High Frequency Data - Part I

12th European Summer School in Financial Mathematics

Roberto Renò

September 4-6, 2019

- Frictions
- EXcess Idle Time (EXIT) and zeros
- An economic model of price formation featuring zeros
- Rounding and excess staleness
- Asset pricing applications

Call $X_t = \log(P_t)$.

- The (ubiquitous) Ito semimartingale:

$$X_t = X_0 + \int_0^t \mu_s \, ds + \int_0^t \sigma_s \, dW_s + \sum_{j=1}^{N_t} \gamma_j.$$

- Market microstructure on a partition of n points $t_{0,n} = 0 < t_{1,n} < \ldots < t_{n,n} = 1$.

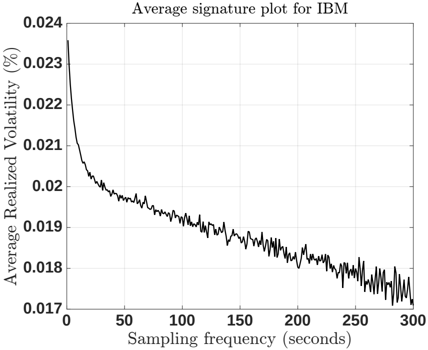
$$\widetilde{X}_{t_{j,n}} = X_{t_{j,n}} + \varepsilon_{j,n}.$$

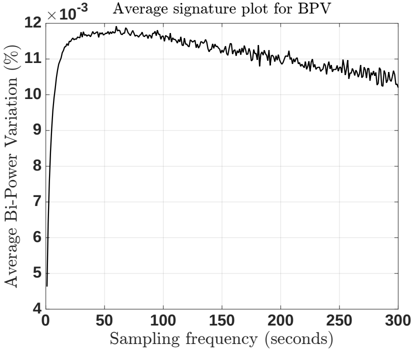
- If we really observe $\widetilde{X}_{t_{j,n}}$ then

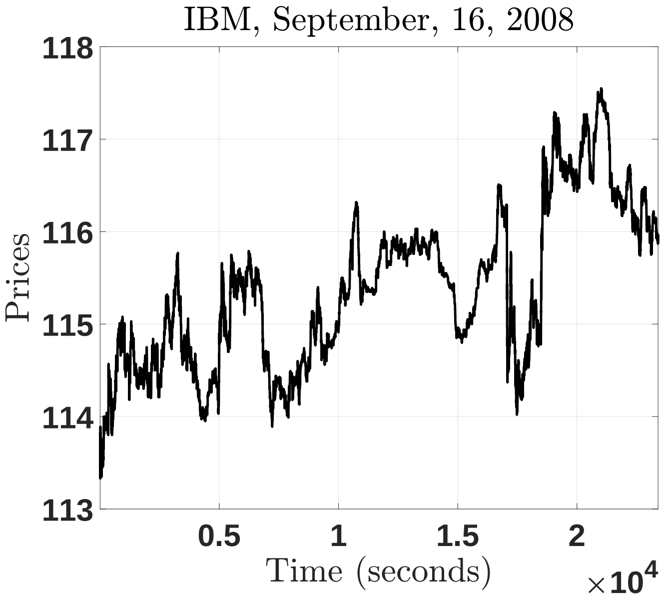
$$\sum_{j=1}^{n} \left(\widetilde{X}_{t_{j,n}} - \widetilde{X}_{t_{j-1,n}} \right)^2 \xrightarrow{\text{p}} +\infty, \quad \sum_{j=1}^{n-1} \left| \widetilde{X}_{t_{j,n}} - \widetilde{X}_{t_{j-1,n}} \right| \left| \widetilde{X}_{t_{j+1,n}} - \widetilde{X}_{t_{j,n}} \right| \xrightarrow{\text{p}} +\infty.$$

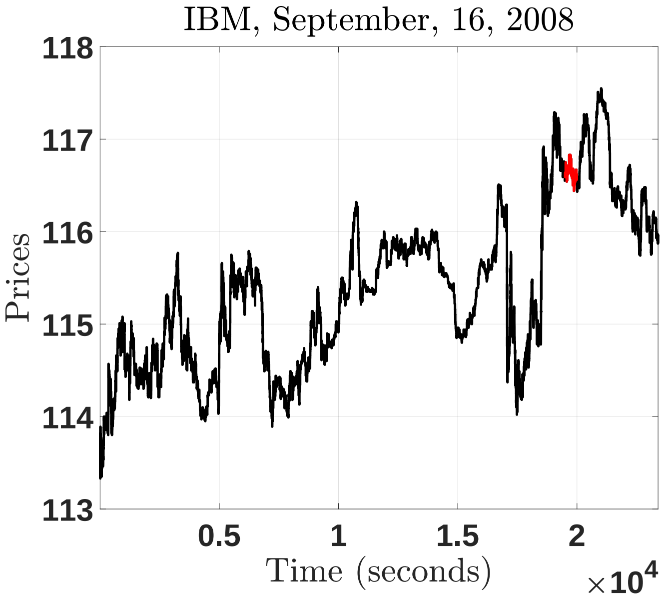
What do data tell us?

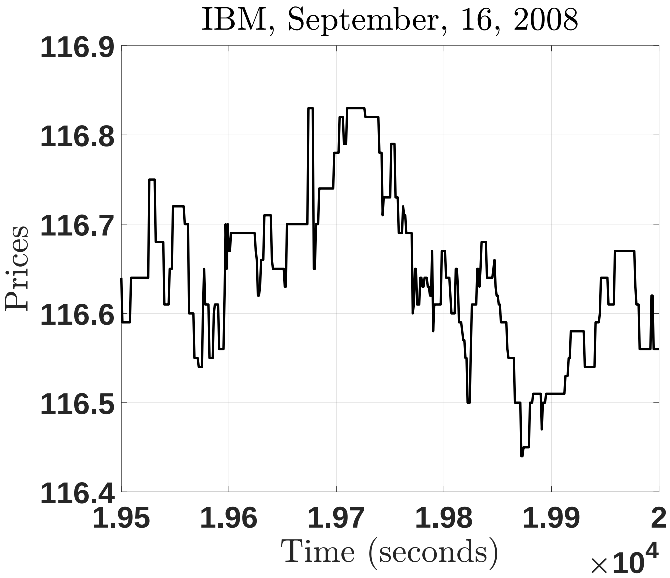
Empirical evidences

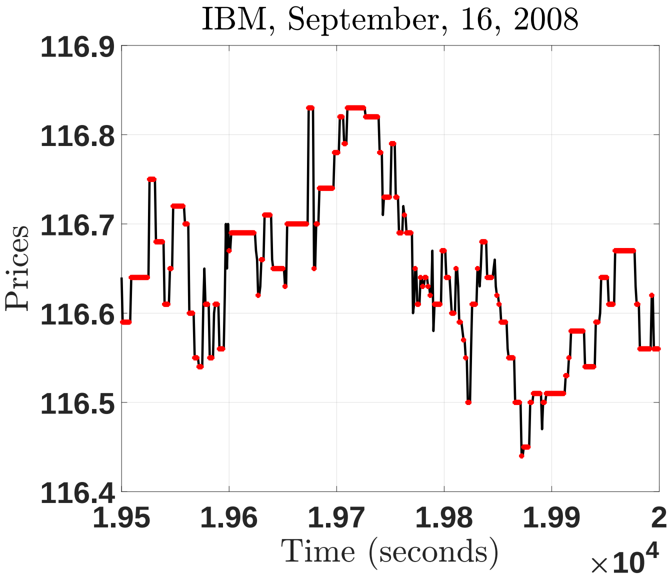




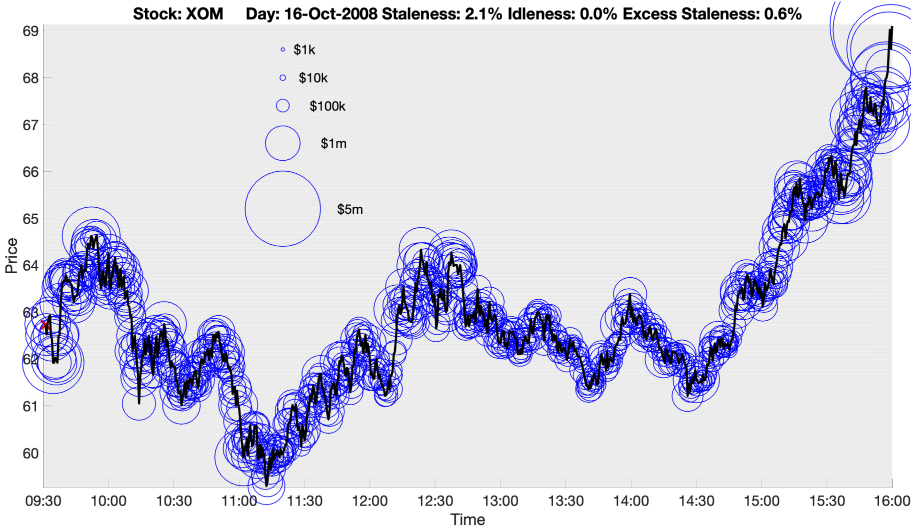

IBM, September, 16, 2008
Prices
118
117
116
115
114
113
0.5
1
1.5
2
Time (seconds)
×10^4


IBM, September, 16, 2008
Prices
118
117
116
115
114
113
0.5
1
1.5
2
Time (seconds)
×10^4

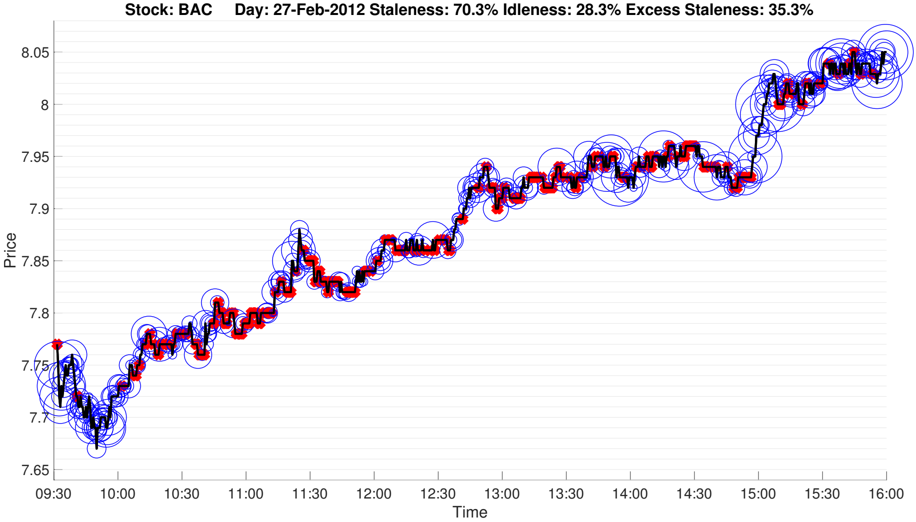

IBM, September, 16, 2008
116.9
116.8
116.7
116.6
116.5
116.4
Prices
1.95
1.96
1.97
1.98
1.99
2
Time (seconds)
×10^4

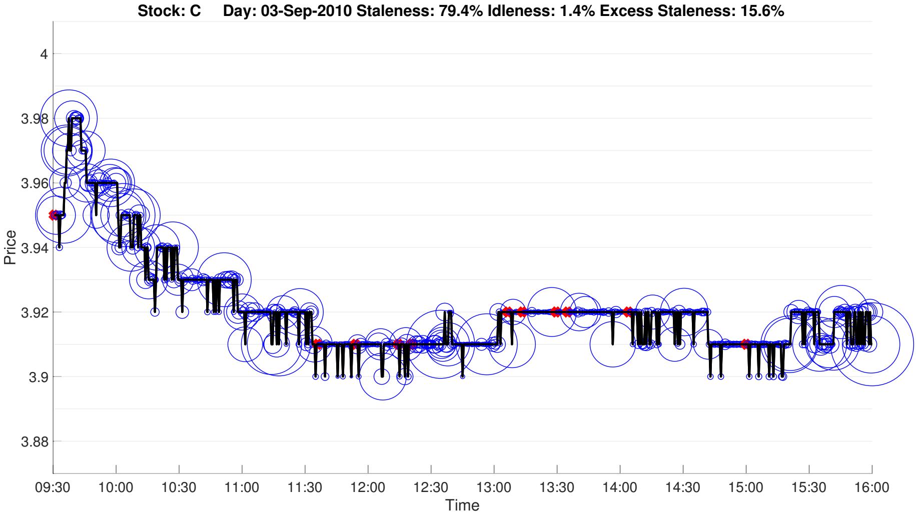

IBM, September, 16, 2008
Prices
116.9
116.8
116.7
116.6
116.5
116.4
1.95
1.96
1.97
1.98
1.99
2
Time (seconds)
×10^4

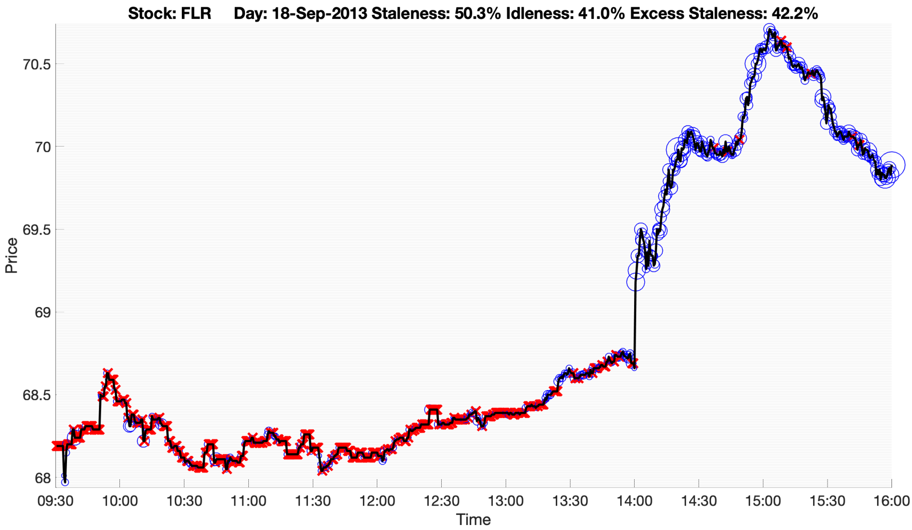
- Our data consists of all trades of 249 NYSE-listed stocks, recorded from 9:30 a.m. to 4 p.m., for the years 2006-2014.
- We employ the 249 stocks with the largest average traded volume during the period.
- Given our focus on the NYSE-listed stocks with the largest traded volume, i.e., those whose prices are expected to be the least inactive, our findings can be viewed as being conservative.



Stock: XOM Day: 16-Oct-2008 Staleness: 2.1% Idleness: 0.0% Excess Staleness: 0.6%
$1k
$10k
$100k
$1m
$5m
Price
69
68
67
66
65
64
63
62
61
60
09:30
10:00
10:30
11:00
11:30
12:00
12:30
13:00
13:30
14:00
14:30
15:00
15:30
16:00
Time

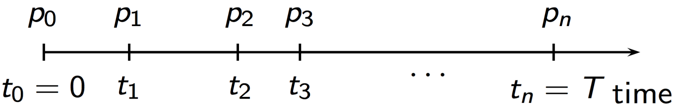

Stock: BAC Day: 27-Feb-2012 Staleness: 70.3% Idleness: 28.3% Excess Staleness: 35.3%
Price
8.05
8
7.95
7.9
7.85
7.8
7.75
7.7
7.65
09:30
10:00
10:30
11:00
11:30
12:00
12:30
13:00
13:30
14:00
14:30
15:00
15:30
16:00
Time


Stock: C Day: 03-Sep-2010 Staleness: 79.4% Idleness: 1.4% Excess Staleness: 15.6%
Price
4
3.98
3.96
3.94
3.92
3.9
3.88
09:30
10:00
10:30
11:00
11:30
12:00
12:30
13:00
13:30
14:00
14:30
15:00
15:30
16:00
Time


Stock: FLR Day: 18-Sep-2013 Staleness: 50.3% Idleness: 41.0% Excess Staleness: 42.2%
Price
70.5
70
69.5
69
68.5
68
09:30
10:00
10:30
11:00
11:30
12:00
12:30
13:00
13:30
14:00
14:30
15:00
15:30
16:00
Time

Consider a stochastic process $X_t = \ln(p_t)$ (the log-price) observed over $[0, T]$.

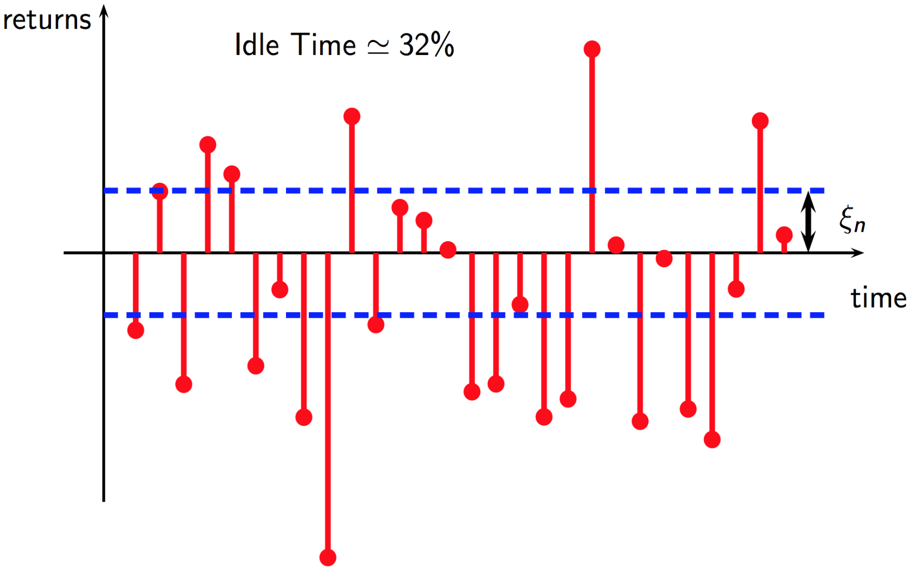
Consider a partition:



(you can assume, for simplicity, times being equally spaced)

Idle Time

$$\mathrm{IT} = \sum_{i=1}^{n} (t_i - t_{i-1}) 1_{\{|X_{t_i} - X_{t_{i-1}}| \leqslant \xi_n\}}.$$


returns
Idle Time ≃ 32%
ξ_n
time

- We think of ξ_n as a vanishing sequence ($\xi_n \to 0$).
- $IT \in [0, 1]$
- Under a frictionless Itō semimartingale null, there is still a (vanishing, as $n \to \infty$) probability of sluggish behavior!

EXcess Idle Time

The quantity

$$IT - \underbrace{bias(IT)}_{\text{Under the semimartingale null}}$$

is named EXcess Idle Time (EXIT) and is designed to be centered around zero.

Assume $\mu \equiv 0$ and $\sigma_s \equiv \sigma_0$.

$$X_t = X_0 + \sigma_0 W_t.$$

Consider an equispaced partition $\Delta_n = 1/n$ and a sequence $\xi_n \to 0$. Expand in ξ_n

$$\begin{aligned} F(\xi_n) &\doteq \mathbb{P}\left[|X_{\Delta_n} - X_0| \leqslant \xi_n\right] = \mathbb{P}\left[\sigma_0 \, |W_{\Delta_n}| \leqslant \xi_n\right] = \mathbb{P}\left[\frac{|W_{\Delta_n}|}{\sqrt{\Delta_n}} \leqslant \frac{\xi_n}{\sigma_0 \sqrt{\Delta_n}}\right] \\ &= \mathbb{P}\left[|u| \leqslant \frac{\xi_n}{\sigma_0 \sqrt{\Delta_n}}\right] = 2 \int_0^{\frac{\xi_n}{\sigma_0 \sqrt{\Delta_n}}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, dx. \end{aligned}$$

$$G(\xi) = \int_0^{\frac{\xi}{\sigma_0 \sqrt{\Delta_n}}} e^{-\frac{x^2}{2}} \, dx \Rightarrow \begin{cases} G(0) &= 0 \\ G'(\xi) &= \frac{1}{\sigma_0 \sqrt{\Delta_n}} e^{-\frac{\xi^2}{2\sigma_0^2 \Delta_n}} \Rightarrow G'(0) = \sigma_0^{-1} \Delta_n^{-1/2} \\ G''(\xi) &= -\frac{\xi}{\sigma_0^3 \Delta^{3/2}} e^{-\frac{\xi^2}{2\sigma_0^2 \Delta_n}} \xi \Rightarrow G''(0) = 0 \\ G'''(\xi) &= \frac{e^{-\frac{\xi^2}{2\sigma_0^2 \Delta_n}} \left(\xi^2 - \sigma_0^2 \Delta_n\right)}{\sigma_0^5 \Delta_n^{5/2}} \Rightarrow G'''(0) = -\left(\sigma_0^{-1} \Delta_n^{-1/2}\right)^3 \end{cases}$$

$$G(\xi) = \frac{\xi}{\sigma_0 \Delta_n^{1/2}} + O\left(\left(\frac{\xi}{\sigma_0 \Delta_n^{1/2}}\right)^3\right) \Rightarrow F(\xi_n) = \sqrt{\frac{2}{\pi}} \frac{\Delta_n^{-1/2} \xi_n}{\sigma_0} + O\left(\left(\Delta_n^{-1/2} \xi_n\right)^3\right).$$

Let $Y_{j,n} \in \mathcal{F}_{t_{j,n}}$. $\mathbb{E}\left[Y_{t_{j,n}}^2\right] < \infty$. Assume that the following conditions are satisfied $\forall t \in [0, T]$:

1. $\sum_{t_j \leqslant t} \mathbb{E}\left[Y_{t_{j,n}} \mid \mathcal{F}_{t_{j-1,n}}\right] \xrightarrow{p} 0$,
2. $\sum_{t_j \leqslant t} \left(\mathbb{E}\left[Y_{t_{j,n}}^2 \,\middle|\, \mathcal{F}_{t_{j-1,n}}\right] - \left(\mathbb{E}\left[Y_{t_{j,n}} \mid \mathcal{F}_{t_{j-1,n}}\right]\right)^2\right) \xrightarrow{p} \int_0^t \left(v_s^2 + w_s^2\right) \, ds$,
3. $\sum_{t_j \leqslant t} \mathbb{E}\left[Y_{t_{j,n}} \left(W_{t_{j,n}} - W_{t_{j-1,n}}\right) \mid \mathcal{F}_{t_{j-1,n}}\right] \xrightarrow{p} \int_0^t v_s \, ds$,
4. $\sum_{t_j \leqslant t} \mathbb{E}\left[Y_{t_{j,n}}^2 \, \mathbb{1}_{\left\{\left|Y_{t_{j,n}}\right| > \varepsilon\right\}} \,\middle|\, \mathcal{F}_{t_{j-1,n}}\right] \xrightarrow{p} 0 \quad \forall \varepsilon > 0$,
5. $\sum_{t_j \leqslant t} \mathbb{E}\left[Y_{t_{j,n}} \left(N_{t_{j,n}} - N_{t_{j-1,n}}\right) \mid \mathcal{F}_{t_{j-1,n}}\right] \xrightarrow{p} 0$, for all bounded $\mathcal{F}_t$-martingales with $N_0 = 0$ and $[W, N]_s \equiv 0$,

where v_s and w_s are predictable processes, W a brownian motion. Then

$$\sum_{t_j \leqslant t} Y_{t_{j,n}} \xrightarrow{st} \int_0^t v_s \, dW_s + \int_0^t w_s \, dW'_s,$$

where W' is a Brownian motion independent of W.

$$X_{\Delta_n} - X_0 = \int_0^{\Delta_n} \mu_s\, ds + \int_0^{\Delta_n} \sigma_s\, dW_s.$$

By using a Taylor expansion of the characteristic function of X_{Δ_n} we get

$$\mathbb{E}_0\left[\mathbb{1}_{\{|X_{\Delta_n}-X_0|\leqslant \xi_n\}}\right] = \sqrt{\frac{2}{\pi}}\,\frac{\xi_n\,\Delta_n^{-1/2}}{\sigma_0} + O_p\left(\left(\xi_n\,\Delta_n^{-1/2}\right)^3\right),$$

whence (define $\Delta_j X = X_{t_{j,n}} - X_{t_{j-1,n}}$ and $t = T = 1$)

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^{n}\mathbb{E}_{j-1}\left[\mathbb{1}_{\{|\Delta_j X|\leqslant \xi_n\}}\right] &= \sum_{j=1}^{n}\Delta_n\left(\sqrt{\frac{2}{\pi}}\,\frac{\xi_n\,\Delta_n^{-1/2}}{\sigma_{j-1}} + O_p\left(\xi_n^3\,\Delta_n^{-3/2}\right)\right)\\ &= \left(\xi_n\sqrt{n}\right)\sqrt{\frac{2}{\pi}}\int_0^1\frac{ds}{\sigma_s} + O_p\left(\xi_n^3\,n^{3/2}\right)\xrightarrow{\text{p}} 0.\end{aligned}$$

So if $\xi_n\, n^{7/10} \to 0$

$$\frac{n^{1/4}}{\xi_n^{1/2}}\sum_{j=1}^{n}\left(\Delta_n\,\mathbb{1}_{\{|\Delta_j X|\leqslant \xi_n\}} - \sqrt{\frac{2}{\pi}}\,\frac{\xi_n\,\Delta_n^{1/2}}{\sigma_{j-1}}\right) = O_p\left(\left(\xi_n\,n^{7/10}\right)^{5/2}\right)\xrightarrow{\text{p}} 0.$$

Define $Y_{j,n} = \frac{n^{1/4}}{\xi_n^{1/2}}\left(\Delta_n\,\mathbb{1}_{\{|\Delta_j X|\leqslant \xi_n\}} - \sqrt{\frac{2}{\pi}}\,\frac{\xi_n\,\Delta_n^{1/2}}{\sigma_{j-1}}\right)$ and the first condition of Jacod theorem is met.

Second condition is

$$\sum_{j=1}^{n} \left(\mathbb{E}\left[Y_{t_{j,n}}^2 \,\middle|\, \mathcal{F}_{t_{j-1,n}} \right] - \mathbb{E}\left[Y_{t_{j,n}} \mid \mathcal{F}_{t_{j-1,n}} \right]^2 \right) \xrightarrow{\text{p}} \int_0^T \left(a_s^2 + b_s^2 \right) ds,$$

where the array $Y_{t_{j,n},n}$ is in our case

$$Y_{t_{j,n},n} = \frac{n^{1/4}}{\xi_n^{1/2}} \left(\Delta_n \mathbb{1}_{\{|\Delta_j X| \leqslant \xi_n\}} - \sqrt{\frac{2}{\pi}} \frac{\xi_n \Delta_n^{1/2}}{\sigma_{j-1}} \right).$$

Again, using the Taylor expansion

$$\psi_{j,n} = \mathbb{E}_{j-1}\left[\mathbb{1}_{\{|X_{j\Delta_n} - X_{(j-1)\Delta_n}| \leqslant \xi_n\}} \right] = \sqrt{\frac{2}{\pi}} \frac{\xi_n \Delta_n^{-1/2}}{\sigma_0} + O_p\left(\left(\xi_n \Delta^{-1/2} \right)^3 \right).$$

we get

$$\sum_{j=1}^{n} \left(\mathbb{E}\left[Y_{t_{j,n}}^2 \,\middle|\, \mathcal{F}_{t_{j-1,n}} \right] - \mathbb{E}\left[Y_{t_{j,n}} \mid \mathcal{F}_{t_{j-1,n}} \right]^2 \right) \xrightarrow{\text{p}} \sqrt{\frac{2}{\pi}} \int_0^1 \frac{1}{\sigma_s} ds.$$

The asymptotic variance is inversely proportional to the spot volatility.

The third condition of the Jacod's theorem is

$$\sum_{j=1}^{n} \mathbb{E}\left[Y_{t_{j,n}}\, \Delta_j W \;\middle|\; \mathcal{F}_{t_{j-1,n}}\right] \xrightarrow{\mathrm{p}} \int_0^T a_s\, ds.$$

Notice that $\mathbb{E}\left[Y_{t_{j,n}}\, \Delta_j W \;\middle|\; \mathcal{F}_{t_{j-1,n}}\right] = \frac{n^{1/4}}{\xi_n^{1/2}}\, \mathbb{E}\left[\Delta_n \mathbb{1}_{\{|X_{j\Delta_n} - X_{(j-1)\Delta_n}| \leqslant \xi_n\}}\, \Delta_j W \;\middle|\; \mathcal{F}_{t_{j-1,n}}\right]$ and use Hölder's inequality to get

$$\sum_{j=1}^{n} \mathbb{E}_{j-1}\left[Y_{t_{j,n}}\, \Delta_j W\right] \leqslant \left(\frac{\sqrt{n}}{\xi_n}\right)^{1/2} \sum_{j=1}^{n} \Delta_{n,j}\, (\psi_{j,n})^{1/p} \left(\mathbb{E}_{j-1}\left[|\Delta_j W|^{\frac{p}{p-1}}\right]\right)^{\frac{p-1}{p}}.$$

Since

$$\mathbb{E}_{j-1}\left[|\Delta_j W|^{\frac{p}{p-1}}\right] = \pi^{-1/2}\, 2^{p/(2\,(p-1))} \left(\Delta_{n,j}^{1/2}\right)^{p/(p-1)} \Gamma\left(\frac{2\,p-1}{2\,(p-1)}\right),$$

the red quantity becomes

$$2^{1/2}\, \pi^{\frac{1-p}{2\,p}}\, \Gamma\left(\frac{2\,p-1}{2\,(p-1)}\right)^{\frac{p-1}{p}} \left(\frac{\sqrt{n}}{\xi_n}\right)^{1/2} \sum_{j=1}^{n} \Delta_{n,j}\, (\psi_{j,n})^{1/p}\, \Delta_{n,j}^{\frac{1}{2}}$$

$$= \quad O_p(n^{1/4} \xi_n^{-1/2}\, n\, n^{-1} \xi_n^{1/p} n^{1/(2p)}\, n^{-1/2}) = O_p(n^{1/(2p)-1/4} \xi_n^{1/p-1/2}),$$

In summary

$$\sum_{j=1}^{n} \mathbb{E}\left[Y_{t_{j,n}}\,\Delta_j W \mid \mathcal{F}_{t_{j-1,n}}\right] = O_p(n^{1/(2p)-1/4}\xi_n^{1/p-1/2}).$$

Take $p = 3/2$ (any $p > 1$ is ok for Hölder) to have

$$\sum_{j=1}^{n} \mathbb{E}\left[Y_{t_{j,n}}\,\Delta_j W \mid \mathcal{F}_{t_{j-1,n}}\right] = O_p((\xi_n\, n^{1/2})^{1/6}).$$

Since, by hypothesis, $\xi_n\, n^{7/10} \to 0$ then

$$\xi_n\, n^{1/2} = \frac{\xi_n\, n^{7/10}}{n^{1/5}} \to 0.$$

Hence the condition

$$\sum_{j=1}^{n} \mathbb{E}\left[Y_{t_{j,n}}\,\Delta_j W \mid \mathcal{F}_{t_{j-1,n}}\right] \xrightarrow{\text{p}} \int_0^T a_s\, ds$$

is satisfied with $a \equiv 0$.

The Lindeberg condition

$$\sum_{j=1}^{n} \mathbb{E}\left[Y_{t_{j,n}}^2 \, \mathbb{1}_{\left\{\left|Y_{t_{j,n}}\right|>\varepsilon\right\}} \;\middle|\; \mathcal{F}_{t_{j-1,n}} \right] \xrightarrow{\text{p}} 0,$$

is verified by proving the stronger condition

$$\sum_{j=1}^{n} \mathbb{E}\left[Y_{t_{j,n}}^4 \;\middle|\; \mathcal{F}_{t_{j-1,n}} \right] \xrightarrow{\text{p}} 0$$

that is

$$\frac{n}{\xi_n^2} \sum_{j=1}^{n} \mathbb{E}\left[\left(\Delta_n \, \mathbb{1}_{\left\{\left|\Delta_j X\right| \leqslant \xi_n\right\}} - \sqrt{\frac{2}{\pi}} \, \frac{\xi_n \, \Delta_n^{1/2}}{\sigma_{j-1}} \right)^4 \;\middle|\; \mathcal{F}_{t_{j-1,n}} \right] \xrightarrow{\text{p}} 0.$$

(tedious but easy)

Theorem

Itô-Clark representation. *Let* X *be a random variable and* W *be a Brownian motion defined on the filtered probability space* $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P}\right)$*. If* X *is* $\mathcal{F}_T$*-measurable, then there exists an almost unique process* $\eta \in \mathbb{L}^2(\mathcal{F}_t)$ *such that*

$$X = \mathbb{E}[X] + \int_0^T \eta_s \, dW_s.$$

To prove $\sum_{j=1}^n \mathbb{E}_{j-1}\left[Y_{t_{j,n}}\,\Delta_j N\right] \xrightarrow{\text{p}} 0$ for all N such that $[W,N]_s \equiv 0$ consider

$$\frac{n^{1/4}}{\xi_n^{1/2}} \sum_{j=1}^n \mathbb{E}_{j-1}\left[\left(\Delta_n \mathbb{1}_{\{|\Delta_j X| \leqslant \xi_n\}} - \sqrt{\frac{2}{\pi}}\,\frac{\xi_n\,\Delta_n^{1/2}}{\sigma_{j-1}}\right)\Delta_j N\right] =$$

$$\frac{n^{1/4}}{\xi_n^{1/2}} \sum_{j=1}^n \mathbb{E}_{j-1}\Bigg[\Bigg(\Delta_n \mathbb{1}_{\{|\Delta_j X| \leqslant \xi_n\}} - \mathbb{E}_{j-1}\left[\Delta_n \mathbb{1}_{\{|\Delta_j X| \leqslant \xi_n\}}\right] +$$

$$+\mathbb{E}_{j-1}\left[\Delta_n \mathbb{1}_{\{|\Delta_j X| \leqslant \xi_n\}}\right] - \sqrt{\frac{2}{\pi}}\,\frac{\xi_n\,\Delta_n^{1/2}}{\sigma_{j-1}}\Bigg)\Delta_j N\Bigg] = \sum_{j=1}^n \mathbb{E}_{j-1}\Big[\left(A_{j,n} + B_{j,n}\right)\Delta_j N\Big].$$

$$A_{j,n} = \frac{n^{1/4}}{\xi_n^{1/2}}\left(\Delta_n \mathbb{1}_{\{|\Delta_j X| \leqslant \xi_n\}} - \mathbb{E}_{j-1}\left[\Delta_n \mathbb{1}_{\{|\Delta_j X| \leqslant \xi_n\}}\right]\right) = \int_{t_{j-1,n}}^{t_{j,n}} \eta_s^{(n)}\, dW_s \Rightarrow$$

and $\sum_{j=1}^n \mathbb{E}_{j-1}\Big[A_{j,n}\,\Delta_j N\Big] = 0$. By Taylor expansion...

$$\sum_{j=1}^n \mathbb{E}_{j-1}\Big[B_{j,n}\,\Delta_j N\Big] = \frac{n^{1/4}}{\xi_n^{1/2}} \sum_{j=1}^n \mathbb{E}_{j-1}\Bigg[\left(\mathbb{E}\left[\Delta_n \mathbb{1}_{\{|\Delta_j X| \leqslant \xi_n\}}\right] - \sqrt{\frac{2}{\pi}}\,\frac{\xi_n\,\Delta_n^{1/2}}{\sigma_{j-1}}\right)\Delta_j N\Bigg] \xrightarrow{\text{p}} 0.$$

Assumption (the price process under the alternative). The observed log-price process $\{X_t; t \geqslant 0\}$ is such that $X_{t_0} = X^e_{t_0}$ and, for $j = 2, \dots, n$,

$$X_{t_{j,n}} = X^e_{t_{j,n}}(1 - B_{j,n}) + B_{j,n} X_{t_{j-1,n}},$$

where $X^e_{t_{j,n}}$ is an Ito semimartingale

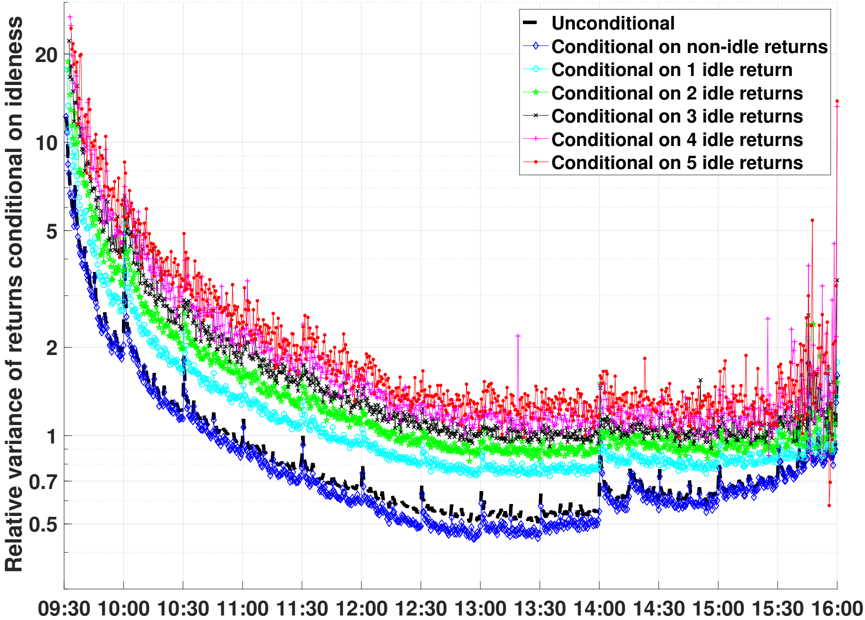
$$dX^e_t = \mu_t \, dt + \sigma_s \, dW_s,$$

and $B_{j,n}$ is a triangular array of (**not necessarily independent**) $\mathcal{F}_{jT/n}-$ measurable Bernoulli variates so that

$$\frac{1}{T} \sum_{j=1}^{n} \Delta_n B_{j,n} \xrightarrow[n \to \infty]{p} p^F,$$

where $p^F \in]0, 1]$. Moreover, denoting by K_n the supremum of the number of consecutive flat trades, we have

$$\frac{K_n}{n} \xrightarrow[n \to \infty]{p} 0.$$


Unconditional
Conditional on non-idle returns
Conditional on 1 idle return
Conditional on 2 idle returns
Conditional on 3 idle returns
Conditional on 4 idle returns
Conditional on 5 idle returns
Relative variance of returns conditional on idleness
20
10
5
2
1
0.7
0.5
09:30
10:00
10:30
11:00
11:30
12:00
12:30
13:00
13:30
14:00
14:30
15:00
15:30
16:00

Under the alternative

$$X_{t_{j,n}} - X_{t_{j-1,n}} = (X^e_{t_{j,n}} - X_{t_{j-1,n}})(1 - B_{j,n})$$

so that we can write

$$\begin{aligned} \mathrm{IT}_n &= \frac{1}{T}\sum_{j=1}^{n} \Delta_n \mathbb{1}_{\{|X_{j\,\Delta} - X_{(j-1)\,\Delta}| \leqslant \xi_n\}} = \frac{1}{T}\sum_{j=1}^{n} \Delta_n \mathbb{1}_{\{|(X^e_{j\,\Delta} - X_{(j-1)\,\Delta})(1-B_{j,n})| \leqslant \xi_n\}} \\ &= \underbrace{\sum_{j=1}^{n} \Delta_n B_{j,n}}_{A_n} + \underbrace{\frac{1}{T}\sum_{j=1}^{n} (1 - B_{j,n}) \Delta_n \mathbb{1}_{\{|X^e_{j\,\Delta} - X_{(j-1)\,\Delta}| \leqslant \xi_n\}}}_{B_n}. \end{aligned}$$

By hypothesis $A_n = \sum_{j=1}^{n} \Delta_n B_{j,n} \xrightarrow{\mathrm{p}} p_F$. Concerning B_n define:

$$K_{j,n} = \min\{k \in \{0, \dots, j\} \,|\, B_{j,n} = 1, B_{j-1,n} = 1, \dots, B_{j-k+1,n} = 1, B_{j-k,n} = 0\}.$$

$K_n \doteq \max_{j=1,\dots,n} K_{j,n}$, which, by assumption

$$\frac{K_n}{n} \xrightarrow{p} 0, \text{ as } n \to \infty,$$

(although this property is used only in the case with noise included in X^e).

Hence, since $t_{j,n} - t_{j-1-K_{j-1},n} \geqslant t_{j,n} - t_{j-1,n} = \Delta_n$, we get

$$\begin{aligned}
& \mathbb{E}\left[1_{\{|X^e_{t_{j,n}} - X_{t_{j-1,n}}| \leqslant \xi_n\}}\right] = \mathbb{E}\left[1_{\{|X^e_{t_{j,n}} - X^e_{t_{j-1-K_{j-1},n}}| \leqslant \xi_n\}}\right] \\
= \; & \mathbb{E}\left[\mathbb{E}\left[1_{\{|X^e_{t_{j,n}} - X^e_{t_{j-1-K_{j-1},n}}| \leqslant \xi_n\}} \,\middle|\, K_{j-1}\right]\right] \\
= \; & \mathbb{E}\left[\sqrt{\frac{2}{\pi}} \frac{\xi_n}{\sqrt{\left(t_{j,n} - t_{j-1-K_{j-1},n}\right)}} \frac{1}{\sigma_{t_{j-1-K_{j-1},n}}} + O_p\left(\frac{\xi_n^3}{\left(t_{j,n} - t_{j-1-K_{j-1},n}\right)^{3/2}}\right)\right] \\
\leqslant \; & \mathbb{E}\left[\sqrt{\frac{2}{\pi}} \frac{\xi_n}{\sqrt{\Delta_n}} \frac{1}{\sigma_{t_{j-1-K_{j-1,n},n}}} + O_p\left(\frac{\xi_n^3}{\Delta_n^{3/2}}\right)\right] \leqslant \sqrt{\frac{2}{\pi}} \frac{1}{C_2} \frac{1}{\sqrt{C_4}} \xi_n n^{1/2} = C\, \xi_n n^{1/2}.
\end{aligned}$$

Hence

$$B_n = \frac{1}{T}\sum_{j=1}^{n}(1 - B_{j,n})\Delta_n \mathbb{1}_{\left\{|X^e_{j\,\Delta} - X_{(j-1)\,\Delta}| \leqslant \xi_n\right\}} \leqslant \frac{1}{T}\sum_{j=1}^{n}\Delta_n \mathbb{1}_{\left\{|X^e_{j\,\Delta} - X_{(j-1)\,\Delta}| \leqslant \xi_n\right\}} \leqslant C\,\xi_n n^{1/2} \xrightarrow{\text{p}} 0.$$

In summary

$$\mathrm{EXIT}_n = \sum_{j=1}^{n}\Bigg(\overbrace{\Delta_n \mathbb{1}_{\left\{|X_{j\,\Delta} - X_{(j-1)\,\Delta}| \leqslant \xi_n\right\}}}^{\xrightarrow{\text{p}} p_F} - \overbrace{\sqrt{\frac{2}{\pi}} \frac{\xi_n \sqrt{\Delta_n}}{\sigma_{t_{j-1}}}}^{\xrightarrow{\text{p}} 0}\Bigg) \xrightarrow{\text{p}} p_F \text{ under } \mathcal{H}_A$$

$$\mathcal{H}_0 : X_t = X_t^e = \int_0^t \mu_s \, ds + \int_0^t \sigma_s \, dW_s, \quad \mathcal{H}_A : X_{j\,\Delta} = X^e_{j\,\Delta}(1 - B_{j,n}) + B_{j,n} X_{(j-1)\,\Delta}$$

Theorem 1. *(Consistency) As $n \to \infty$, let $\xi_n \to 0$ in such a way that $\xi_n \sqrt{n} \to 0$. Then,*

$$\mathrm{IT} = \sum_{j=1}^{n} \Delta_n 1_{\{|X_{t_{j,n}} - X_{t_{j-1,n}}| \leqslant \xi_n\}} \xrightarrow{p} \begin{cases} 0 & \text{under } \mathcal{H}_0 \\ p^F & \text{under } \mathcal{H}_A \end{cases}.$$

(Stable conv.) As $n \to \infty$, let $\xi_n \to 0$ in such a way that $n^{7/10}\xi_n \to 0$ and $\xi_n n^{3/2} \to \infty$. Under $\mathcal{H}_0$:

$$\frac{n^{1/4}}{\xi_n^{1/2}} \mathrm{EXIT} = \frac{n^{1/4}}{\xi_n^{1/2}} \sum_{j=1}^{n} \left(\Delta_n 1_{\{|X_{t_{j,n}} - X_{t_{j-1,n}}| \leqslant \xi_n\}} - \sqrt{\frac{2}{\pi}} \frac{\xi_n \sqrt{\Delta_{n,j}}}{\sigma_{t_{j-1,n}}} \right) \stackrel{stably}{\Longrightarrow} MN\left(0, \sqrt{\frac{2}{\pi}} \int_0^1 \frac{1}{\sigma_s} ds \right).$$

Under $\mathcal{H}_A$:

$$\frac{n^{1/4}}{\xi_n^{1/2}} \mathrm{EXIT} \xrightarrow{p} \infty.$$

$$\mathcal{H}_0' : X_t = \widetilde{X}_t^e = \int_0^t \mu_s \, ds + \int_0^t \sigma_s \, dW_s + \eta_t, \quad \mathcal{H}_A' : X_{j\,\Delta} = \widetilde{X}_{j\,\Delta}^e (1 - B_{j,n}) + B_{j,n} X_{(j-1)\,\Delta}$$

Theorem 2. *(Consistency) As $n \to \infty$, let $\xi_n \to 0$. Then,*

$$\mathrm{IT} = \sum_{j=1}^{n} \Delta_n 1_{\{|X_{t_{j,n}} - X_{t_{j-1,n}}| \leqslant \xi_n\}} \xrightarrow{p} \begin{cases} 0 & \text{under } \mathcal{H}_0' \\ p^F & \text{under } \mathcal{H}_A' \end{cases}$$

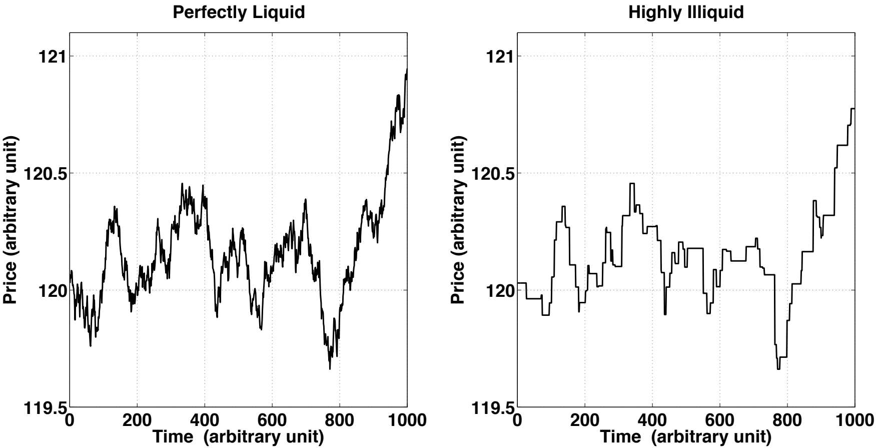
(Stable conv.) As $n \to \infty$, let $\xi_n \to 0$ in such a way that $n^5 \xi_n \to 0$ and $\xi_n n \to \infty$.
Under $\mathcal{H}_0$:

$$\frac{n^{1/2}}{\xi_n^{1/2}} \mathrm{EXIT} = \frac{n^{1/2}}{\xi_n^{1/2}} \sum_{j=1}^{n} \left(\Delta_n 1_{\{|X_{t_{j,n}} - X_{t_{j-1,n}}| \leqslant \xi_n\}} - \sqrt{\frac{2}{\pi}} \frac{\xi_n \sqrt{\Delta_{n,j}}}{\sqrt{2}\,\sigma_\eta} \right) \overset{stably}{\Longrightarrow} N\left(0, \sqrt{\frac{2}{\pi}} \frac{1}{\sqrt{2}\,\sigma_\eta}\right).$$

Under $\mathcal{H}_A$:

$$\frac{n^{1/2}}{\xi_n^{1/2}} \mathrm{EXIT} \xrightarrow{p} \infty.$$

- Transaction costs: assets are not easily exchanged $\Rightarrow$ prices staler than in a frictionless world.
- Informed traders: if transaction costs increase $\Rightarrow$ less trading opportunity $\Rightarrow$ prices even more stale!
- Staleness should be correlated with execution costs and asymmetric information.
- Measure illiquidity as the excess (w.r.t. a frictionless world) of "small" log-price returns.


Perfectly Liquid
Price (arbitrary unit)
121
120.5
120
119.5
0
200
400
600
800
1000
Time (arbitrary unit)
Highly Illiquid
Price (arbitrary unit)
121
120.5
120
119.5
0
200
400
600
800
1000
Time (arbitrary unit)

- Efficient unobserved log-price process $dp_t^e = \sigma\, dW_t$.
- Mid-quote is common knowledge: $m_t = m_{t-1} + \delta(p_t^e - m_{t-1})$. PAIT $\stackrel{def}{=}$ Probability of arrival of informed traders.
- Noise traders (arrive with probability $1 - \text{PAIT}$) and toss a coin

$$p_t = m_t \pm s, \quad s = \text{ half bid-ask spread.}$$

- Informed traders know p_t^e and face total execution costs

$$c = s + f, \quad f = \text{ “funding liquidity”}.$$

  Informed traders trade **iff** they face a profit net of execution costs c, hence

$$p_t = m_t + (p_{t-1} - m_t)\, 1_{\{|p_t^e - m_t| \leqslant c\}} + s\, 1_{\{p_t^e - m_t > c\}} - s\, 1_{\{p_t^e - m_t < -c\}}$$

- $p_F = \text{function}\,(\mathrm{PAIT}, s, f, \sigma)$.

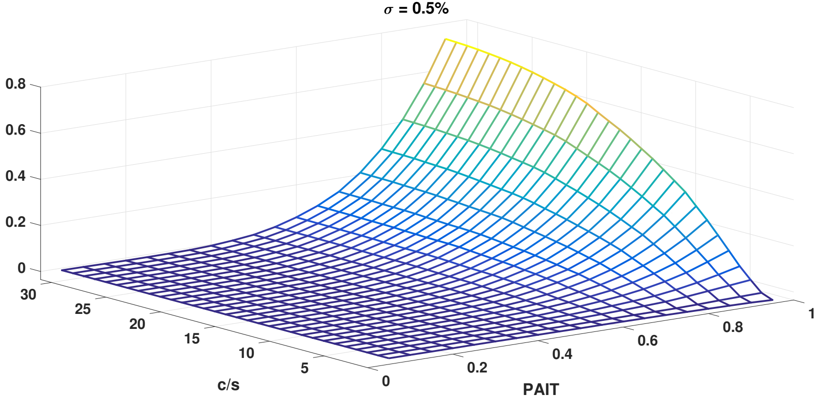


Figure: EXIT is computed on 5-minute intervals (with $\xi = \frac{1}{20}\sigma$) and plotted as a function of the execution cost c (standardized by the bid-ask spread s) and PAIT, the probability of arrival of informed traders. The parameters δ and s are mean (daily) estimates from data: $\delta = 0.0118$ and $s = 1.8654 \cdot 10^{-4}$.

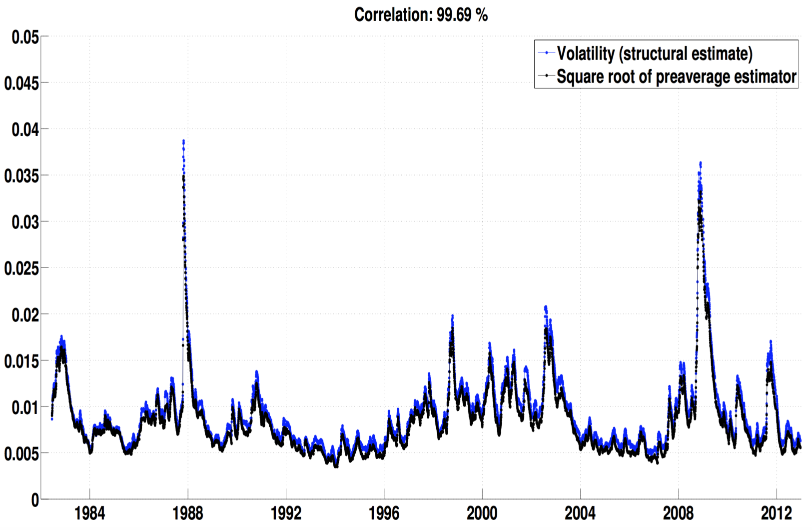
Assume to have $\{p_1, ..., p_t\}$ observed intra-day (e.g. 5-minutes) transaction prices.

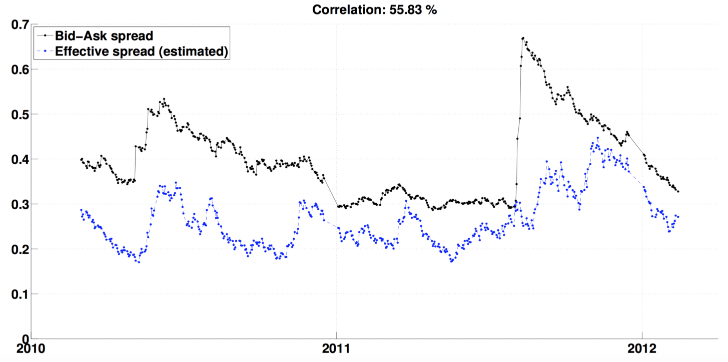
Indirect inference á la Gourieroux, Monfort, Renault (1993)

Estimate parameters of micro-structural model by matching observed moments with model-implied ones.

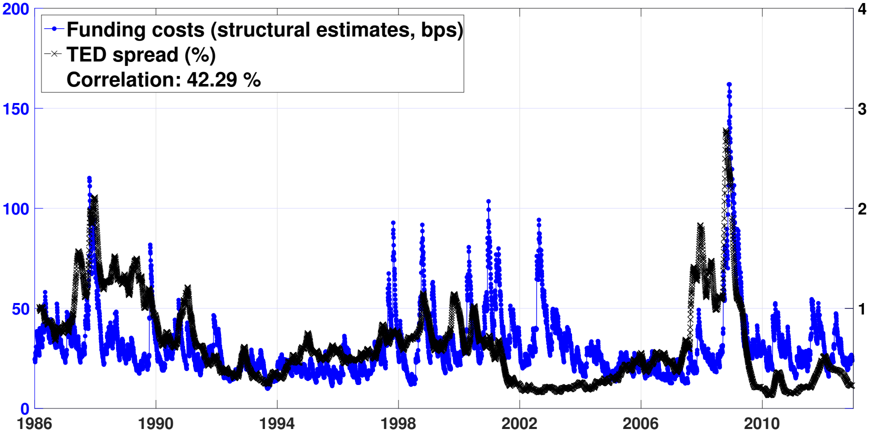
- The variance of one-minute returns, $\mathcal{M}_1$.
- The first-order auto-covariance of one-minute returns, $\mathcal{M}_2$.
- The variance of five-minute returns, $\mathcal{M}_3$.
- EXIT computed with one-minute returns, $\mathcal{M}_4$.
- EXIT computed with two-minute returns, $\mathcal{M}_5$.

Empirical application: indirect inference.





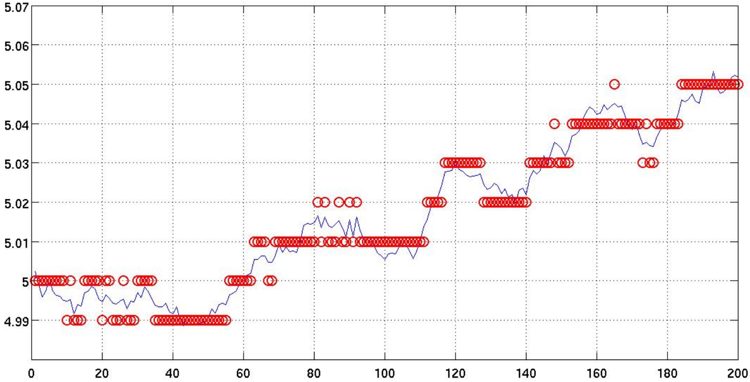
Trades may occur within the spread ⇒ it is unsurprising to find estimated effective spreads which are, on average and for virtually each day, smaller than the corresponding half quoted spreads.


Funding costs (structural estimates, bps)
TED spread (%)
Correlation: 42.29 %
200
150
100
50
0
4
3
2
1
1986
1990
1994
1998
2002
2006
2010

- Express the latent price P_t as the discretized analogue (over an interval of length Δ) of the local martingale

$$P_t = \int_0^t \sigma_s P_s dW_s.$$

- Assume the traded price $\overline{P}_t$ can be only observed as a multiple of a fixed quantity d (e.g., for NYSE-listed stocks d is one cent) and $\overline{P}_t$ is obtained by rounding P_t to the nearest value in the set $\{0, d, 2d, \ldots, kd, \ldots\}$.


5.07
5.06
5.05
5.04
5.03
5.02
5.01
5
4.99
0
20
40
60
80
100
120
140
160
180
200

For the time being, we de-activate the Bernoulli variates and set them equal to zero. The probability that $\overline{P}_{(i+1)\Delta_n} - \overline{P}_{i\Delta_n} = kd$, conditional on spot volatility $\sigma_{i\Delta_n}$, the rounded price $\overline{P}_{i\Delta_n}$ and the displacement value $x_{i\Delta_n}$, where $\overline{P}_{i\Delta_n} = P_{i\Delta_n} - x_{i\Delta_n}$ with $-d/2 < x_{i\Delta_n} < d/2$, is

$$
\begin{aligned}
& \mathbb{P}\left[\overline{P}_{i\Delta_n} + kd - \frac{d}{2} < P_{(i+1)\Delta_n} < \overline{P}_{i\Delta_n} + kd + \frac{d}{2} \,\middle|\, \sigma_{i\Delta_n}, \overline{P}_{i\Delta_n}, x_{i\Delta_n}\right] \\
= \quad & \mathbb{P}\left[P_{i\Delta_n} + kd - \frac{d}{2} - x_{i\Delta_n} < P_{(i+1)\Delta_n} < P_{i\Delta_n} + kd + \frac{d}{2} - x_{i\Delta_n} \,\middle|\, \sigma_{i\Delta_n}, \overline{P}_{i\Delta_n}, x_{i\Delta_n}\right] \\
\approx \quad & \int_{kd-\frac{d}{2}-x_{i\Delta_n}}^{kd+\frac{d}{2}-x_{i\Delta_n}} \frac{e^{-\frac{z^2}{2\sigma^2_{i\Delta_n}(\overline{P}_{i\Delta_n}+x_{i\Delta_n})^2\Delta_n}}}{\sqrt{2\pi\sigma^2_{i\Delta_n}(\overline{P}_{i\Delta_n}+x_{i\Delta_n})^2\Delta_n}}dz = \int_{kd-\frac{d}{2}-x_{i\Delta_n}}^{kd+\frac{d}{2}-x_{i\Delta_n}} \frac{e^{-\frac{z^2\zeta^2_{i\Delta_n}}{2d^2}}}{\sqrt{2\pi\frac{d^2}{\zeta^2_{i\Delta_n}}}}dz,
\end{aligned}
$$

where $\zeta_{i\Delta_n} = d/(\sigma_{i\Delta_n} P_{i\Delta_n}\sqrt{\Delta_n})$ is a "rounding impact ratio" defined as the magnitude of price discreteness (i.e., d) relative to the volatility of the return process over Δ_n (i.e., $\sigma_{i\Delta_n} P_{i\Delta_n}\sqrt{\Delta_n}$).

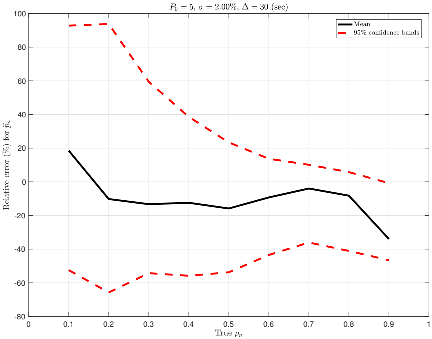
- Assume that $x = P - \overline{P}$ is uniformly distributed over $[-d/2, d/2]$.
- We set $\zeta_{i\Delta_n} = \zeta_{\Delta_n}$.
- Integrating x out, we evaluate the probability $p^{k,R}_{\Delta_n}(\zeta_{\Delta_n})$ of observing a k-tick movement over an interval Δ_n by virtue of

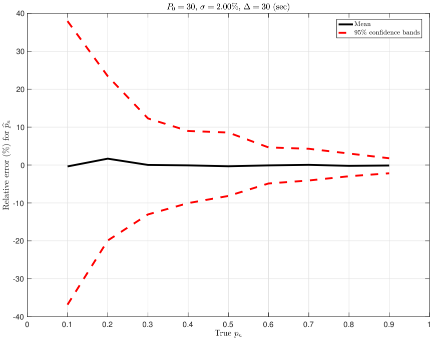
$$
\begin{aligned}
p^{k,R}_{\Delta_n}(\zeta_{\Delta_n}) = \quad & (k-1)\mathrm{erf}\left(\frac{(k-1)\zeta_{\Delta_n}}{\sqrt{2}}\right) - 2k \cdot \mathrm{erf}\left(\frac{k\zeta_{\Delta_n}}{\sqrt{2}}\right) \\
& +(1+k)\mathrm{erf}\left(\frac{(k+1)\zeta_{\Delta_n}}{\sqrt{2}}\right) \\
& +\sqrt{\frac{2}{\pi}}\left(e^{-\frac{1}{2}(1+k)^2\zeta^2_{\Delta_n}}\left(1 + e^{2k\zeta^2_{\Delta_n}} - 2e^{-\frac{1}{2}(1+2k)\zeta^2_{\Delta_n}}\right)\right)\frac{1}{\zeta_{\Delta_n}},
\end{aligned}
$$

  where $k = 0, \pm 1, \ldots$ and the symbol $\mathrm{erf}(x)$ defines the Gaussian error function.
- Re-activating staleness through independent Bernoulli variates, the probability of a k-tick movement becomes

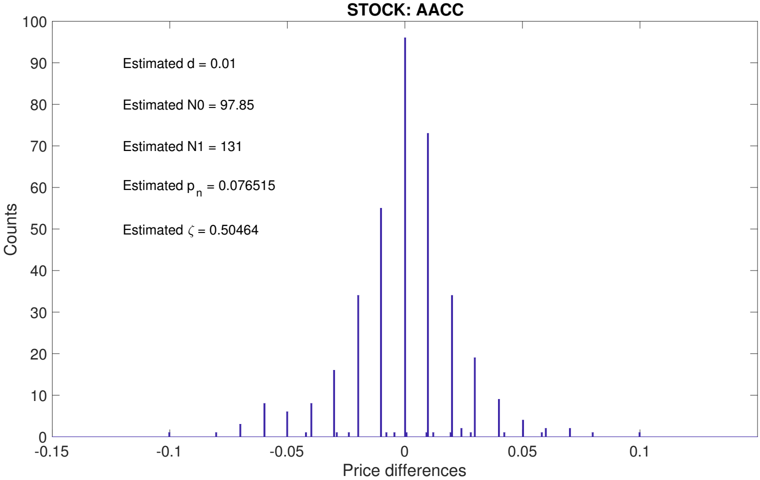
$$
\tilde{p}^{k,R}_{\Delta_n}(\zeta_{\Delta_n}, p_n) = p_n \mathbb{1}_{\{k=0\}} + \sum_{j=0}^{\infty}(1-p_n)^2 p_n^j \cdot p^{k,R}_{\Delta_n}\left(\zeta_{\Delta_n}/\sqrt{j+1}\right). \qquad (0.1)
$$

Simulations

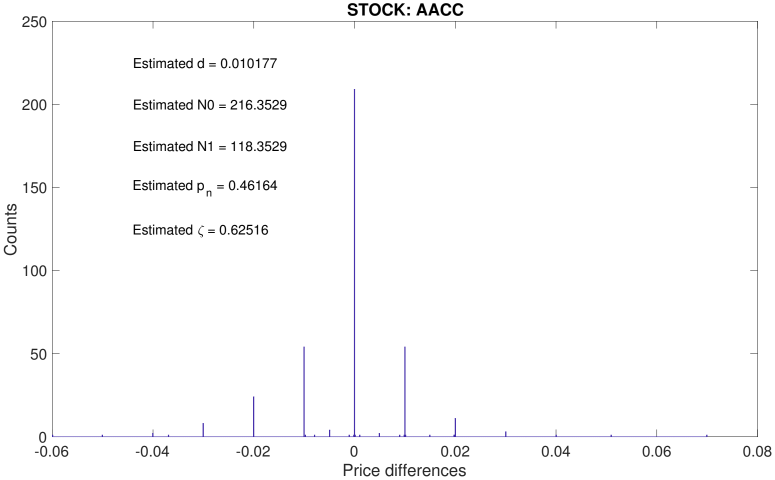




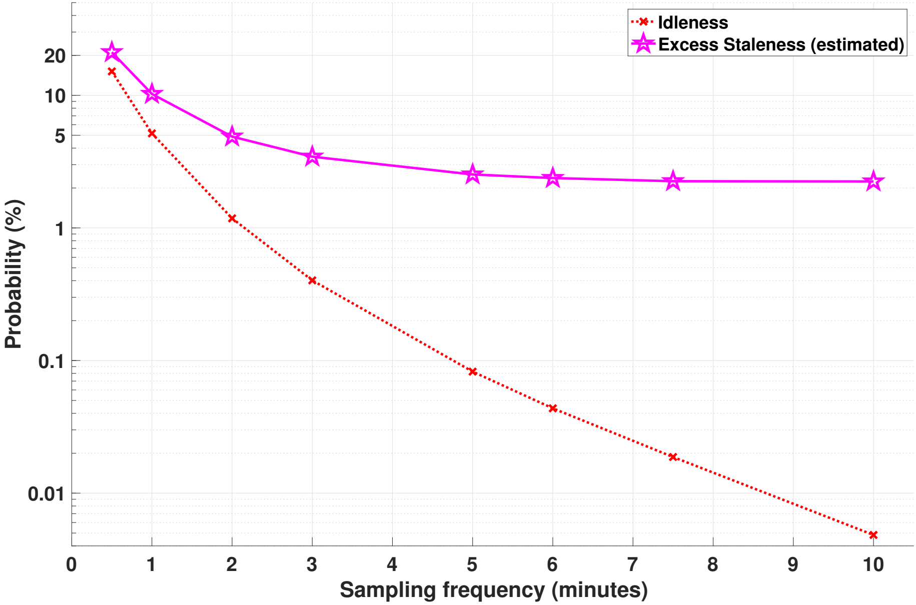
The "liquid" case

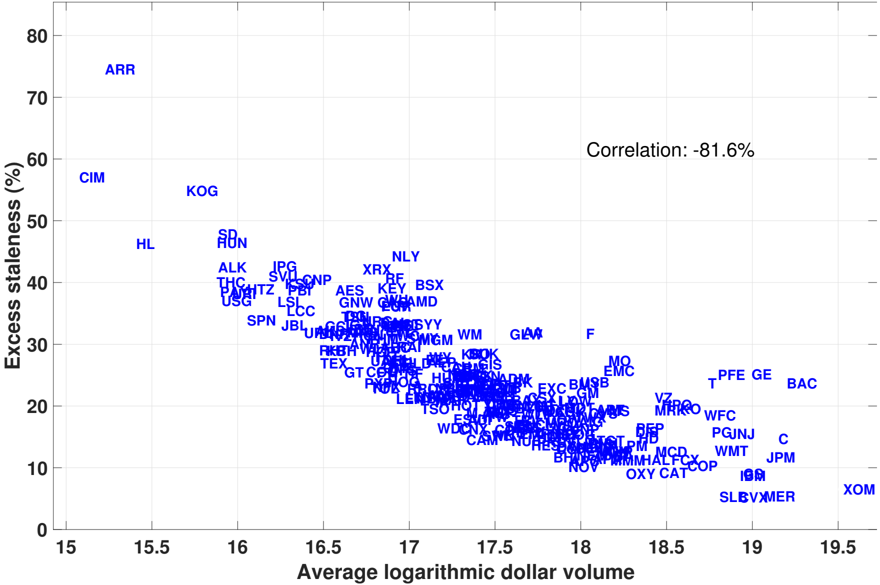


The "illiquid" case



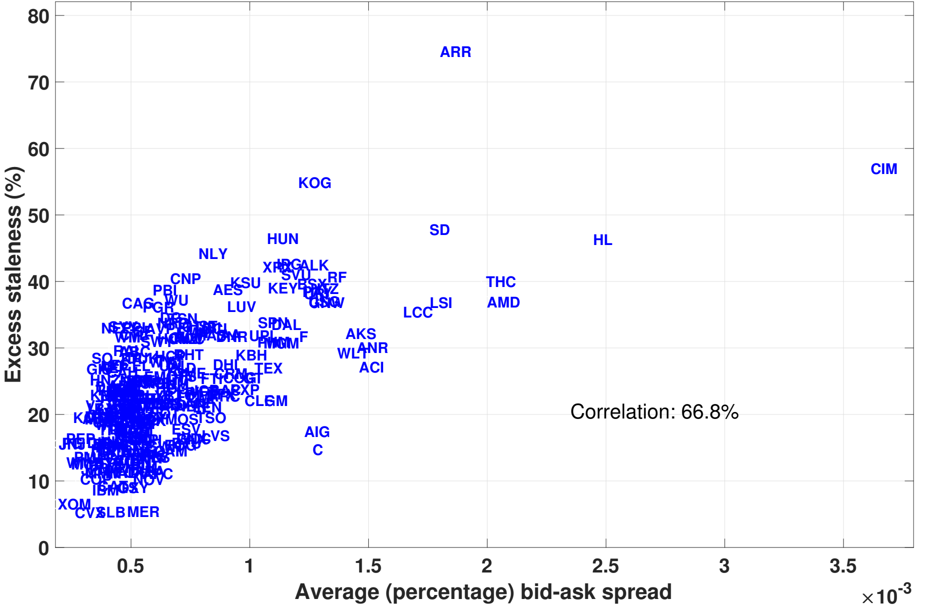
Average excess staleness vs average idleness





Correlation: -81.6%
Excess staleness (%)
Average logarithmic dollar volume
0
10
20
30
40
50
60
70
80
15
15.5
16
16.5
17
17.5
18
18.5
19
19.5
ARR
CIM
KOG
HL
HUN
ALK
IPG
CNP
THC
HTZ
USG
SPN
LCC
JBL
TEX
NLY
XRX
RF
BSX
AES
KEY
GNW
AMD
WM
F
MO
EMC
EXC
PFE
GE
BAC
T
VZ
WFC
PG
JNJ
C
WMT
JPM
MCD
COP
OXY
CAT
NOV
XOM
SLB
MER

Relation with bid-ask spread



- Excess staleness satisfies the following cross-sectional regression:

$$\hat{p}_n = \underset{(7.00)}{0.807} - \underset{(-18.10)}{0.085} < \log DV > + \underset{(11.53)}{0.012} < \text{bid-ask} > - \underset{(-10.90)}{0.065} < \log RV > + \hat{\varepsilon},$$

where we regress (across stocks) averages of daily estimates of 30-second excess staleness, $\hat{p}_n$, on averages of daily logarithmic dollar volume, $< \log DV >$, averages of logarithmic bid-ask spreads, $< \text{bid-ask} >$ in basis points, and averages of daily logarithmic 5-minute realized variances, $< \log RV >$.

- This is slightly puzzling.

- But if we take into account volume clustering:

$$\hat{p}_n = \underset{(12.98)}{3.389} - \underset{(-23.03)}{0.091} < \log DV > + \underset{(2.00)}{0.003} < \text{bid-ask} > - \underset{(-0.26)}{0.002} < \log RV >$$
$$+ \underset{(10.63)}{0.276} < \log CV > + \hat{\varepsilon}.$$

where we insert the logarithmic coefficient of variation $< \log CV >$ of volume.

- We construct an *excess staleness factor* in a traditional way. At time t, we sort stocks into deciles using the excess staleness observed over the previous month. We then construct equally-weighted decile portfolios. The excess staleness factor is the difference between the return on the top-decile portfolio and the return on the bottom-decile portfolio, which we label as $R_{ES,t}$.
- We use monthly rebalancing (22 days) and regress the monthly returns of this long-short strategy on a state-of-the-art, 5-factor, Fama-French model in which risk is captured by the market ($R_{M,t} - R_{F,t}$, with $R_{F,t}$ denoting the risk-free rate), size (SMB_t), value (HML_t), profitability (RMW_t) and investment (CMA_t).
- The output of the regression is

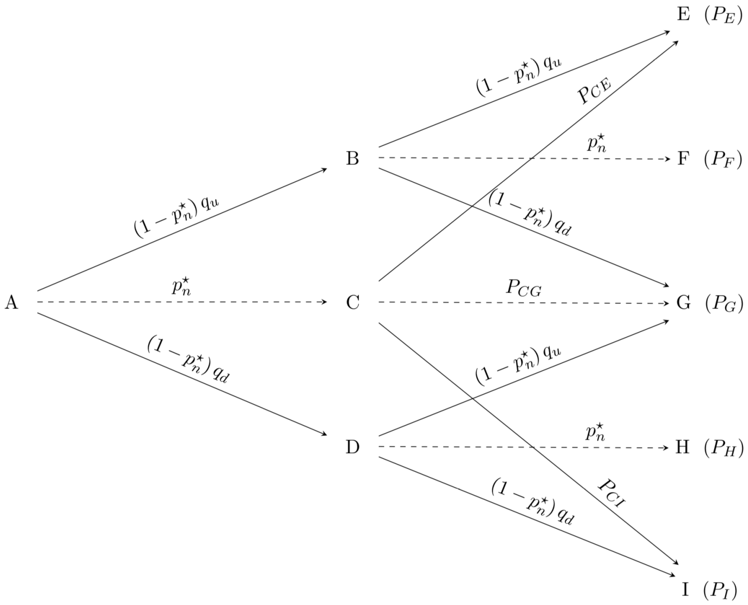
$$R_{ES,t} - R_{F,t} = \underset{(3.15)}{0.0108} - \underset{(-0.81)}{0.0770}\,(R_{M,t} - R_{F,t}) + \underset{(3.51)}{0.8056}\,\mathrm{SMB}_t + \underset{(1.16)}{0.2466}\,\mathrm{HML}_t - \underset{(-1.66)}{0.6817}\,\mathrm{RMW}_t + \underset{(1.83)}{0.8286}\,\mathrm{CMA}_t + \hat{\varepsilon}_t.$$

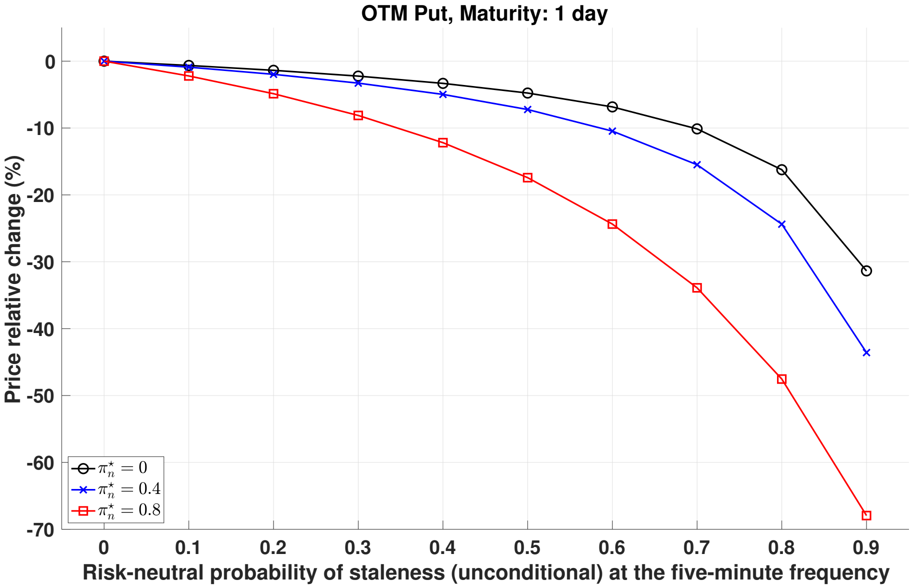
- The positive and significant value of the intercept suggests that excess staleness leads to yet another anomaly hardly explained by well-accepted risk factors.

In light of the dependence between excess staleness and volume (levels and variability), we now relate the returns on the long-short excess staleness portfolio to the returns on a long-short volume portfolio (i.e., a portfolio long stocks in the high volume decile and short stocks in the low volume decile) and the returns on a long-short volume-CV portfolio (i.e., a portfolio long stocks in the high volume-CV decile and short stocks in the low volume-CV decile):

$$R_{ES,t} - R_{F,t} = \underset{(0.62)}{0.0019} - \underset{(-7.74)}{0.5423}\ R_{DV,t} + \underset{(5.42)}{0.2833}\ R_{CV,t} + \hat{\varepsilon}_t.$$

The premium disappears, showing that staleness carries information about the dynamics of trading volume also from a pricing perspective.


A
B
C
D
E (P_E)
F (P_F)
G (P_G)
H (P_H)
I (P_I)
$(1-p_n^\star)\,q_u$
$p_n^\star$
$(1-p_n^\star)\,q_d$
$(1-p_n^\star)\,q_u$
P_{CE}
$p_n^\star$
$(1-p_n^\star)\,q_d$
P_{CG}
$(1-p_n^\star)\,q_u$
$p_n^\star$
P_{CI}
$(1-p_n^\star)\,q_d$


OTM Put, Maturity: 1 day
Price relative change (%)
0
-10
-20
-30
-40
-50
-60
-70
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Risk-neutral probability of staleness (unconditional) at the five-minute frequency
$\pi_n^\star = 0$
$\pi_n^\star = 0.4$
$\pi_n^\star = 0.8$