.716	1.23041	75.1	1.23043	75.1
4	2018.03.05 13:49:02.900	1.23041	100.1	1.23043	75.1
5	2018.03.05 13:49:03.060	1.23042	25	1.23044	80
6	2018.03.05 13:49:03.541	1.23042	25	1.23043	25.1
7	2018.03.05 13:49:03.887	1.23042	25	1.23043	75.1
8	2018.03.05 13:49:04.318	1.23042	25	1.23043	74.9
9	2018.03.05 13:49:04.770	1.23042	24.9	1.23043	74.9
10	2018.03.05 13:49:05.300	1.23041	50	1.23043	75.1
11	2018.03.05 13:49:05.733	1.23042	25	1.23043	50
12	2018.03.05 13:49:06.284	1.23042	25	1.23043	75.1
13	2018.03.05 13:49:06.455	1.23041	50	1.23043	75.1
14	2018.03.05 13:49:06.973	1.23042	25	1.23043	50

Empirical Market Impact Estimation Algorithm ("Get Real" paper)

0.1 Stylized facts about order market impact

Market impact of order placement is a expected to grow as a function of order volume. For each time interval τ, define $V_{\text{buy},\tau}$ and $V_{\text{ask},\tau}$ to be buy and sell order volumes in τ respectively. Define participation of volume in τ as

$$P_\tau = \frac{|V_{\text{buy},\tau} - V_{\text{ask},\tau}|}{V_{\text{buy},\tau} + V_{\text{ask},\tau}}.$$

Note that $0 \le P_\tau \le 1$. Also define Δm_τ to be the observable mid-price move in τ. Discretize the range for P_τ into bins $B_i, i = 1, \ldots, N$ such that $B_i = \{\tau : \frac{i-1}{N} \le P_\tau \le \frac{i}{N}\}$. For each B_i, define

$$M_i = \frac{1}{|B_i|} \sum_{\tau \in B_i} \Delta m_\tau \quad \text{and} \quad P_i = \frac{1}{|B_i|} \sum_{\tau \in B_i} \Delta P_\tau$$

to be the average price move and average participation of volume in bins with similar volume participation. One can then fit the relationship of the form $M_i \sim \alpha P_i^\beta$ through the data [1, 2, 3].

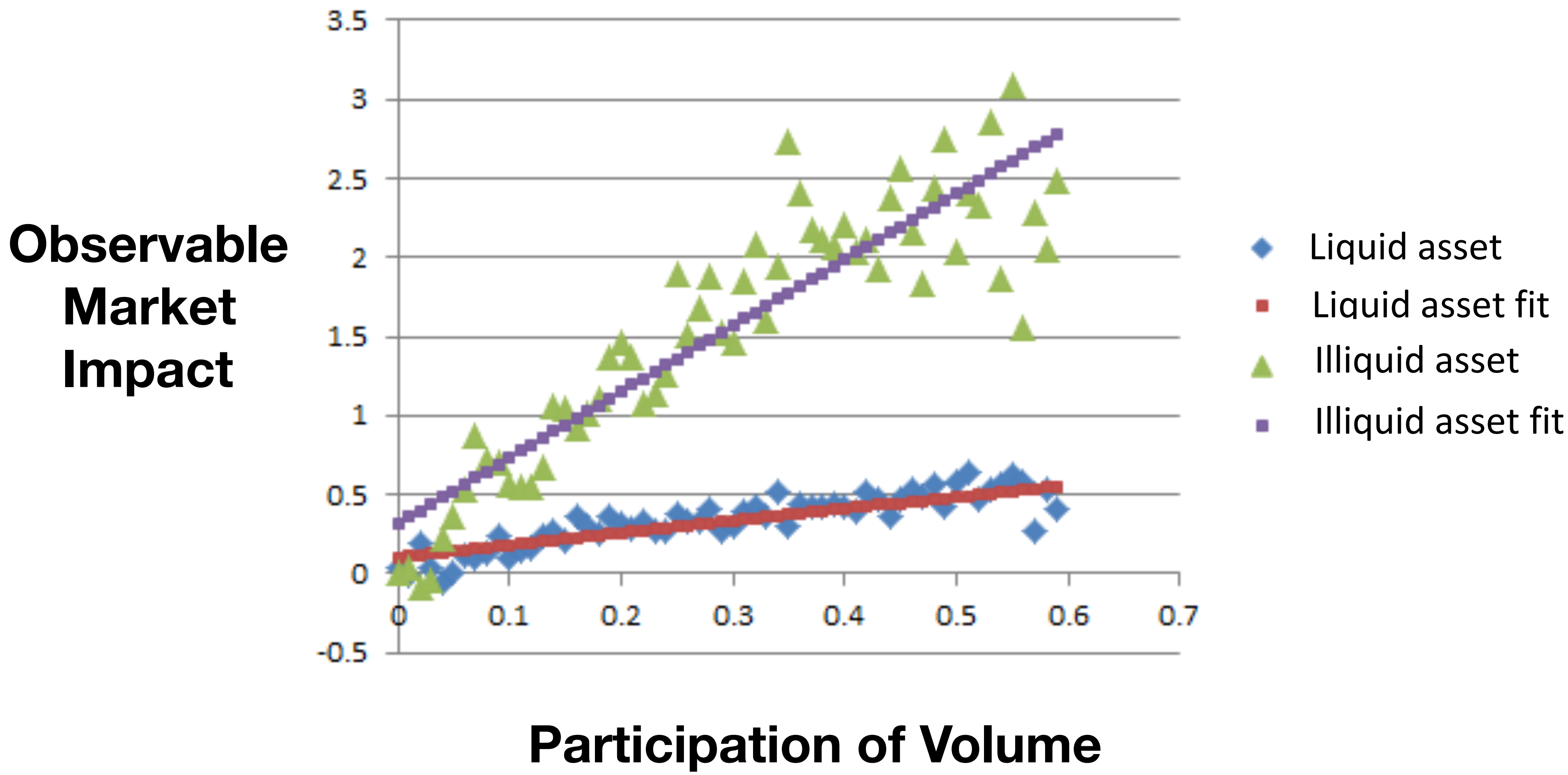
References

[1] Robert Almgren, Chee Thum, Emmanuel Hauptmann, and Hong Li. Direct estimation of equity market impact. *RISK*, 18, 04 2005.

[2] J Farmer, Paolo Patelli, and Ilija Zovko. The predictive power of zero intelligence in financial markets. *Proceedings of the National Academy of Sciences of the United States of America*, 102:2254–9, 03 2005.

[3] Jean-Philippe Bouchaud. Price impact. *Encyclopedia of quantitative finance*, 2010.

Empirical Market Impact Estimation



- A LOT OF market data is required to calibrate
- Non-constant variance of error terms
    - Weighted least squares
        - *i*th response has average of n_i observations with variance δ_i
        - Weight of *i*th response is inversely proportional to δ_i

Limit Order Book Data

- https://bcourses.berkeley.edu/courses/1518016/files/folder/LOB%20data

5224948A	34S	100APA	595000Y
5227220A	36B	780MSFT	564400Y
5238879A	39S	100LEA	402000Y
5242003X	36	780	
5242090A	40B	780MSFT	564400Y
5242542A	41S	1000SWY	271000Y
5250681A	42S	100AEG	159900Y
5256049A	43S	1250INTC	174960Y
5257095X	40	780	
5257113A	44S	1000ABT	392500Y
5257181A	45B	780MSFT	564400Y
5258156X	43	1250	
5258190A	46S	1250INTC	175040Y
5259749A	48B	1000COCO	359700Y
5260115A	49B	30QLGC	380300Y
5260125A	50S	30QLGC	443000Y
5260127A	51B	600PSFT	200400Y
5260129A	52B	100QCOM	355400Y
5260130A	53B	100MPET	07100Y
5260132A	54B	1POOL	250000Y
5260134A	55S	100MPET	09400Y
5260135A	56S	200MPET	09900Y
5260136A	57S	100MPET	10400Y
5260138A	58S	200MPET	10900Y
5260140A	59B	1MHK	525100Y
5260141A	60B	1100NT	21900Y

?

	A	B	C	D
1	Time	Bid	BidVol	Ask
2	2018.03.05 13:49:00.674	1.23041	50	1.23042
3	2018.03.05 13:49:02.716	1.23041	75.1	1.23043
4	.03.05 13:49:02.900	1.23041	100.1	1.23043
5	20 3.05 13:49:03.060	1.23042	25	1.23044
6	03.05 13:49:03.541	1.23042	25	1.23043
7	18.03.05 13:49:03.887	1.23042	25	1.23043
8	2018.03.05 13:49:04.318	1.23042	25	1.23043
9	18.03.05 13:49:04.770	1.23042	24.9	1.23043
10	2018.03.05 13:49:05.300	1.23041	50	1.23043
11	2018.03.05 13:49:05.733	1.23042	25	1.23043
12	2018.03.05 13:49:06.284	1.23042	25	1.23043
13	2018.03.05 13:49:06.455	1.23041	50	1.23043
14	2018.03.05 13:49:06.973	1.23042	25	1.23043

ITCH Protocol

- Market agents “talk” to exchange via messages
- ITCH is the outbound protocol NASDAQ uses to communicate market data to its clients, that is, all information including market status, orders, trades, circuit breakers, etc. with nanosecond timestamps for each day and each exchange.

Order Replace Message

Name	Offset	Length	Value	Notes
Message Type	0	1	“U”	Order Replace Message
Stock Locate	1	2	Integer	Locate code identifying the security
Tracking Number	3	2	Integer	Nasdaq internal tracking number
Timestamp	5	6	Integer	Nanoseconds since midnight
Original Order Reference Number	11	8	Integer	The original order reference number of the order being replaced
New Order Reference Number	19	8	Integer	The new reference number for this order at time of replacement Please note that the Nasdaq system will use this new order reference number for all subsequent updates
Shares	27	4	Integer	The new total displayed quantity
Price	31	4	Price (4)	The new display price for the order Please refer to Data Types for field processing notes

LOBSTER Data

- https://lobsterdata.com/

LOBSTER
Limit Order Book System – The Efficient Reconstructor

academic data.

LOBSTER? output. samples. access options. join. how? meet. docs. research.

sample files.

The sample files contain an 'orderbook' file, a 'message' file and a readme summarizing the data's properties. All sample files are based on the official NASDAQ Historical TotalView-ITCH sample.

AMZN_2012-06-21_342000

2239500	100	2231800	100
2239500	100	2238100	21
2239500	100	2237500	100
2239500	100	2237500	74
2239500	100	2237500	74
2239500	100	2237500	74
2239500	100	2237500	74

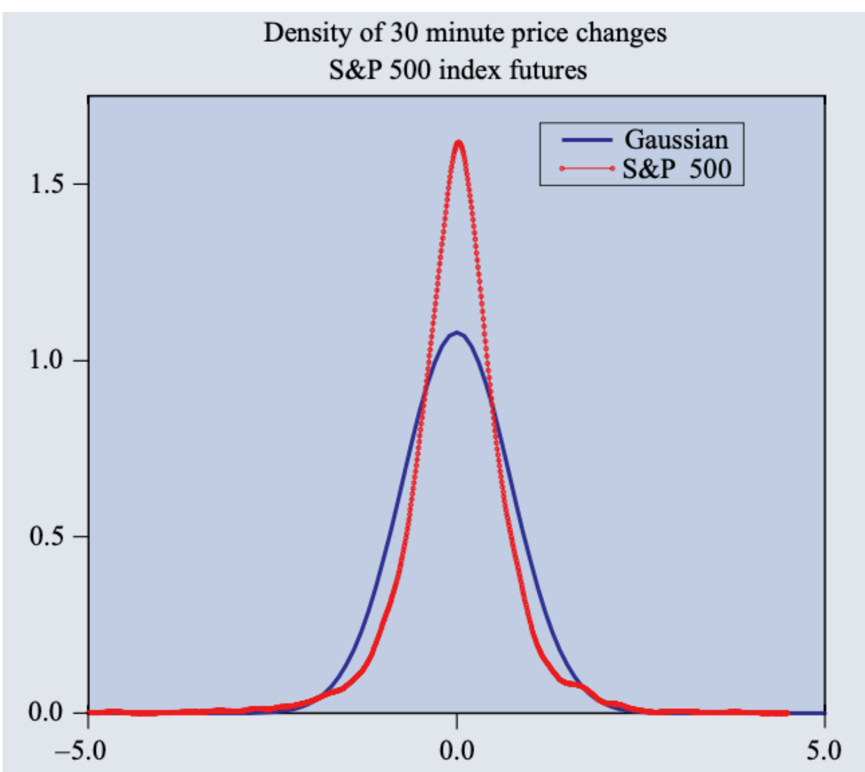
AMZN_2012-06-21_34200000_57600000_me

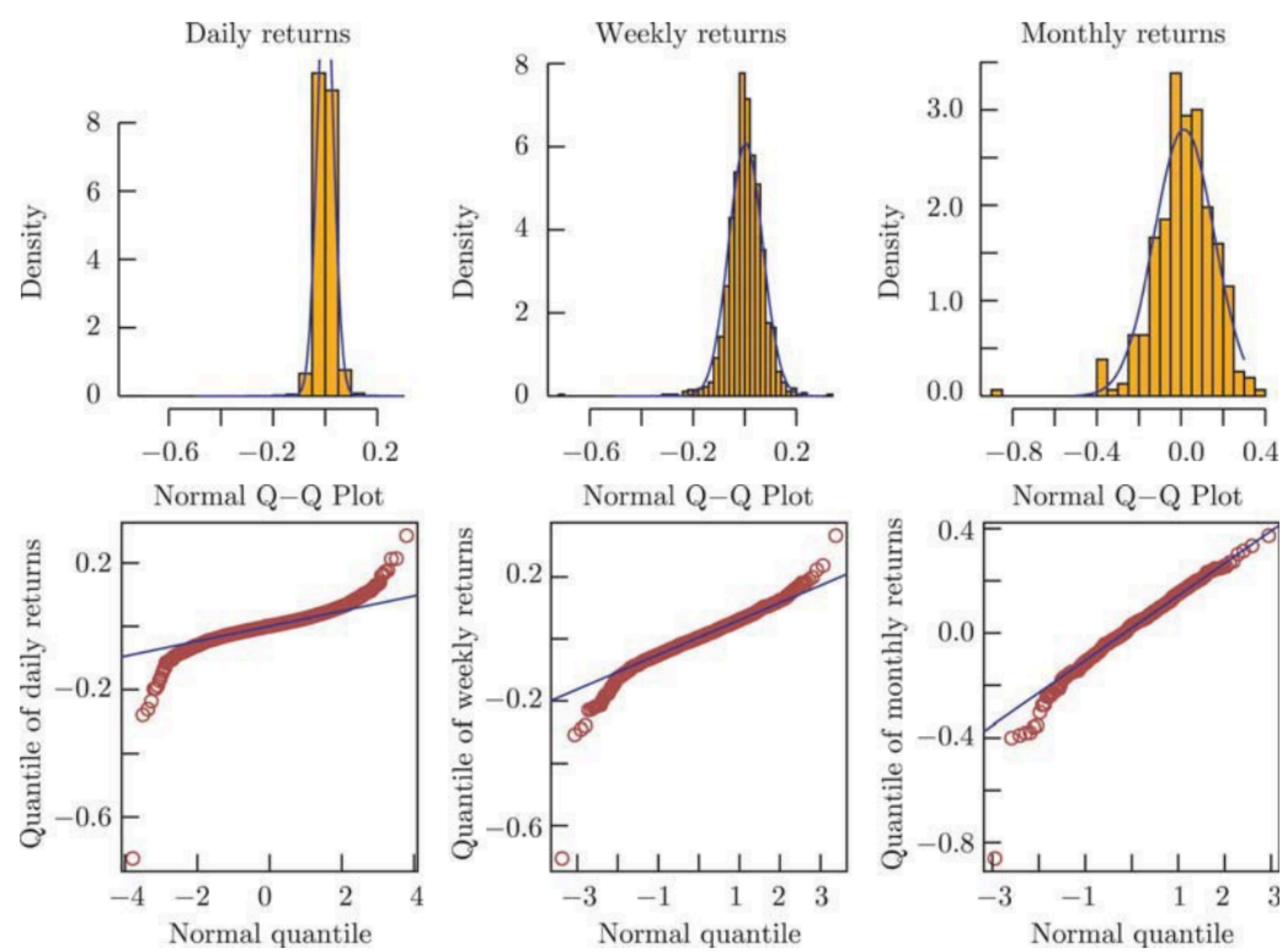
34200.017459617	5	0	1	2238200	-1
34200.18960767	1	11885113	21	2238100	1
34200.190226476	4	11885113	21	2238100	1
34200.190226476	4	11534792	26	2237500	1
34200.372779672	5	0	100	2238400	-1
34200.375671205	5	0	100	2238400	-1
34200.383971366	5	0	100	2238600	-1

Stylized Facts

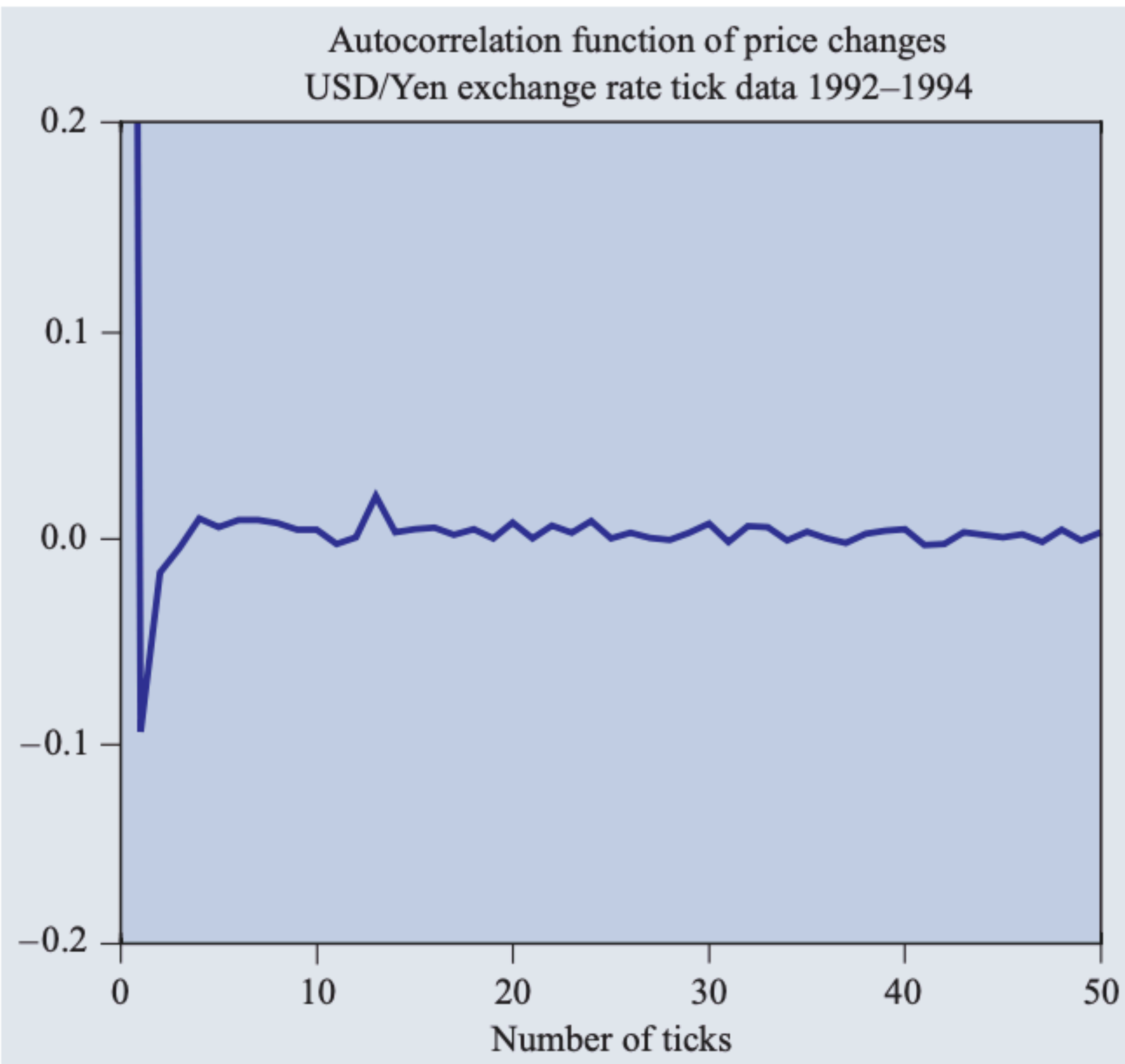
Asset Return Distributions

- The distribution of daily asset price returns shows fat tails
- As one increases the period of time Δt over which these returns are calculated, asset returns show lower tails
- One way to quantify deviation from normal distribution is to calculate its kurtosis.





Asset Return Distributions: Absence of Autocorrelations



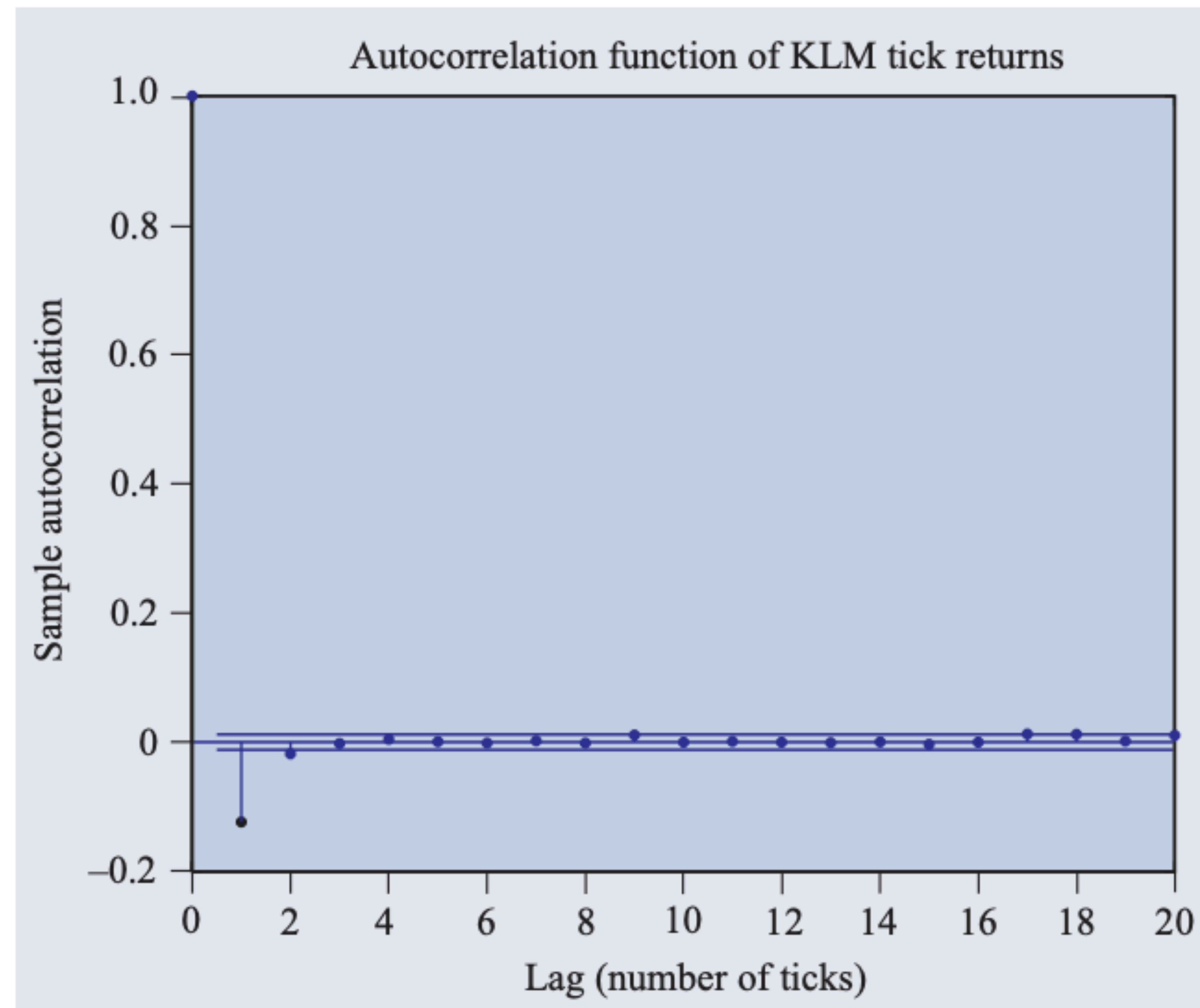
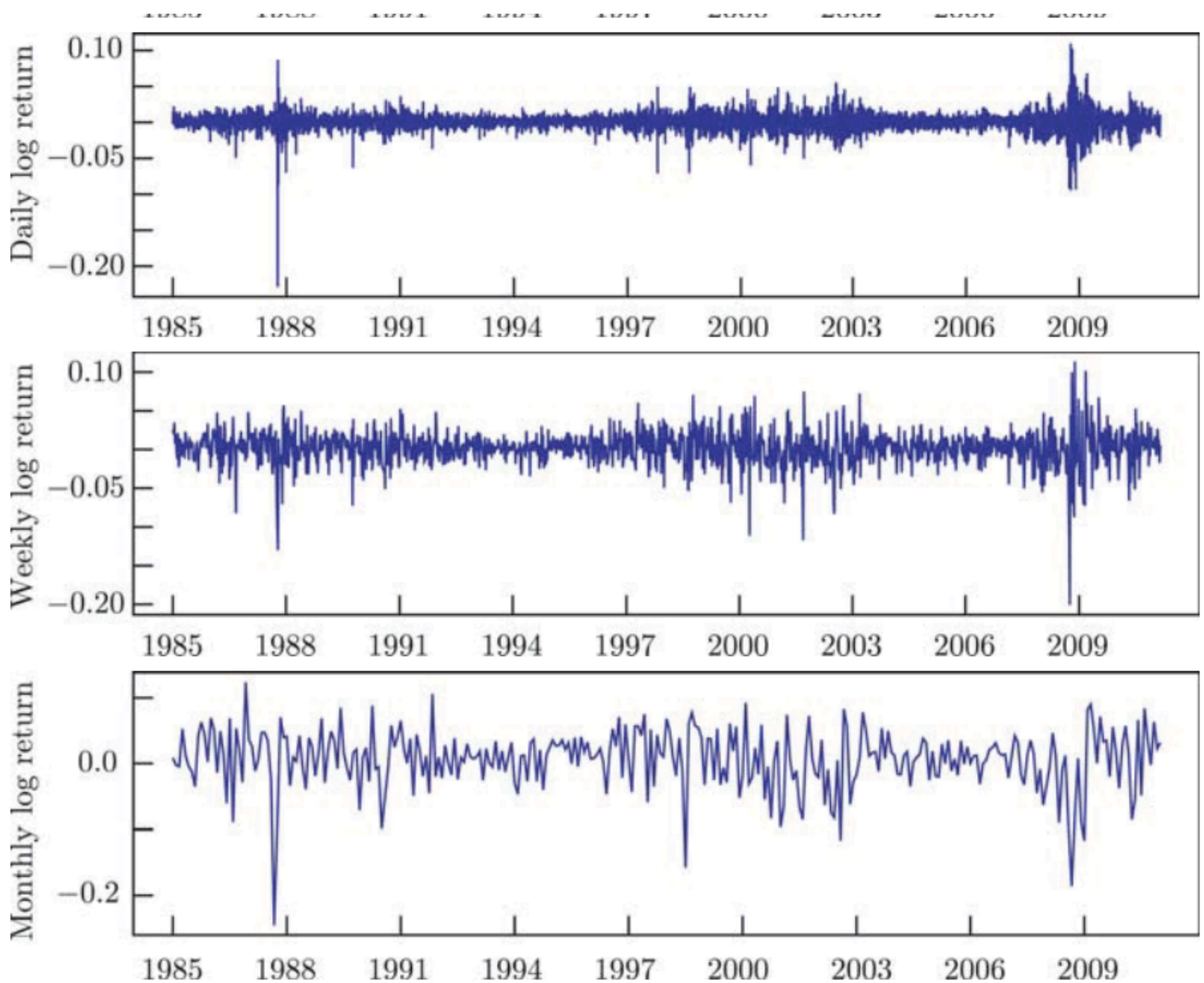


Figure 7. Autocorrelation function of tick by tick returns on KLM shares traded on the NYSE. Time scale: ticks.

Volatility Clustering

- High volatility events tend to cluster in time.



Returns and Volatility are Negatively Correlated

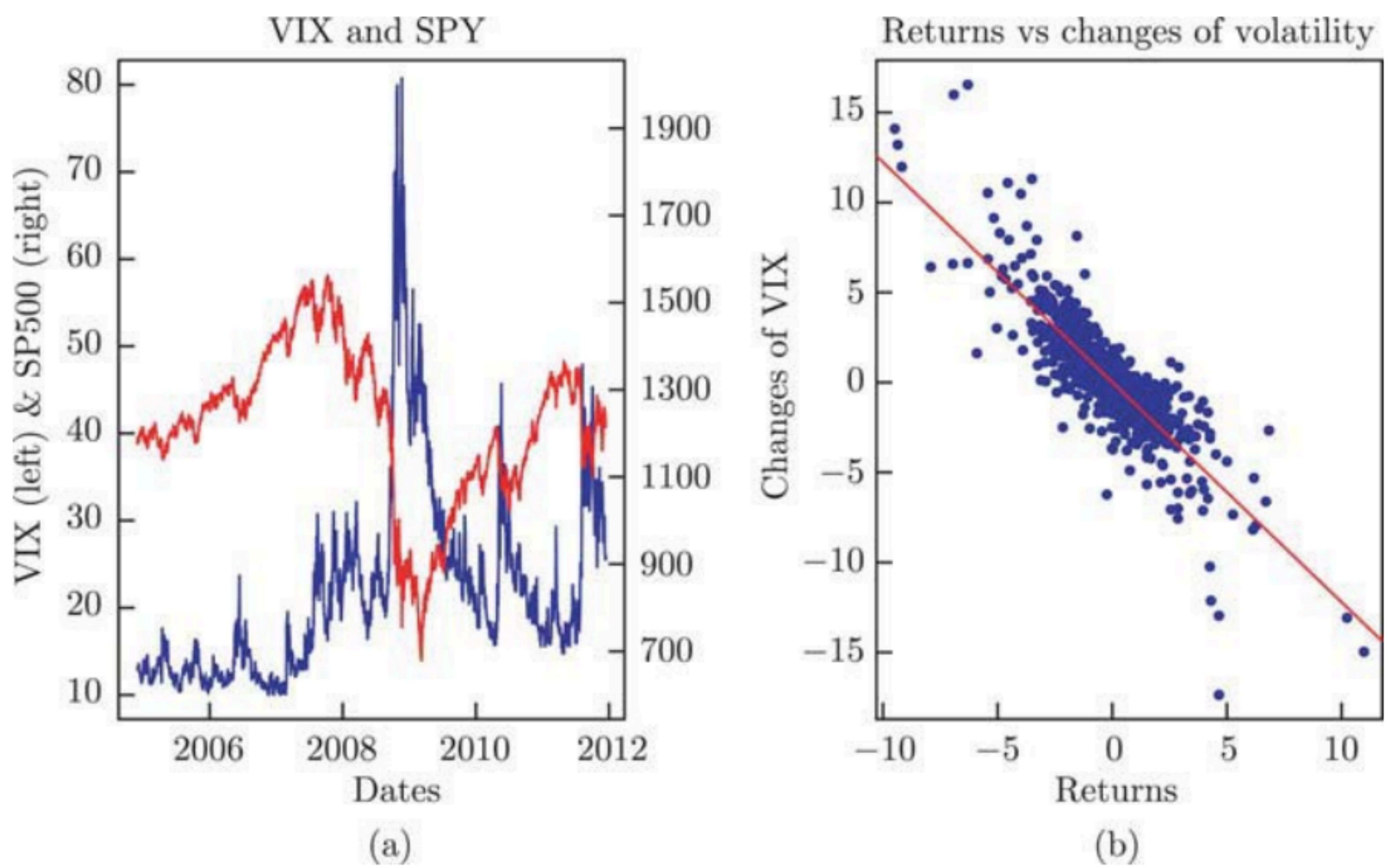


Figure 1.10 Time series plot of VIX (red) and the S&P 500 index (blue) in Nov. 29, 2004 – Dec. 14, 2011 (the left panel), and the plot of the daily S&P 500 returns (in percent) against the changes of VIX (the right panel).

Stylized Facts

- Similarity of statistical behaviors **across assets**.
- Properties of market behavior that are repeated across a wide range of instruments, markets and time are referred to as **stylized facts.**
- Can use stylized facts as a quantitative metric of realism to assess the quality of simulator.

Other Stylized Facts About Asset Return Distributions

- **Intermittency** At any micro or macro time scale, asset price returns must display high degree of volatility.

- **Long range dependence** If one looks at autocorrelation function of absolute returns as a function of time lag $f(\tau) = corr(|r_{t+\tau,\Delta t}|, |r_{t,\Delta t}|)$, it is empirically shown that it decays according to the power law distribution $f(\tau) \sim \tau^{-\beta}$ with exponent $\beta \in [0.2, 0.4]$ [8].

- **Gain/loss asymmetry** Gain/loss asymmetry is prevalent for equity price returns as stocks lose value faster than they grow [8]. However, this trend is not as pronounced for foreign exchange and rates products. Skewness is a metric that can be used to quantify the asymmetry of probability distribution about its mean.

- **Volume/volatility positive correlation** Volume and volatility are positively correlated. Linear regression relationship $\mu V_\tau \sim \alpha + \beta\sigma_{\tau,\Delta t}$ can be derived from the data [34]

- **Asymmetric causal information flow** Coarse-scaled volatility predicts fine-scaled volatility better than fine-scaled volatility predicts coarse scaled-volatility.

Stylized Facts About Volumes and Order Flow

2.3 Stylized facts about volumes and order flow

- **Order book volumes** Volumes at best bid V_b (and respectively volumes at best ask V_a) are distributed according to Gamma distribution for $\gamma \leq 1$ [35]:
$$P(V_b) \sim \exp^{-V_b} V_b^{-1+\gamma}.$$
- **Order sizes** Order sizes are power-law distributed [7]. For instance,[36] show examples when limit order sizes are distributed as $P(x) \sim x^{-(1+\mu)}$ with exponent $1 + \mu \approx 2$ and market order sizes are distributed as $P(x) \sim x^{-(1+\mu)}$ with exponent $1 + \mu \approx 2.3 - 2.7$. Orders tend to have round number of shares (i.e. multiples of 10, 100, etc. are more common than neighboring sizes); in general, power-law distribution fit is product-specific.
- **Number of orders in a fixed time window** Number of orders in a fixed time window can be approximated by gamma or lognormal distributions [36].
- **Order inter-arrival times** In the literature, LOB order inter-arrival times are suggested to be fit into exponential [28], lognormal, and Weibull distributions [36].
- **New order prices** Prices at which new limit orders are placed, are power-law distributed around bid-ask [36]. Specifically, $P(\Delta) \sim \Delta^{-(1+\mu)}$ with $1 + \mu \approx 1.6$ [35].
- **Cancellation time, time-to-first-fill and time-to-execution** Lifetimes of both cancelled and executed limit orders are power-law distributed, $P(T) \sim T^{-(1+\mu)}$ with $1 + \mu$ ranging between 1.3 and 1.6 for both canceled and executed limit orders [36]. Since an order can require multiple fills to be completed, one must distinguish between time-to-first-fill and time-to-completion statistics. Generalized gamma distribution with accelerated failure time can be used to model time-to-first-fill and time-to-completion distributions [37].
- **Time correlation of order flow** Individual agent's order placement decisions depend on other agents' actions [10].

Distribution of order interval times

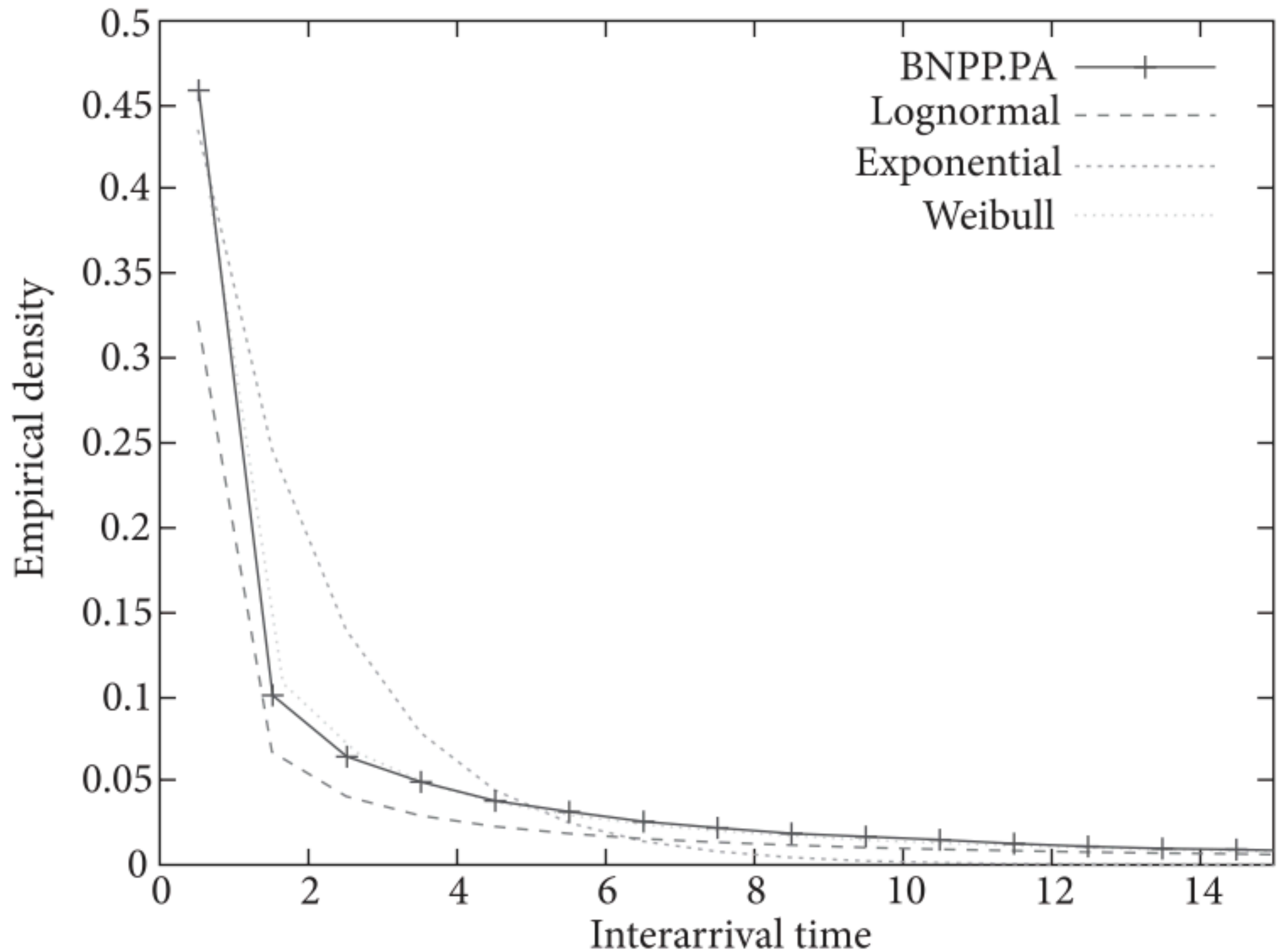


Fig. 2.2 Distribution of interarrival times for stock BNPP.PA (Main body, linear scale). Extracted from Chakraborti et al. (2011a)

Distribution of the number of trades in a given time period

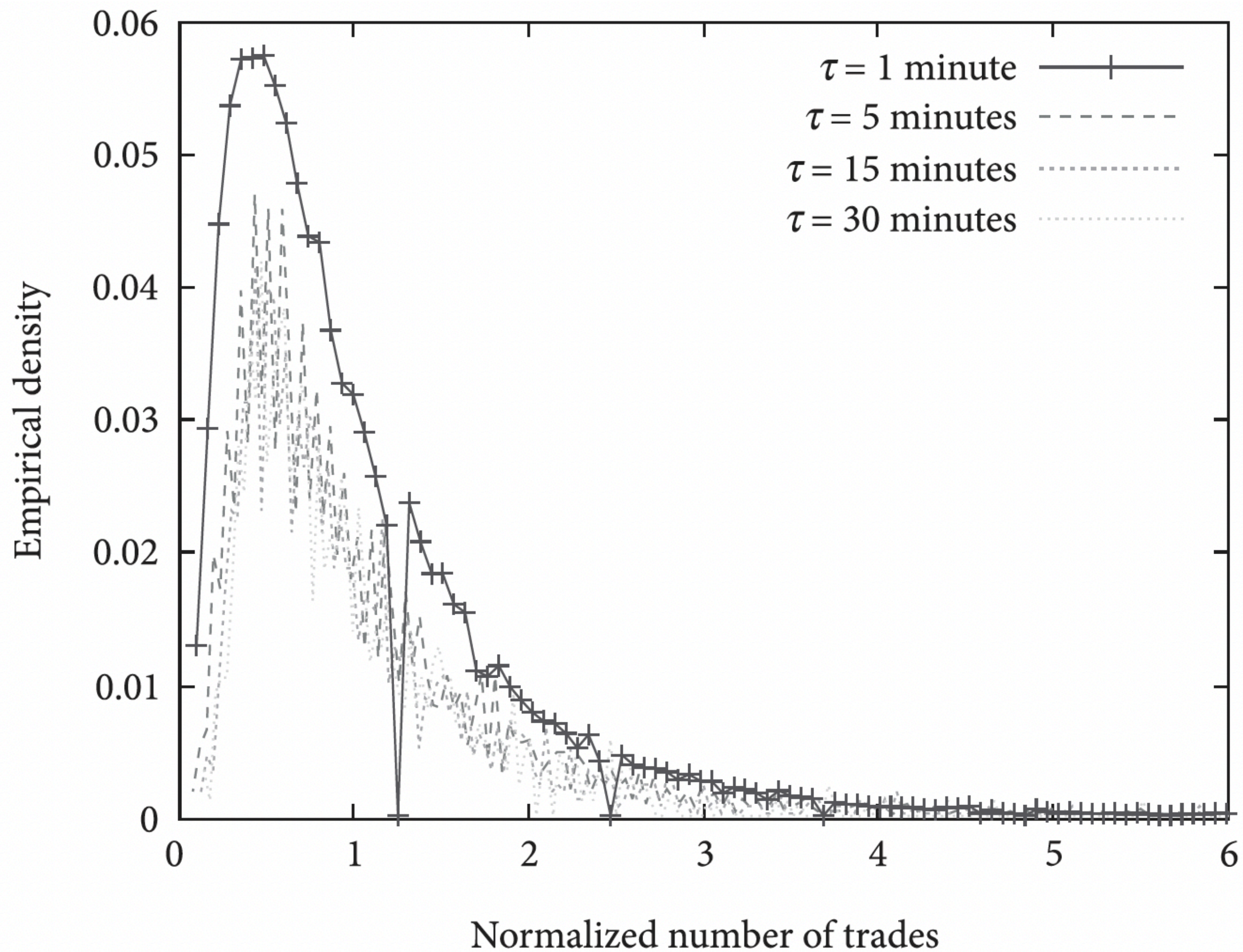


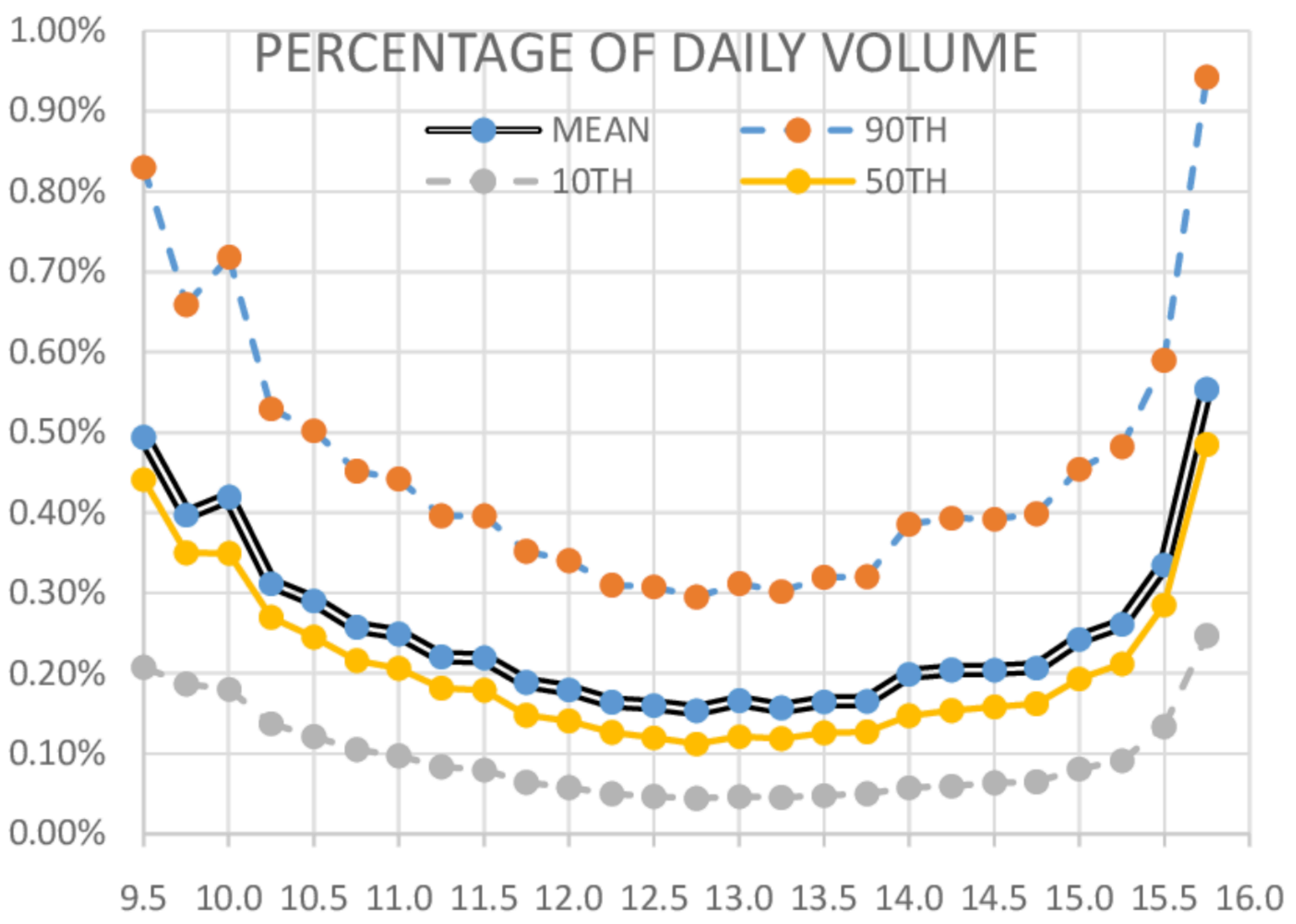
Fig. 2.3 Distribution of the number of trades in a given time period τ for stock BNPP.PA. This empirical distribution is computed using data from 2007, October 1st until 2008, May 31st. Extracted from Chakraborti et al. (2011a)

Stylized Facts About the Non-Stationary Patterns

- **Intraday volume patterns** LOB volumes are known to exhibit strong intraday patterns. For instance, historical foreign exchange trading volumes can be approximated by fifth-degree polynomial "U-shaped" regional sessions that correspond to New York, London, and Tokyo trading [38]. Similarly, in most equity markets, volumes are highest in the beginning of trading day, followed by a period of lower activity, and then spike again at the end of the trading day [7]. Note that making a transformation from physical time to tick (or transaction) time may help adjusting for intraday non-stationarity [39].
- **Seasonal volume patterns** Some assets, especially those consumer demand for which is seasonal (e.g., electricity futures), display strong seasonal volume patterns.
- **Intraday sensitivity to macro economic events/holidays** Due to product sensitivity to macro factors, volume spikes are known to occur in foreign exchange and rates markets during economic announcements. Equities trading is also sensitive to economic events [40]. Additionally, lower trading volumes are observed on holidays throughout all asset classes.
- **Intraday volume/spread negative correlation** Lower spreads are typically observed during periods of higher trading volumes.

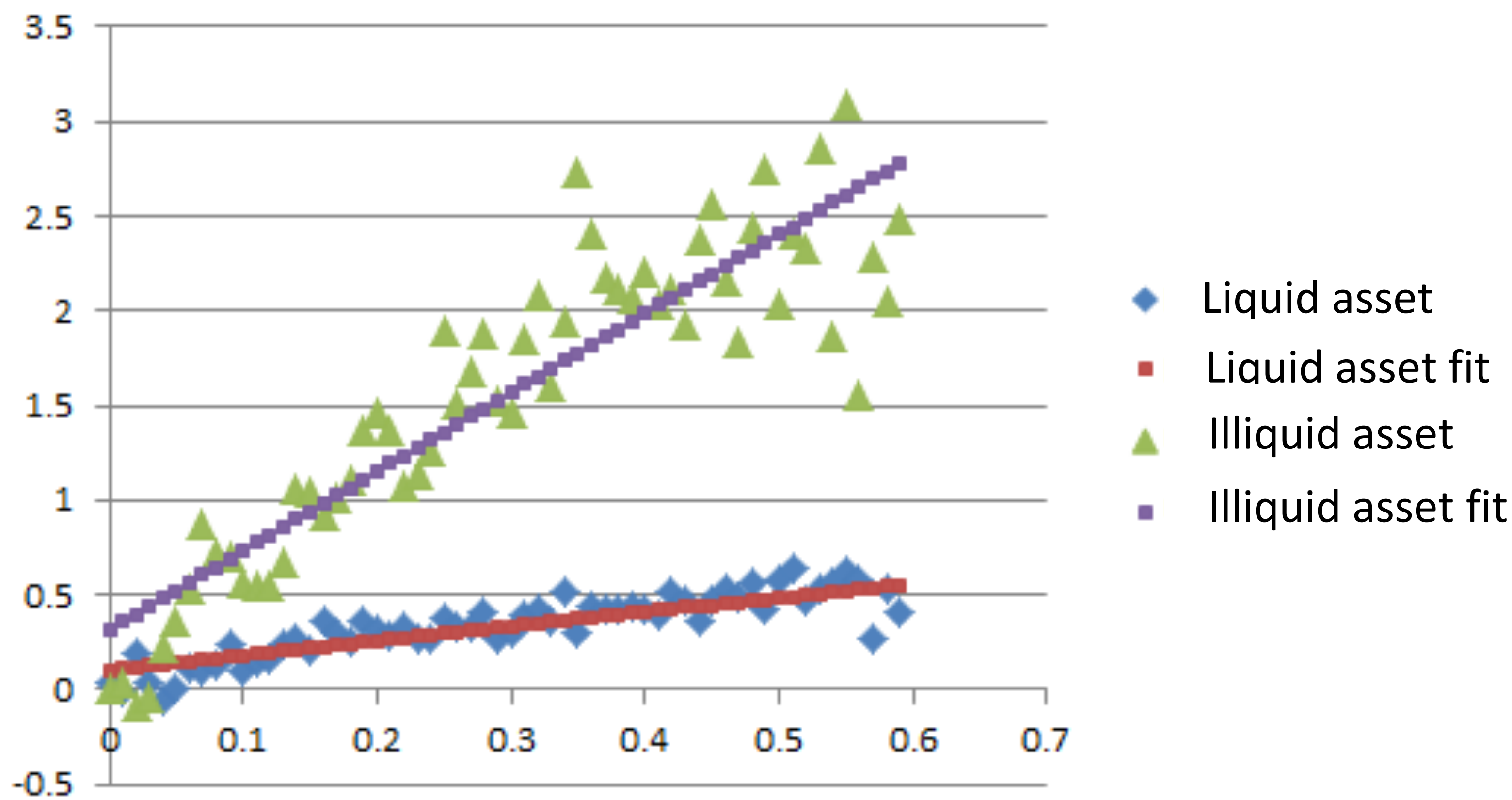
Electricity: Seasonality of price volatility

Intraday Trading Volumes



Time of the day

Stylized Facts About Market Impact



Observable Market Impact

Participation of Volume

- Similar shape across assets
- Difference in liquidity patterns

Stylized Facts About Cross-Asset Correlations

- When simulating multiple assets, cross asset correlation properties must hold.

Get Real: Realism Metrics for Robust Limit Order Book Market Simulations

Svitlana Vyetrenko[1], David Byrd[1,3], Nick Petosa[3], Mahmoud Mahfouz[1,2], Danial Dervovic[1], Manuela Veloso[1], and Tucker Hybinette Balch[1]

[1]J. P. Morgan AI Research, New York, NY, USA *
[2]Department of Electrical and Electronic Engineering, Imperial College, London, UK
[3]School of Interactive Computing, College of Computing, Georgia Institute of Technology, Atlanta, GA, USA

Homework 1