Modelling and forecasting trading volume

Mstislav Elagin

with Profs. Nikolaus Hautsch and Vladimir Spokoiny

QPL

2010-02-11

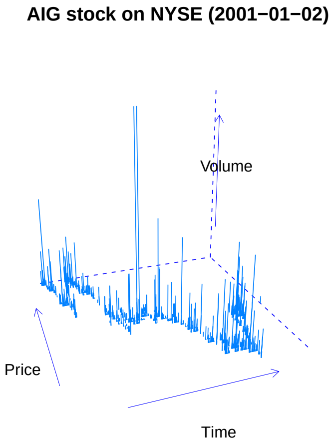
Forecasting for algo trading

- 2006: over 40% of all orders at the London Stock Exchange were entered by algo traders
- 2009: high frequency trading firms account for 73% of all US equity trading volume

(Source: `http://en.wikipedia.org/wiki/Algorithmic_trading`)

- Successful algo trading requires accurate forecasting

VWAP and volume forecasting

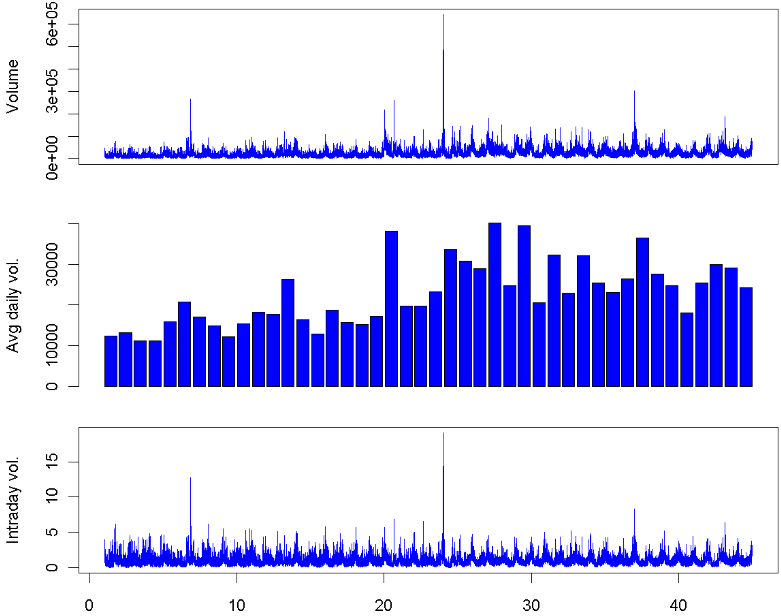


$$\text{VWAP} = \frac{\sum_j p_j v_j}{\sum_j v_j}$$

Stylized facts: Apple

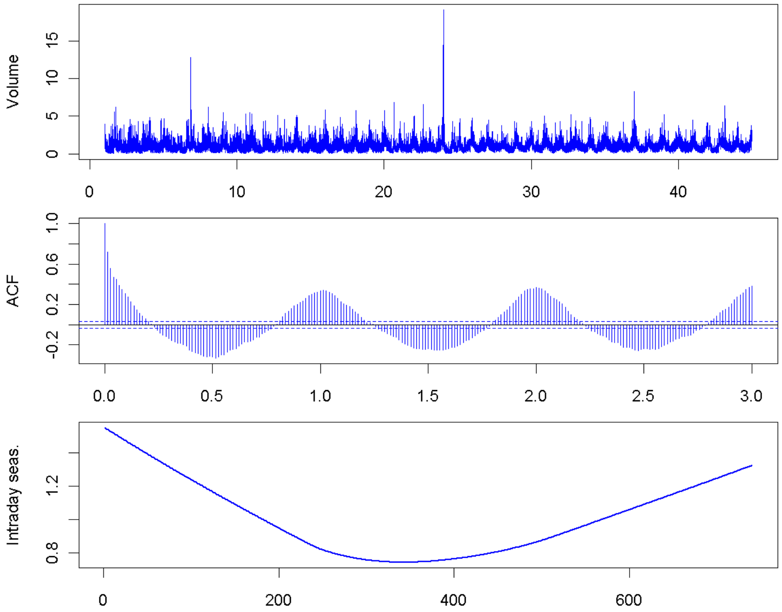
- AAPL.OQ (Apple stock on NASDAQ)
- Time frame: 2009-09-02 – 2009-10-30
- Aggregation: 30 sec

Daily trend

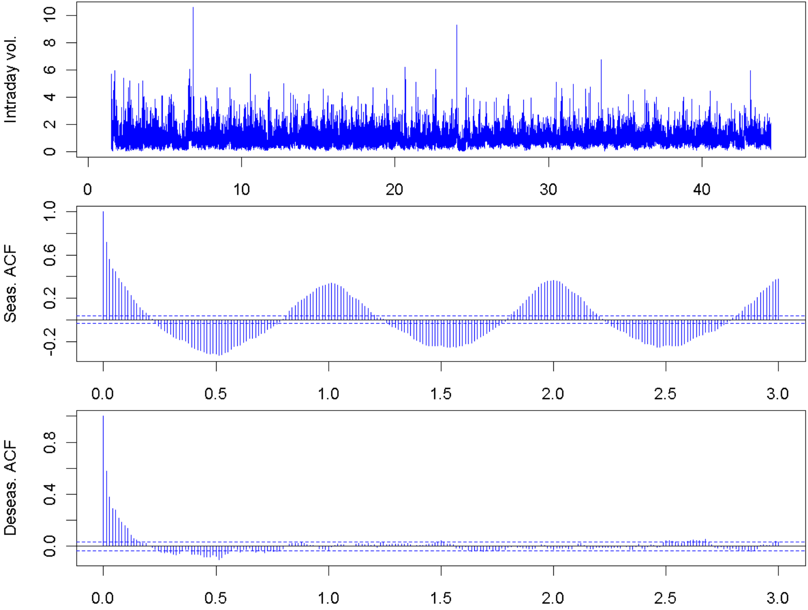


Volume
0e+00
3e+05
6e+05
Avg daily vol.
0
10000
30000
Intraday vol.
0
5
10
15
0
10
20
30
40

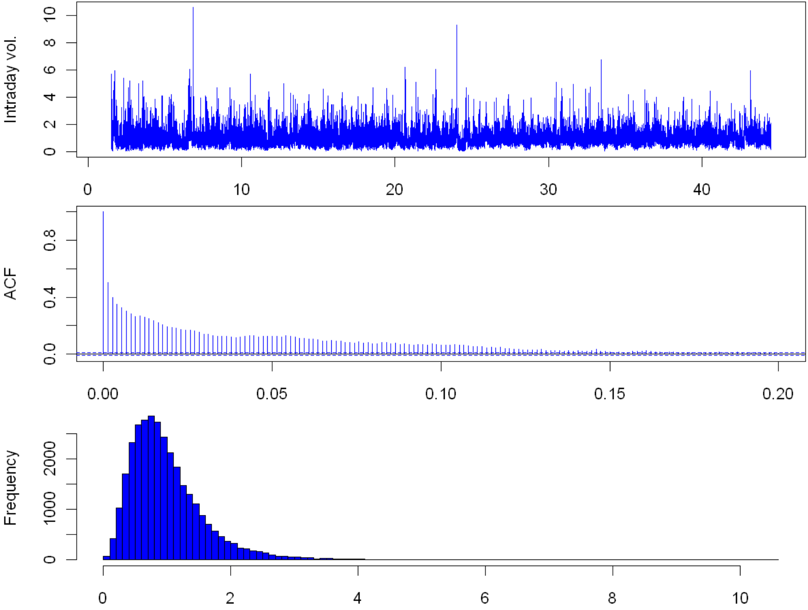
Intraday component and seasonality



Deseasonalised intraday component



Intraday non-periodic component



Autoregressive conditional durations (ACD) model

- Multiplicative error model:

$$x_t = \psi_t \epsilon_t$$

- Conditional distribution of the errors:

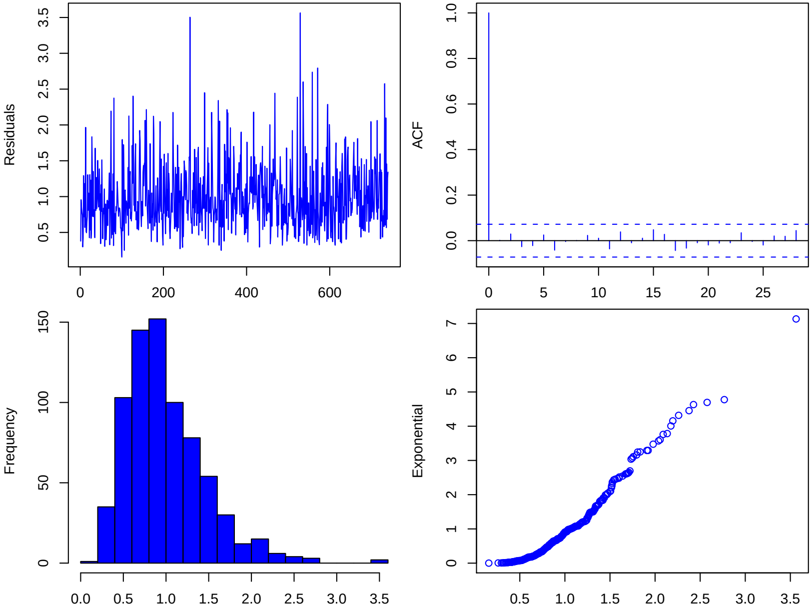
$$\mathcal{L}\big(\epsilon_t | \mathcal{F}_{t-1}\big) = \mathrm{Exp}(1)$$

- ARMA-type dynamics of the conditional durations:

$$\psi_t = \omega + \sum_{j=1}^{r} \gamma_j x_{t-j} + \sum_{j=1}^{s} \omega_j \psi_{t-j}$$

- Conditions on coefficients (non-negativity)
  $\theta = (\omega_0, \gamma, \omega) \quad \widehat{x} = \arg\max_\theta L(\theta)$

ACD(2,1) in-sample diagnostics (day 20)



Adaptive modelling

- x_t – 1-dimensional stochastic process in discrete time, progressively measurable w.r.t. $(\mathcal{F}_t)$ (filtration generated by past data)
- Conditional distribution given the past:

$$\mathcal{L}(x_t|\mathcal{F}_{t-1}) = \mathrm{Exp}(\theta_t), \quad \theta_t \in \Theta \subset \mathbb{R}$$

- Local constant parametric assumption:

$$\theta_t \approx \theta, \quad t \in \mathcal{I} = [T - N_{\mathcal{I}}, T],$$

  T is the point of estimation.
- Goal: identify $\mathcal{I}$ and estimate θ.

"Weak" maximum likelihood estimates

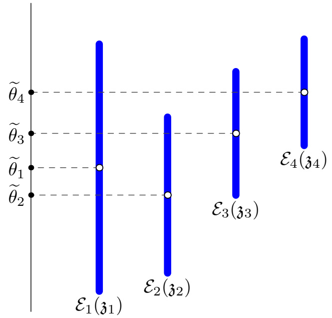
- Maximum likelihood estimate:

$$\widetilde{\theta}_{\mathcal{I}} = \arg\max_{\theta} \sum_{t \in \mathcal{I}} \ell(Y_t, \theta),$$

where ℓ is the log-density of the exponential family

- $\mathcal{I}_1 \subset \mathcal{I}_2 \subset \ldots \subset \mathcal{I}_K$ – sequence of candidate intervals
- $\widetilde{\theta}_1, \ldots, \widetilde{\theta}_K$ – associated weak ML estimates

Description of the method



- Confidence set:

$$\mathcal{E}_k(\mathfrak{z}) = \{\theta \in \Theta : L(\widetilde{\theta}_k, \theta) \leq \mathfrak{z}\}$$

- $\widetilde{\theta}_k$ is accepted if $\widetilde{\theta}_k \in \mathcal{E}_\ell$ for all $\ell < k$
- $\widehat{\theta} = \widetilde{\theta}_{\widehat{k}}$, where $\widehat{k} = \max k$ satisfying the previous condition.

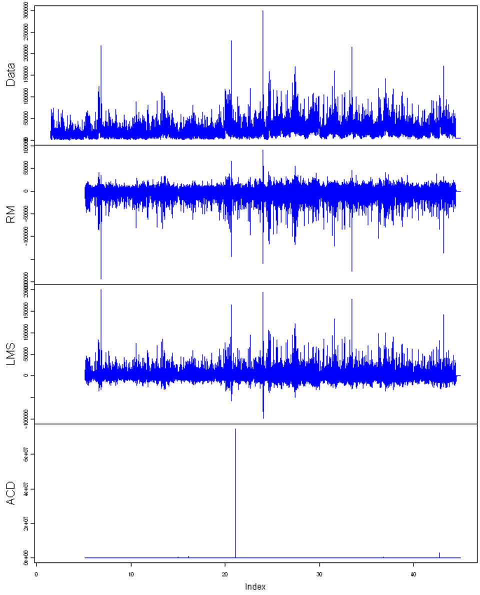
Diebold – Mariano test

- Forecasting errors:
  $\varepsilon^{(i)}_{t+h|t} = x_{t+h} - \widehat{x}^{(i)}_{t+h|t}, \quad i = 1, 2. \quad t = t_0, \ldots, T$
- Loss function $L(\cdot) : (\cdot)^2, \quad |\cdot|$
- $H_0 : \boldsymbol{E} L(\varepsilon^{(1)}_{t+h|t}) = \boldsymbol{E} L(\varepsilon^{(2)}_{t+h|t}) \quad$ vs. $\quad \neq$
- Test statistic:

$$S = \frac{\overline{d}}{[\widehat{\mathrm{avar}}\{\overline{d}\}]^{1/2}}, \quad \overline{d} = \mathrm{mean}_t(L^{(1)}_t - L^{(2)}_t)$$

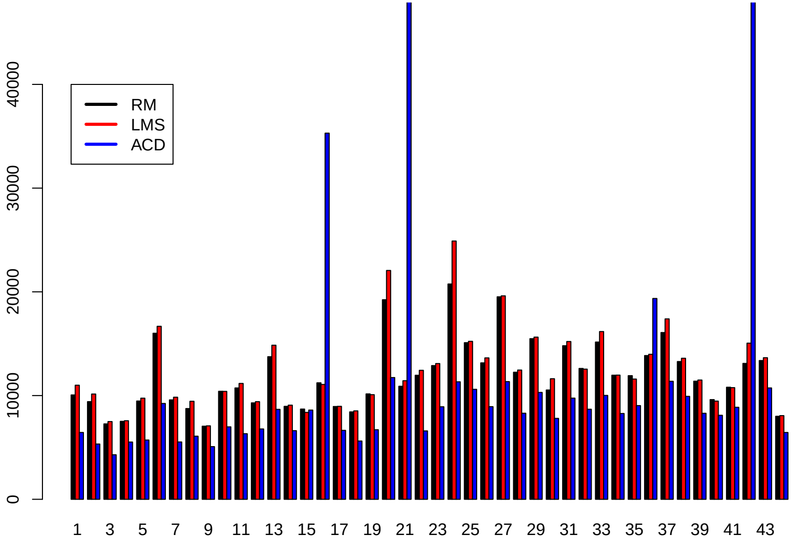
- Under H_0 the test statistic is asymptotically $\mathrm{N}(0, 1)$.

Forecasting performance (full set)

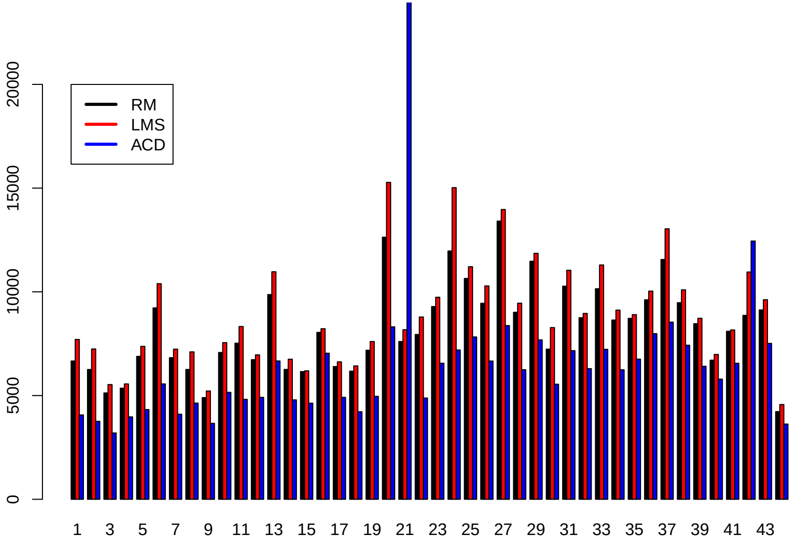


Criterion	RM	LMS
RMSE	12642	12945***
MAE	8478	8939***

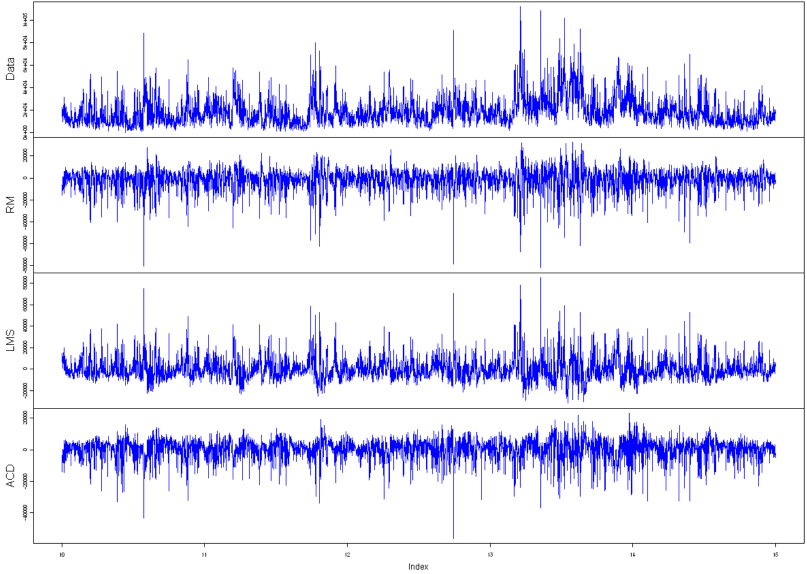
Forecasting performance (RMSE)



Forecasting performance (MAE)



Forecasting performance (days 10-14)



Forecasting performance (days 10-14)

Model	RMSE	MAE
RM	10763	7492
LMS	10985*	7944***
ACD	7122***	5269***

Static VWAP replication

- Notation: $x_{t,i}$ is the i-th observation on day t
- $\mathrm{VWAP} = \frac{\sum_i p_i v_i}{\sum_i v_i}$
- Slicing strategy:

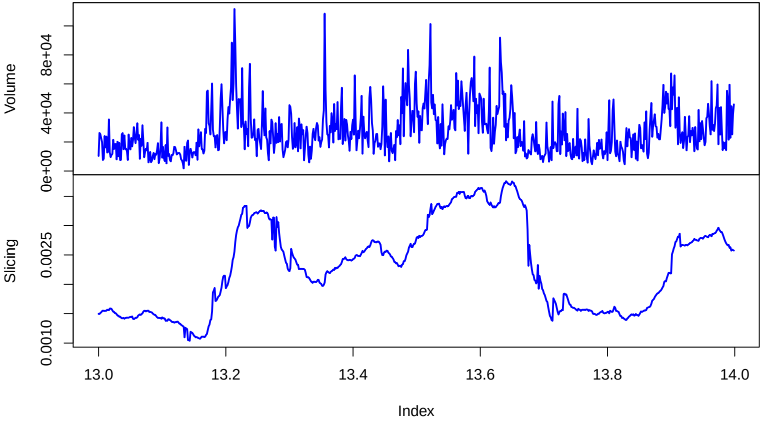
$$\widehat{w}_{t,i|t-1} = \frac{\widehat{x}_{t,i|i-1}}{\sum_{j=1}^{I} \widehat{x}_{t,j|t-1}}, \quad i = 1 \dots I$$

- Average execution price:

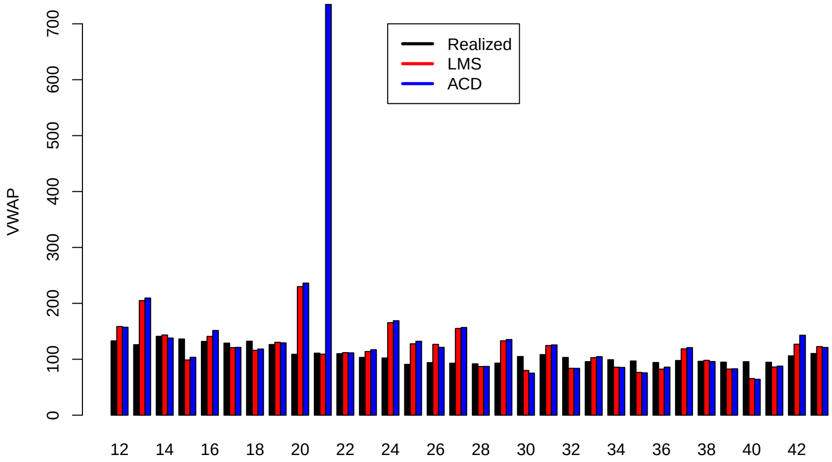
$$\mathrm{AEP_t} = \sum \widehat{w}_{t,i|t-1} \overline{p}_{t,i},$$

where $\overline{p}_{t,i}$ is the realized VWAP.

Slicing on day 13



Forecasting quality in terms of VWAP



Conclusion

- ACD is better on “calm” data, but possibly unstable
- local linear, quadratic etc. models; trend forecasting
- simultaneous modelling of multiple characteristics