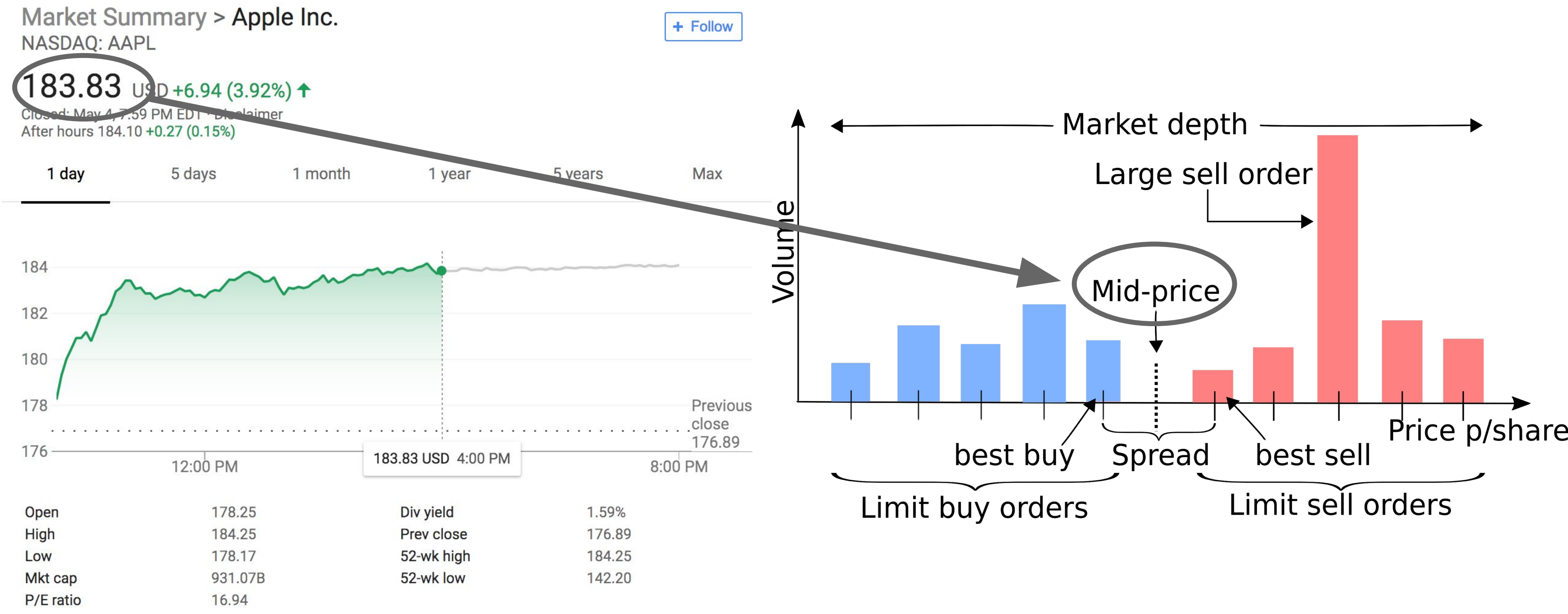
High Frequency Price Movement Strategy

Adam, Hujia, Joel, Jorge

Limit Order Book (LOB)

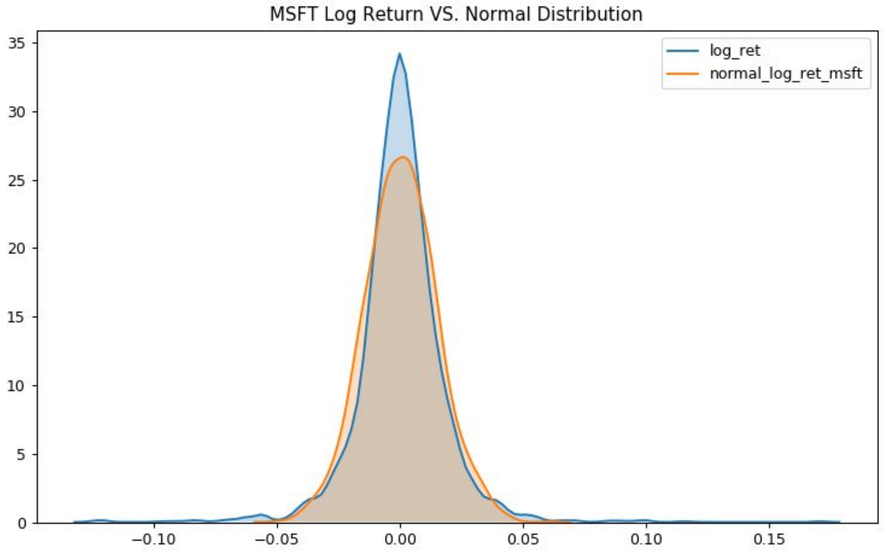
- **Top of the Book** - highest bid and the lowest ask orders
- **Price levels** - several orders at the same price
- **Book depth** - number of price levels available at a particular time in the book
- The LOB data gives traders insight into supply and demand of market microstructure, and short-term price movements

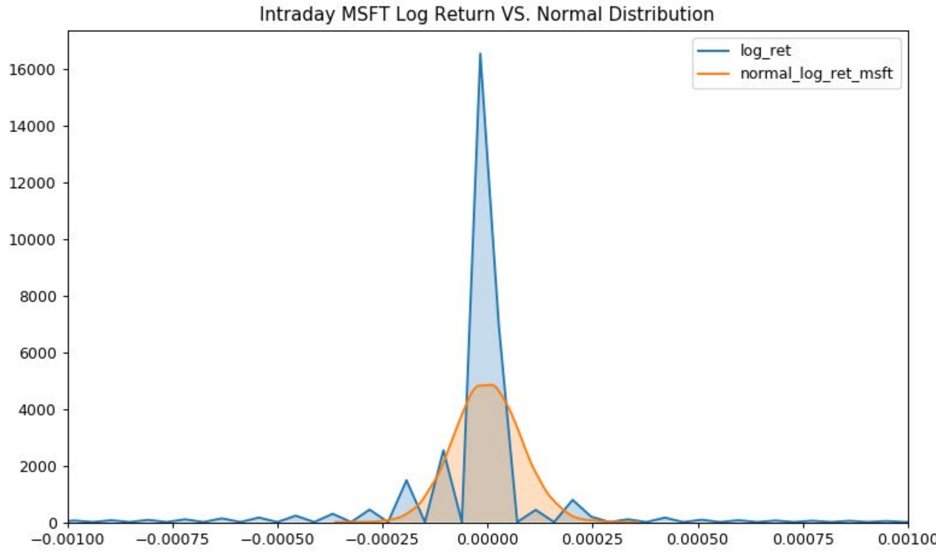


Limit Order Book [https://nms.kcl.ac.uk/rll/enrique-miranda/index.html]

High Frequency Price vs. Daily Price (MSFT)

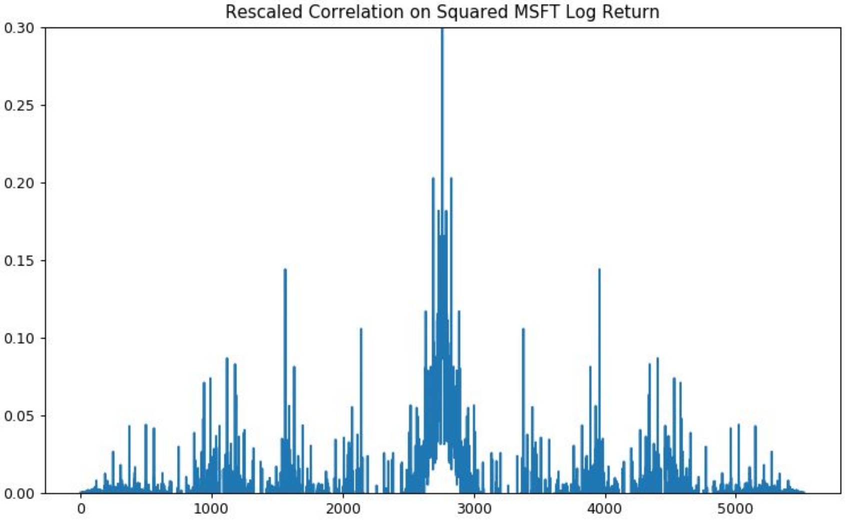
- HF return - significantly smaller mean and variance, but sharper peak and fatter tail
- [Left] Daily return: μ: 3.1e-4, σ: 0.0174 [right]: High frequency return: μ: 8.7e-08, σ: 7.9e-05

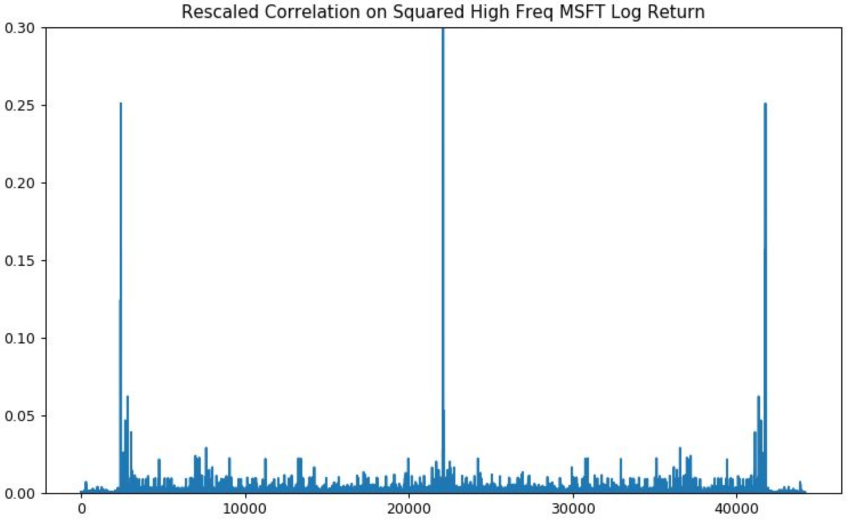




Autocorrelation

- [Left] Daily Price [Right] High Frequency Price
- High frequency log return - **significantly less autocorrelation** - fails to meet strong autocorrelation assumption of time series models.





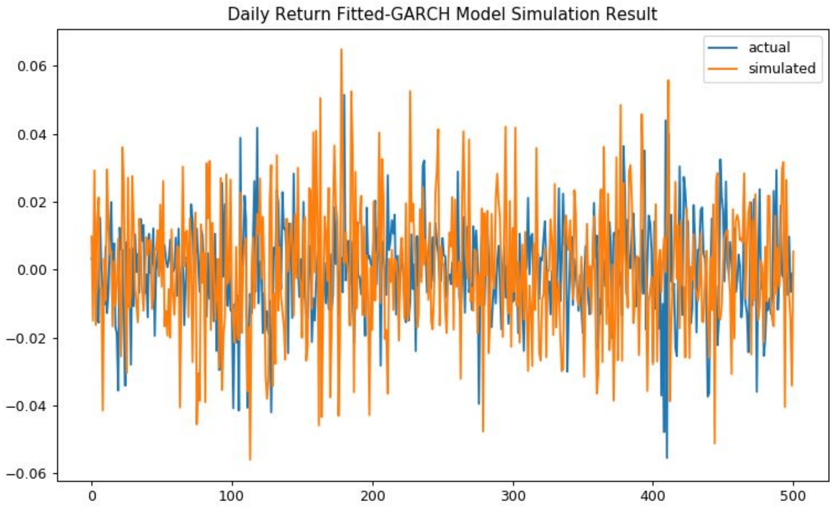
GARCH Simulation

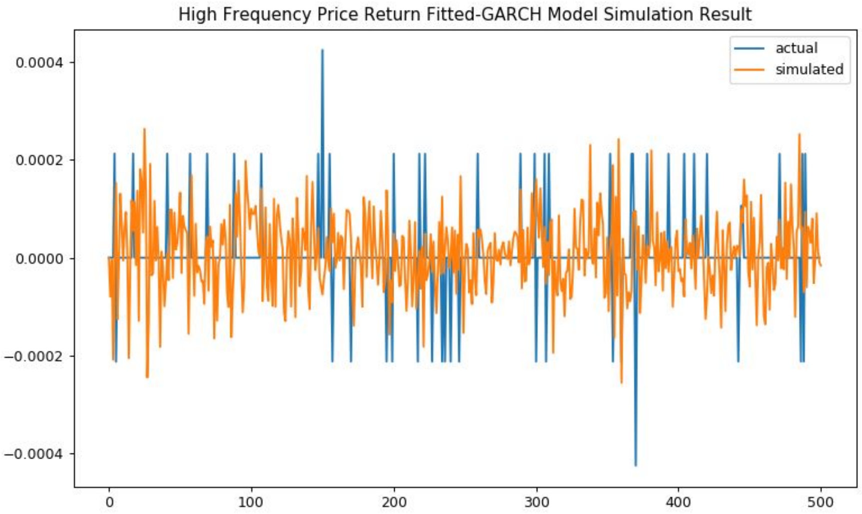
$$\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta\sigma_{t-1}^2$$

DOES NOT CONVERGE !!! :(

- params_daily = [6.36e-06, 0.05, 0.93]

params_intraday = [6.11e-10, 0.05, 0.85]





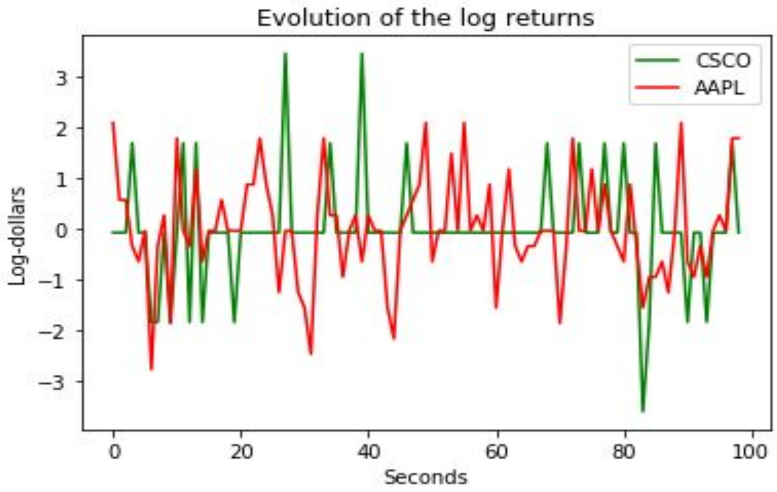
Conclusion: Time series models can still be fitted to high frequency data. **Cons:** (1) suboptimal parameters due to failure to converge. (2) can't model discrete / tick-size or zero price return.

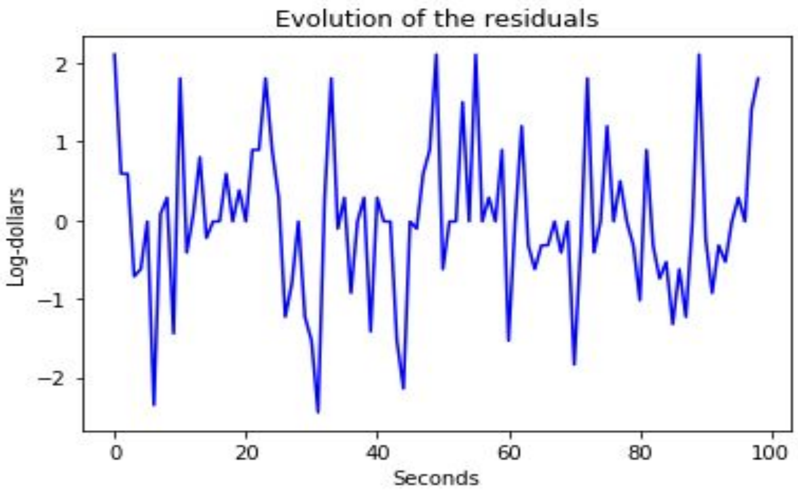
Stat Arb: Pairs trading based on Avellaneda-Lee

- High-frequency pairs trading [1]
- Requires correlation between returns
- Three steps of the algorithm:
  - Identify pairs with high correlation
  - Regress the returns and model residuals
  - Identify temporary mispricings

Measuring correlations

- Linearly **regress the midprice** returns of a pair of stocks
- **Obtain residuals** of the regression





Identifying mispricings

- Arbitrage opportunity if residuals “significatively” **diverge from 0**
- Residuals follow **Ornstein-Uhlenbeck (OU)** process:

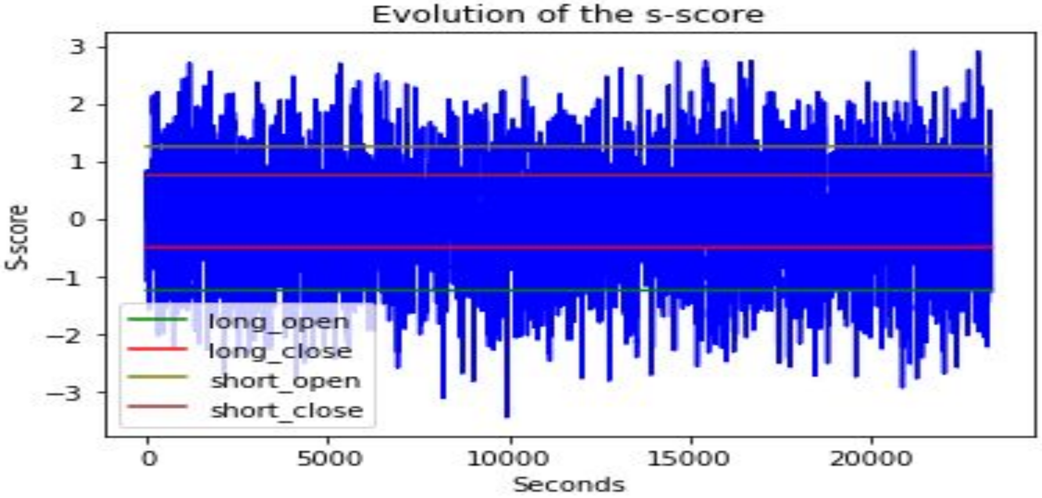
$$dr_t = \kappa(a - r_t)dt + \sigma dW_t$$

- Fit the parameters every second via an **AR(1) model**
- **Mispricing** if the last observation is far from the equilibrium

|S-score| = |(r_100 - mean(OU))/standard_deviation(OU)| > threshold

Execution of trades

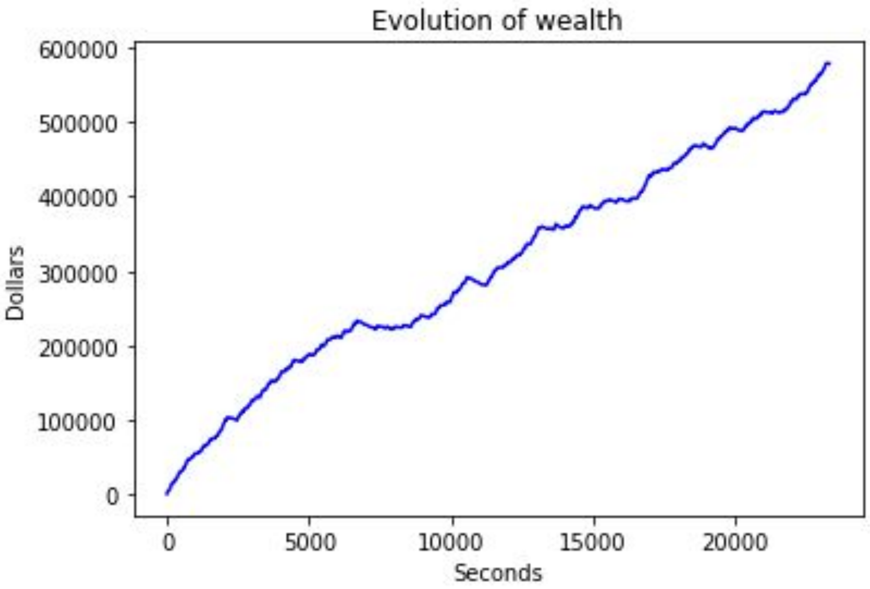
- Execute trades whenever empirical **thresholds are crossed**



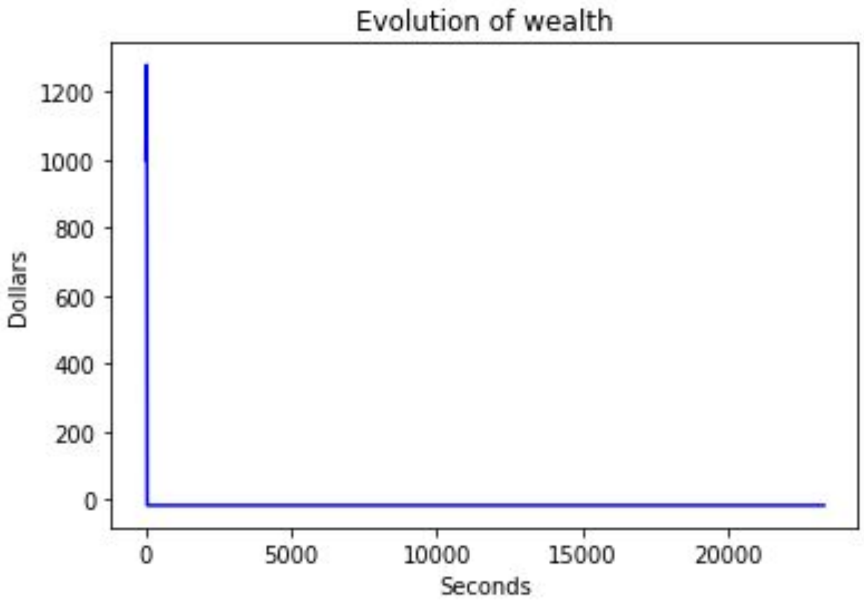
- **Limit the risk** by trading few stocks in a dollar-neutral way and use of stop-loss

Results

AAPL vs CSCO



AAPL vs GOOG



Next steps (1): Stochastic Control

- Now **ad-hoc thresholds,** requiring calibration.
- Idea [2]: think about the **thresholds as stopping times maximizing an expected utility function and find them by solving a HJB equation.**

  Eg: the criteria for exiting and entering a long position at time t observing r could be

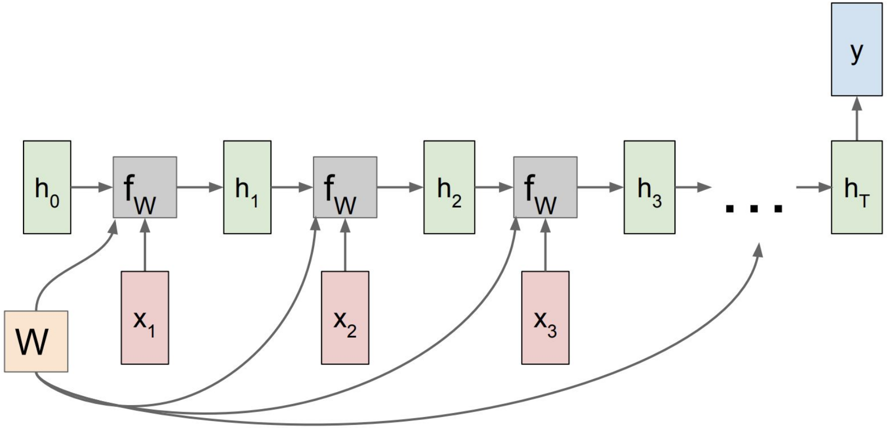
$$H(t,r) = \sup_{\tau} \mathbb{E}_{t,r}[e^{-\rho(\tau-t)}(r_\tau - c)] \quad G(t,r) = \sup_{\tau} \mathbb{E}_{t,r}[e^{-\rho(\tau-t)}(H(\tau,r) - r_\tau - c)]$$

- Next step: **implement this and compare with the naive thresholds.**

Next Steps (2): Predicting Residuals

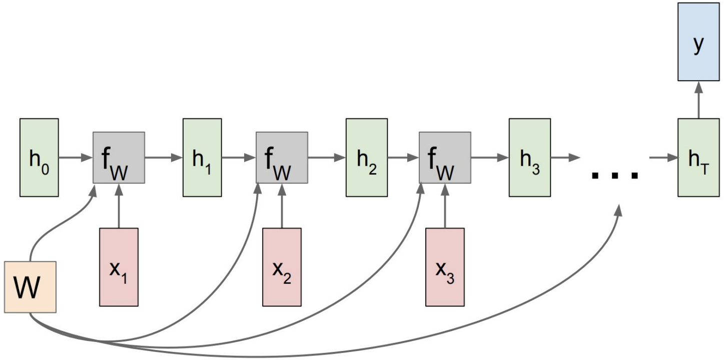
- Use other statistical and machine learning models to predict residuals
  - Other forms of ARIMA models
  - Recurrent Neural Network

$$h_t = f_W(h_{t-1}, x_t)$$



Next Steps (3): Order Book with Deep Learning

- Create feature vectors (proposed by [3]) from the state of the order book at each timestep and formulate strategy using an RNN



Basic Set	Description(i = *level index*, $n = 10$)
$v_1 = \{P_i^{ask},\ V_i^{ask},\ P_i^{bid},\ V_i^{bid}\}_{i=1}^n,$	*price and volume (n levels)*

Time-insensitive Set	Description(i = *level index*)
$v_2 = \{(P_i^{ask} - P_i^{bid}), (P_i^{ask} + P_i^{bid})/2\}_{i=1}^n,$	*bid-ask spreads and mid-prices*
$v_3 = \{P_n^{ask} - P_1^{ask}, P_1^{bid} - P_n^{bid}, \lvert P_{i+1}^{ask} - P_i^{ask}\rvert, \lvert P_{i+1}^{bid} - P_i^{bid}\rvert\}_{i=1}^n,$	*price differences*
$v_4 = \{\frac{1}{n}\sum_{i=1}^n P_i^{ask},\ \frac{1}{n}\sum_{i=1}^n P_i^{bid},\ \frac{1}{n}\sum_{i=1}^n V_i^{ask},\ \frac{1}{n}\sum_{i=1}^n V_i^{bid}\},$	*mean prices and volumes*
$v_5 = \{\sum_{i=1}^n (P_i^{ask} - P_i^{bid}),\ \sum_{i=1}^n (V_i^{ask} - V_i^{bid})\},$	*accumulated differences*

Time-sensitive Set	Description(i = *level index*)
$v_6 = \{dP_i^{ask}/dt,\ dP_i^{bid}/dt,\ dV_i^{ask}/dt,\ dV_i^{bid}/dt\}_{i=1}^n,$	*price and volume derivatives*
$v_7 = \{\lambda_{\Delta t}^{la},\ \lambda_{\Delta t}^{lb},\ \lambda_{\Delta t}^{ma},\ \lambda_{\Delta t}^{mb},\ \lambda_{\Delta t}^{ca},\ \lambda_{\Delta t}^{cb}\ \}$	*average intensity of each type*
$v_8 = \{\mathbf{1}_{\{\lambda_{\Delta t}^{la} > \lambda_{\Delta T}^{la}\}},\ \mathbf{1}_{\{\lambda_{\Delta t}^{lb} > \lambda_{\Delta T}^{lb}\}},\ \mathbf{1}_{\{\lambda_{\Delta t}^{ma} > \lambda_{\Delta T}^{ma}\}},\ \mathbf{1}_{\{\lambda_{\Delta t}^{mb} > \lambda_{\Delta T}^{mb}\}}\},$	*relative intensity indicators*
$v_9 = \{d\lambda^{ma}/dt,\ d\lambda^{lb}/dt,\ d\lambda^{mb}/dt,\ d\lambda^{la}/dt\},$	*accelerations(market/limit)*

Summary

- Statistical Arbitrage using Limit Order Book Data shows initial promising results
- Next Steps:
  - Backtest using Thesys Simulator
  - Test with more pairs
  - Stochastic Control to set thresholds
  - Other methods for predicting residuals
  - RNN strategy

References

[1] Avellaneda, M., & Lee, J. H. (2010). *Statistical Arbitrage in the US Equities Market.* Quantitative Finance, 10(7), p.761-782.

[2] Cartea, A., Jaimungal, S., and Peñalva, J. (2015). *Algorithmic and high frequency trading.* Cambridge University Press, chapter 11.

[3] Kercheval, A. and Zhang, Y. Modeling high-frequency limit order book dynamics with support vector machines. University of Florida, 2013

High Frequency Data Visualization

- [Right] Thesys data visualization of order book for AAPL at 3pm and 3:01 pm on January 2nd, 2015
- [Below]