The Random Walk of High Frequency Trading

Eric M. Aldrich[1] Indra Heckenbach[2]
Gregory Laughlin[3]

[1]Department of Economics, UC Santa Cruz

[2]Department of Physics, UC Santa Cruz

[3]Department of Astronomy and Astrophysics, UC Santa Cruz

Motivation

In an ideal world, asset returns would follow a Gaussian distribution.

- Unfortunately, the do not.
- Almost all asset returns are better characterized by a fat-tailed distribution.
- This was first observed in Mandelbrot (1963).
- After decades of research, there is still no consensus regarding which family of fat-tailed distributions best characterizes asset returns.

Contribution

We carry out a ground-level re-examination of the process that generates short-period returns.

- We do this in the context of high-frequency, trade-by-trade data.

En route to our final model we highlight two insights:

1. Returns distributions are effectively categorized into two groups: those occurring during prescheduled news announcement periods, and those that do not.
2. Outside of news announcement periods, returns measured in trade-time (not clock-time) follow a Gaussian distribution.

Contribution

Characteristics of our model:

- Individual asset returns are Gaussian (outside the context of time).
- Intertrade durations follow a modified version of the Markov-Switching Multifractal Duration (MSMD) model of Chen, Diebold and Schorfheide (2013).
- Resulting clock-time returns are mixtures of Gaussians.
- Provides a very good fit to the data.

We apply our model to volatility data to investigate potential conditions of market failure.

Data

We focus our analysis on the CME E-mini S&P 500 Futures contract.

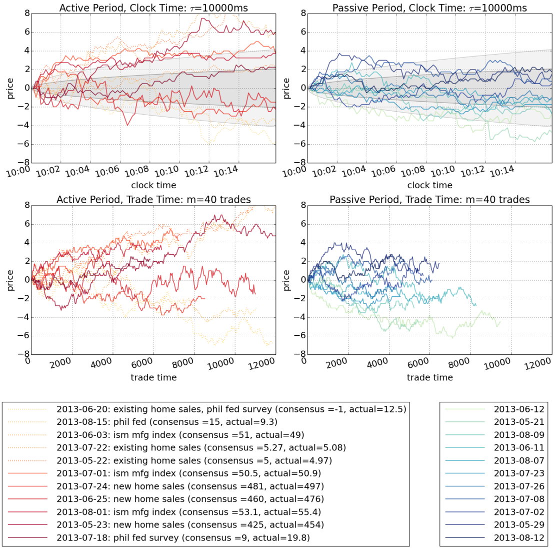
- This is a futures contract traded on the value of the S&P 500 Index.
- We obtained the full record of tick-by-tick trades for the period 18 May 2013 to 18 August 2013.
- It is exemplary of a large class of assets (including equities).
  - This is attributed to its liquidity and the relationship of price formation between the futures and equities exchanges (Laughlin, Aguirre and Grundfest (2013)).

Data

Our data sample contains 6,832,305 trade records (no quotes).

- We subsample trades that occurred in the 1000 seconds following 8:30 am and 10:00 am on each day of our sample (common news announcement times).
- We used the EconoDay calendar to determine days with news announcements scheduled for 8:30 am or 10:00 am.
- If a news announcement was scheduled for either time, the ensuing trades were classified as news affected (active) or not news affected (passive).
- The resulting sorted subsamples are roughly equal in size: 191,127 active trades and 174,041 passive trades.

Data



Active Period, Clock Time: τ=10000ms
Passive Period, Clock Time: τ=10000ms
Active Period, Trade Time: m=40 trades
Passive Period, Trade Time: m=40 trades
price
clock time
trade time
10:00
10:02
10:04
10:06
10:08
10:10
10:12
10:14
0
2000
4000
6000
8000
10000
12000
8
6
4
2
0
−2
−4
−6
−8
2013-06-20: existing home sales, phil fed survey (consensus =-1, actual=12.5)
2013-08-15: phil fed (consensus =15, actual=9.3)
2013-06-03: ism mfg index (consensus =51, actual=49)
2013-07-22: existing home sales (consensus =5.27, actual=5.08)
2013-05-22: existing home sales (consensus =5, actual=4.97)
2013-07-01: ism mfg index (consensus =50.5, actual=50.9)
2013-07-24: new home sales (consensus =481, actual=497)
2013-06-25: new home sales (consensus =460, actual=476)
2013-08-01: ism mfg index (consensus =53.1, actual=55.4)
2013-05-23: new home sales (consensus =425, actual=454)
2013-07-18: phil fed survey (consensus =9, actual=19.8)
2013-06-12
2013-05-21
2013-08-09
2013-06-11
2013-08-07
2013-07-23
2013-07-26
2013-07-08
2013-07-02
2013-05-29
2013-08-12

Returns

We define returns in two ways.

- Clock-time returns:

$$r_\tau(t) = p(t) - p(t - \tau)$$

  where τ is the clock-time duration.

- Trade-time returns:

$$r_m(n) = p(n) - p(n - m),$$

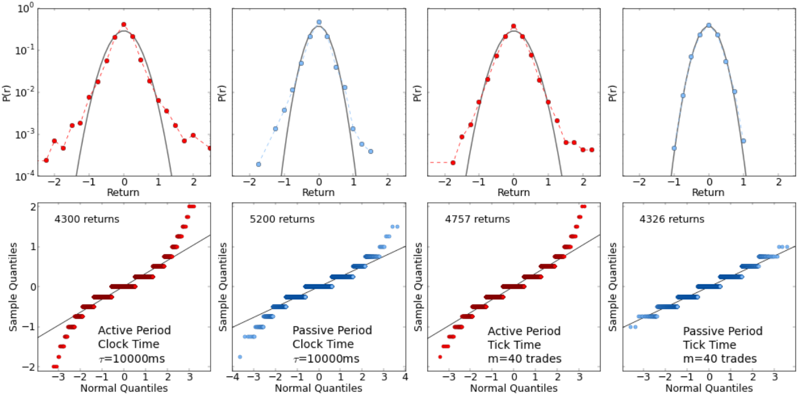
  where n denotes the n-th trade and m is the number of trades in a unit of time.

Trade Time

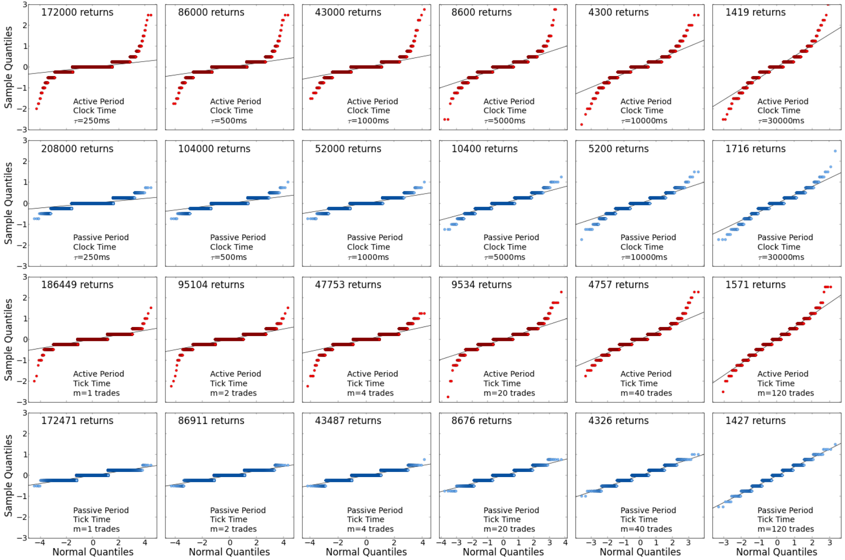
Trade time fixes a certain number of trades as a unit of time.

- The clock time between trade-time intervals may be variable.
- In our data there are 54 passive 1000-second intervals with 174,041 observations.
  - This is an average of 3.22 trades per second.
- There are 43 active 1000-second intervals with 191,127 observations.
  - This is an average of 4.44 trades per second.
- So we roughly equate a trade-time interval of $m = 1$ trade to a clock-time interval of $\tau = 0.25$ seconds.

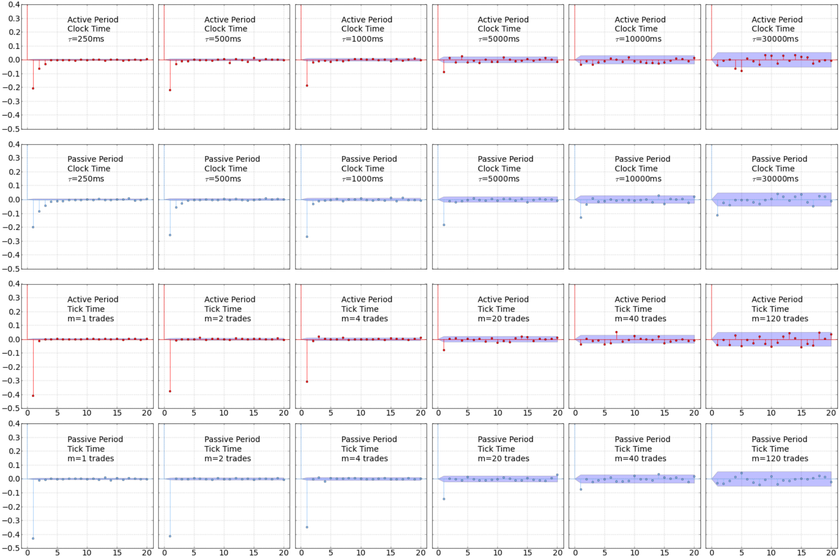
Empirical Densities, $\tau = 10,000$ ms and $m = 40$



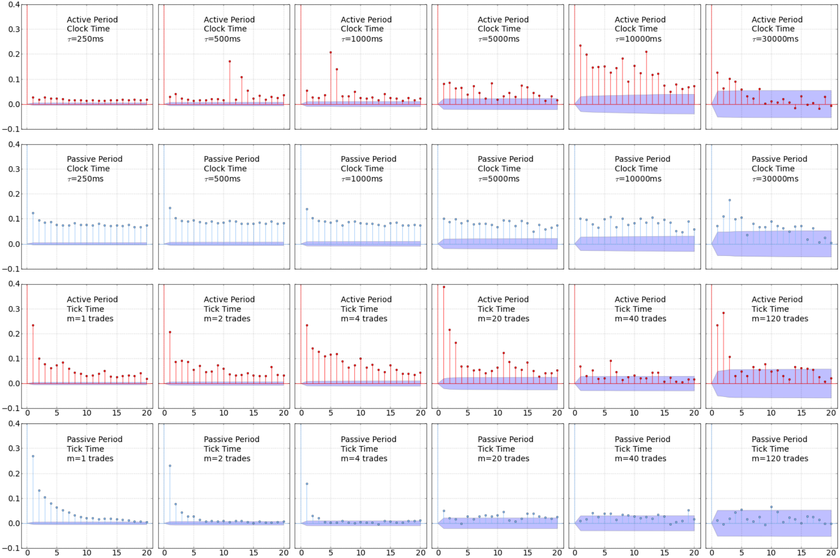
Q-Q Plots of Clock-Time and Trade-Time Returns



ACFs of Clock-Time and Trade-Time Returns



ACFs of Clock-Time and Trade-Time Squared Returns



General Model

We develop a hierarchical model of clock-time returns that mixes a distribution of trade-time returns with a distribution of trade arrival.

- Given m, assume trade-time returns are i.i.d. Gaussian:

$$r_m(n) \overset{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma), \forall n.$$

Trade arrivals will be distributed according to some counting process.

- For clock-time duration τ, denote the number of m-period executed trades as $N_m(\tau)$ with probability $P(N_m(\tau) = k)$.

General Model

We are interested in the random variable

$$r_\tau(t) = \sum_{i=1}^{N_m(\tau)} r_m(n).$$

The probability density function of $r_\tau(t)$ is,

$$p(r_\tau(t)|\mu, \sigma) = \sum_{k=1}^{\infty} p\left(\sum_{i=1}^{k} r_m(n) \middle| N_m(\tau) = k, \mu, \sigma\right) P(N_m(\tau) = k).$$

- This is a finite Normal mixture model.
- The mixture weights vary according to the probability distribution of $N_m(\tau)$.

General Model

The mixture model can also be viewed as a hierarchical model.

- In the first stage the number of trades is drawn from the distribution of $N_m(\tau)$.
- In the second stage a single τ-period return is drawn from the normal distribution:

$$r_\tau(t) = \sum_{i=1}^{N_m(\tau)} r_m(n) \sim \mathcal{N}\left(N_m(\tau)\mu, \sqrt{N_m(\tau)}\sigma\right)$$

- The resulting distribution of $r_\tau(t)$ will have fatter tails than a Gaussian.
- The tail fatness will be intimately related to the distribution of $N_m(\tau)$.

Poisson Trade Arrival

A starting point for modeling trade arrivals would be to assume they follow a Poisson process:

$$N_m(\tau) \sim \mathrm{Poisson}(\lambda\tau)$$

or

$$P\left(N_m(\tau) = k\right) = \frac{(\lambda\tau)^k}{k!} \exp -\lambda\tau,$$

where λ is the arrival intensity parameter.

Compound Poisson Process

The probability density function $r_\tau(t)$ is

$$p(r_\tau(t)|\mu,\sigma) = \sum_{k=1}^{\infty} \frac{1}{\sigma\sqrt{2\pi k}} \exp\left\{-\frac{1}{2}\frac{\left(\sum_{i=1}^{k} r_m(n) - k\mu\right)^2}{k\sigma^2}\right\}$$
$$\times \exp\{-\lambda\tau\}\frac{(\lambda\tau)^k}{k!}.$$

- This is known as a compound Poisson process.
- The density function cannot be obtained in closed form, but we can approximate it via Monte Carlo simulation.
- Poisson trade arrivals are associated with Exponential inter-trade arrival times.

MSMD Model

As an alternative, we use the Markov-Switching Multifractal Duration (MSMD) Model for inter-trade durations.

- This model is due to Chen, Diebold and Schorfheide (2013).

The core components of the MSMD model are:

- $\bar{k}$ latent state variables, $M_{k,i}$, that obey two-state Markov-switching processes.
- Persistence parameters, γ_k, for each latent variable $M_{k,i}$, $k = 1, 2, \ldots, \bar{k}$.

MSMD Model

The distribution of MSMD trade durations, d_i, is governed by the equations:

$$d_i = \frac{\varepsilon_i}{\lambda_i}$$

$$\varepsilon_i \sim Exp(1)$$

$$\lambda_i = \lambda \prod_{k=1}^{\bar{k}} M_{k,i}$$

$$M_{k,i} = \begin{cases} M & \text{with probability } \gamma_k \\ M_{k,i-1} & \text{otherwise} \end{cases}$$

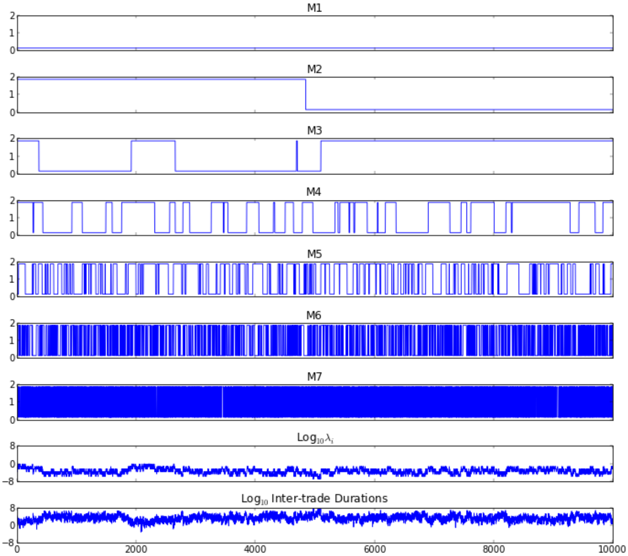
$$\gamma_k = 1 - (1 - \gamma_{\bar{k}})^{b^{k-\bar{k}}}$$

$$M = \begin{cases} m_0 & \text{with probability } 1/2 \\ 2 - m_0 & \text{otherwise.} \end{cases}$$

MSMD Model

- The MSMD model is characterized by five parameters: $\bar{k} \in \mathbb{N}$, $\lambda > 0$, $\gamma_{\bar{k}} \in (0,1)$, $b \in (1,\infty)$ and $m_0 \in (0,2]$.
- Conditional on knowing the values of the latent state variables, inter-trade durations are Exponentially distributed with rate parameter λ_i.
- As time evolves the latent states, $M_{k,i}$, switch values with varying degrees of persistence, γ_k.
- This causes the actual distribution of intra-day trade durations to be a mixture of Exponentials.
- The latent states can be interpreted as shocks that have varying impacts over diverse timescales.

MSMD Components



Compound Multifractal Process

Note that the MSMD model is a duration model (d_i) and not a trade arrival model ($N_m(\tau)$).

- We can infer the counting process, $N_m(\tau)$, associated with the MSMD durations.
- We can pair simulations from the counting density with Gaussian random variables to obtain simulations for clock-time returns $r_\tau(t)$ associated with the MSMD model.
- We refer to the resulting process as a compound multifractal process.

Compound Truncated Multifractal Process

We also model intertrade durations with a modification of the MSMD model:

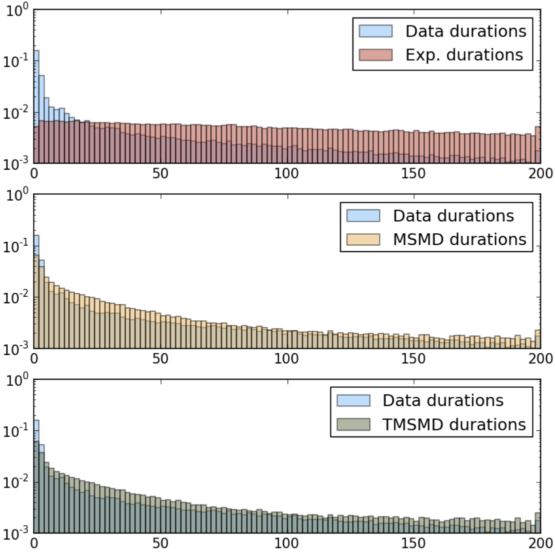
$$d_i = \min\{d_{MSMD}, d_{\overline{Exp}}\} \tag{1a}$$

$$d_{MSMD} \sim MSMD(\bar{k}, \lambda, \gamma_{\bar{k}}, b, m_0) \tag{1b}$$

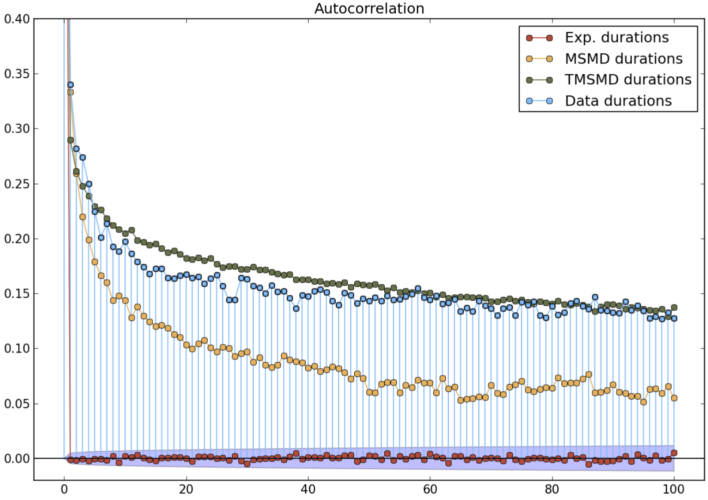
$$d_{\overline{Exp}} \sim Exp(\nu_{max}). \tag{1c}$$

- ν_{max} is chosen so that $\mathrm{E}\left[\max\{d_{\overline{Exp}}\}\right]$ equals the maximum duration observed in the data.
- Conceptually: two types of traders (electronic and human) arriving according to different processes.
- We pair Gaussian returns with these durations to arrive at a compound truncated multifractal process.

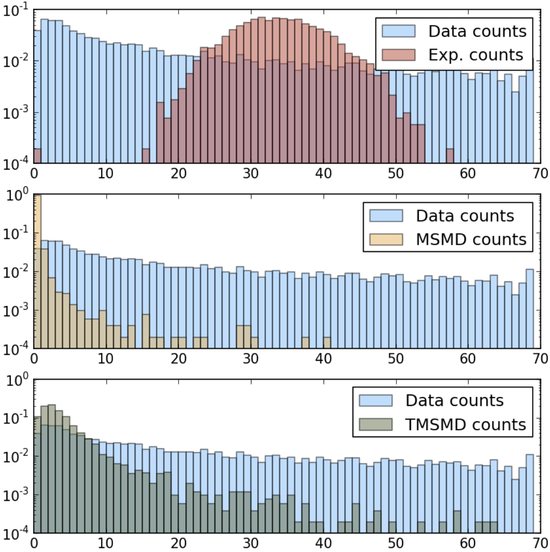
Duration Process Densities



Autocorrelations of Durations



Counting Process Densities



Estimation

We use the passive-market E-mini data to estimate the Exponential, MSMD and TMSMD duration models.

- Estimation is done via maximum likelihood.
- Since the MSMD density cannot be obtained in closed form, we resort to the nonlinear filtering method of Hamilton (1989) to obtain MLEs.
- We estimate the model using trade-time unit $m = 1$.
- We optimize over $\bar{k}$ and fix $\bar{k} = 7$.

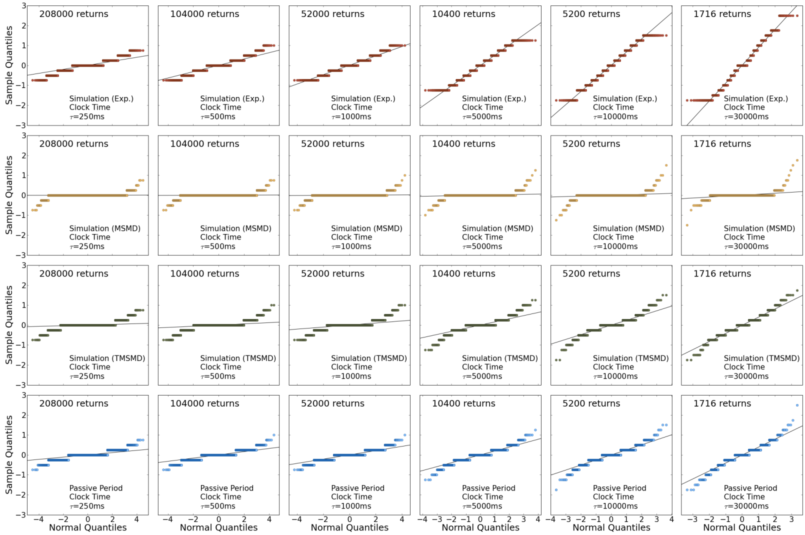
We assume that the trade-time returns, $r_m(n)$, follow a Gaussian density.

Simulations of Clock-Time Returns

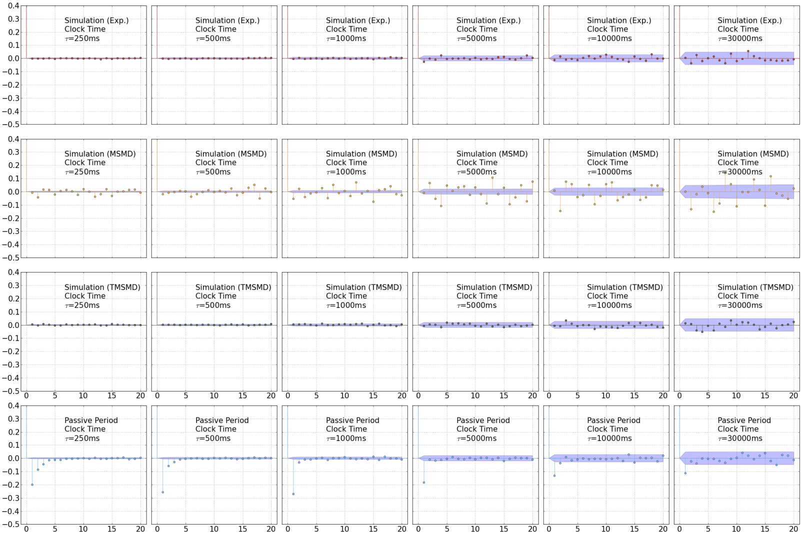
With estimates of the component distributions in hand, we can simulate clock-time returns.

- First, simulate inter-trade durations from each of the duration models.
- Second, pair the durations with independent draws of trade-time returns from the estimated Gaussian density.
- Third, aggregate individual returns within a fixed clock time interval.

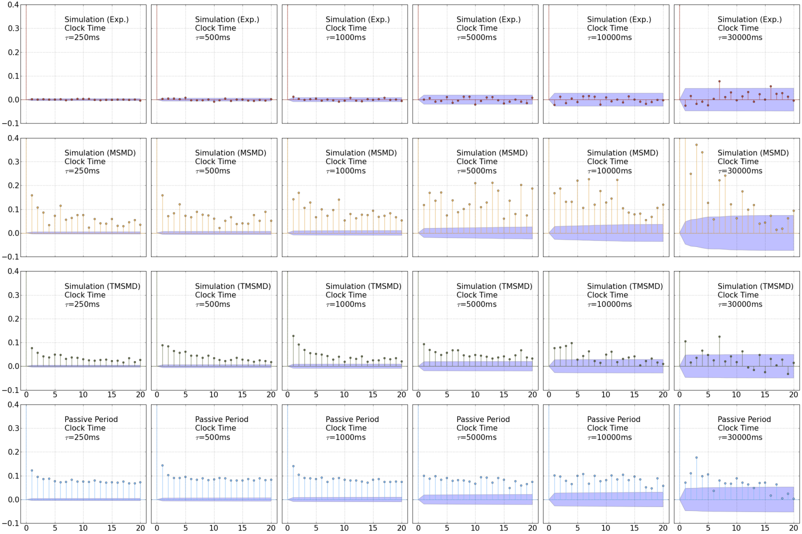
Q-Q Plots of Simulated Clock-Time Returns



ACFs of Simulated Clock-Time Returns



ACFs of Simulated Clock-Time Squared Returns



Goodness of Fit Tests

	τ					
	250	500	1000	5000	10000	30000
χ^2 Exp	35035.0	63146.0	106900.0	77396.0	48035.0	8680.0
χ^2 MSMD	24747.0	20079.0	15229.0	7519.8	5252.7	3039.9
χ^2 TMSMD	15462.0	10963.0	6577.7	762.14	85.534	50.68
5% χ^2_k critical values	12.592	14.067	14.067	18.307	22.362	24.996
KL Exp	0.032975	0.089006	0.20827	0.50814	0.55751	0.4007
KL MSMD	0.43104	0.62514	0.81613	1.2171	1.2877	1.365
KL TMSMD	0.089029	0.11033	0.11152	0.041454	0.0076904	0.014179

Table 4: Goodness of fit statistics for simulated returns under the Exponential and MSMD duration models. The first three rows report χ^2 goodness of fit measures relative to the observed data for each clock time interval τ that we consider. The latter three rows report Kullback-Leibler divergences, again, relative to the observed data.

Goodness of Fit Tests

	τ					
	250	500	1000	5000	10000	30000
LB r_{Exp}	25.2	28.707	25.64	19.6	18.261	18.848
LB r_{MSMD}	860.02	483.01	311.56	95.955	95.207	76.485
LB r_{TMSMD}	45.206	53.868	47.937	23.578	20.614	28.109
LB r_{Data}	1930.1	351.88	110.36	18.059	28.569	20.303
LB r^2_{Exp}	32.536	11.004	22.702	16.973	19.24	16.647
LB r^2_{MSMD}	2397.5	2435.3	1855.0	409.8	452.78	135.69
LB r^2_{TMSMD}	6559.7	4135.7	1780.7	624.93	345.7	68.076
LB r^2_{Data}	27412.0	16306.0	6767.5	1249.4	625.47	190.64

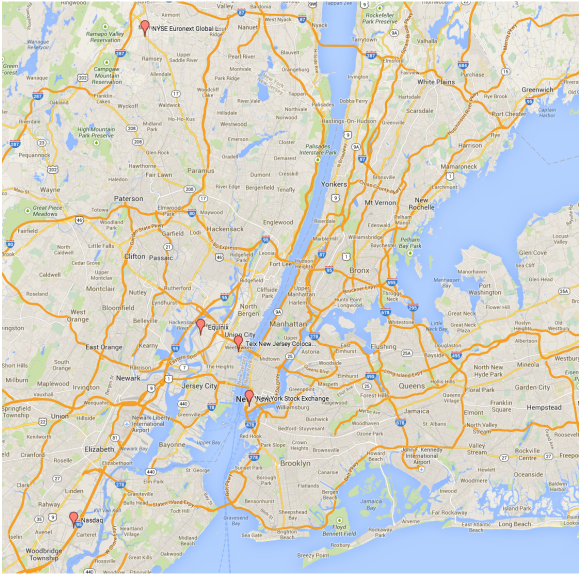
Table 5: Ljung-Box statistics for simulated returns under the Exponential, MSMD and TMSMD duration models as well as the observed data for each clock time interval τ that we consider. The first four rows report Ljung-Box statistics for the ACFs of returns and the last four rows report Ljung-Box statistics for the ACFs of squared returns.

Market Fragmentation

Our results will inform the discussion of market fragmentation.

- The equities markets consist of 13 exchanges at 4 locations in metro New York.
    1. Weehawken, NJ (BATS)
    2. Secaucus, NJ (DirectEdge)
    3. Mahwah, NJ (NYSE)
    4. Carteret, NJ (Nasdaq)
- Futures trade at the CME Globex exchange in Aurora, IL.
- Options trade at the CBOE exchange in Secaucus, NJ.
- Currencies trade primarily in Secaucus, NJ.

Exchanges in Metro NY



Exchanges in IL and NY

Exchange Latencies and Market Influence

- The round-trip line time between Chicago and NY is roughly 10 ms.
- The longest round-trip line time between metro NY exchanges is no longer that 666 microseconds (2/3 ms).
- A rough estimate of market influence on price formation: 70% futures (IL), 30% spot (NJ).
- We approximate aggregate round trip market latency as

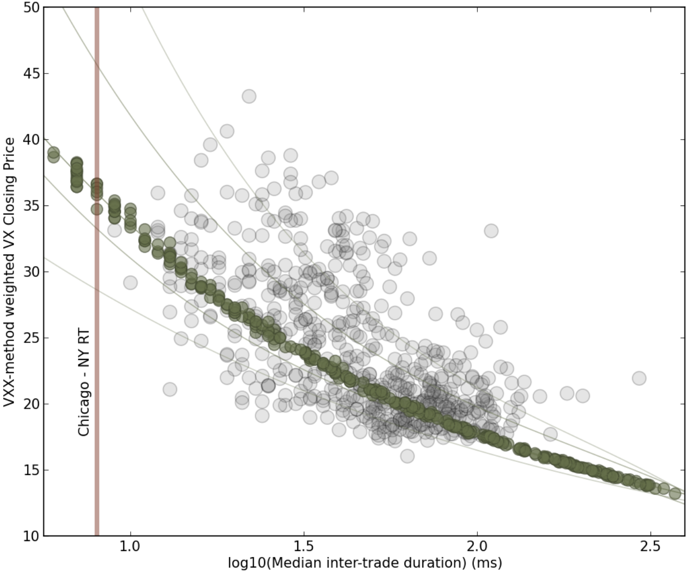
$$0.7 \times 10 \text{ ms} + 0.3 \times 0.66 \text{ ms} \approx 7 \text{ ms}.$$

Latency/Volatility Relationship

There is a relationship between volatility and trade latency in market data.

- Aug 9, 2011 was a heavy trading day for E-mini: roughly 5M shares traded with median inter-trade duration of 18 ms.
- Closing VIX on Aug 9, 2011 was 48.00.
- Jul 11, 2011 was a more typical day: roughly 1M shares traded with median inter-trade duration of 80 ms.
- Closing VIX on Jul 11, 2011 was 18.39.

Latency/Volatility Relationship



Latency/Volatility Relationship

These values suggest that in a stressed market (where latencies approach 7 ms) the maximum attainable VIX would be between 60 and 80.

- The highest observed closing VIX value observed to date is 81.65 on Oct 27, 2008.
- Our model suggests that if median intertrade durations fell to 350 microseconds (round-trip exchange latency in NJ) VIX could reach 200.
- Co-locating all NY exchanges would reduce the total market latency to roughly 1 microsecond, or a VIX upper bound of 6500.
- Market fragmentation forces an implicit threshold which bounds market volatility.

Conclusion

We develop a model that approximates the distribution of high-frequency asset returns quite well.

- A striking result of our analysis is that when measured in trade-time, outside of scheduled news announcements, trade-time returns follow a Gaussian distribution.
- When pairing Gaussian trade-time returns with an accurate model of inter-trade duration, we obtain a good approximation of the returns distribution in clock time.
- The clock-time returns distribution has fat tails that are commensurate with the data and also exhibits volatility clustering.
- Finally, our work suggests that market fragmentation is desirable from the perspective of placing an implicit bound on market volatility.