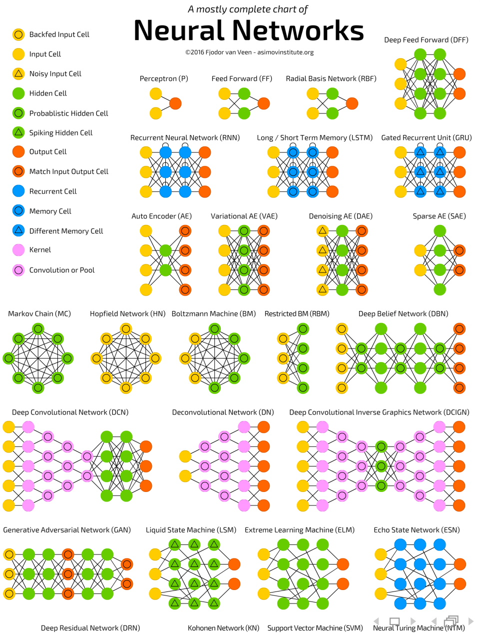
# A High Frequency Trade Execution Model for Supervised Learning

Matthew Dixon[1]

Illinois Institute of Technology

April 20th, 2018
Machine Learning in Finance Workshop
Columbia-Bloomberg


A mostly complete chart of
Neural Networks
©2016 Fjodor van Veen - asimovinstitute.org
Backfed Input Cell
Input Cell
Noisy Input Cell
Hidden Cell
Probablistic Hidden Cell
Spiking Hidden Cell
Output Cell
Match Input Output Cell
Recurrent Cell
Memory Cell
Different Memory Cell
Kernel
Convolution or Pool
Perceptron (P)
Feed Forward (FF)
Radial Basis Network (RBF)
Deep Feed Forward (DFF)
Recurrent Neural Network (RNN)
Long / Short Term Memory (LSTM)
Gated Recurrent Unit (GRU)
Auto Encoder (AE)
Variational AE (VAE)
Denoising AE (DAE)
Sparse AE (SAE)
Markov Chain (MC)
Hopfield Network (HN)
Boltzmann Machine (BM)
Restricted BM (RBM)
Deep Belief Network (DBN)
Deep Convolutional Network (DCN)
Deconvolutional Network (DN)
Deep Convolutional Inverse Graphics Network (DCIGN)
Generative Adversarial Network (GAN)
Liquid State Machine (LSM)
Extreme Learning Machine (ELM)
Echo State Network (ESN)
Deep Residual Network (DRN)
Kohonen Network (KN)
Support Vector Machine (SVM)
Neural Turing Machine (NTM)

# Background on Price Impact Models[2]

- The limit order book is an important source of information for predicting near-term price movements [Parlour, 1998], [Bloomfield, 2005], [Anderson, 2008], [Cao, 2009], [Kearns 2013], [Cont, 2014]
- Regression and machine learning models have been developed to capture linear order flow and price impact relationships [Cont, 2014], [Kearns, 2013], [Kercheval, 2015] and [Sirignano, 2016].
- In practice, the information content of the limit order book does not directly translate to greater economic profits through different high frequency market taking rules [Kozhan, 2012] and [Kearns, 2013]
- How effective are non-linear price impact models for avoiding adverse selection?

# Studying Microstructure is a Data Science Problem

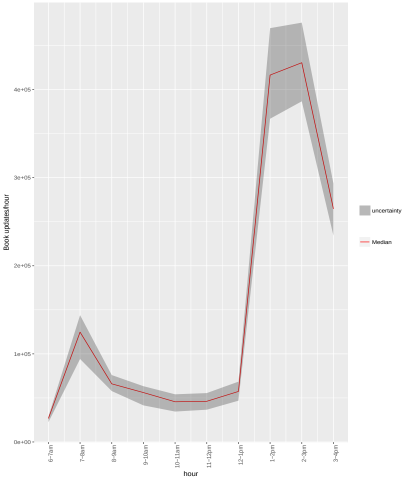



Figure: *The hourly limit order book rates of ESU6 are shown by time of day. A surge of quote adjustment and trading activity is consistently observed between the hours of 7-8am CST and 3-4pm CST.*

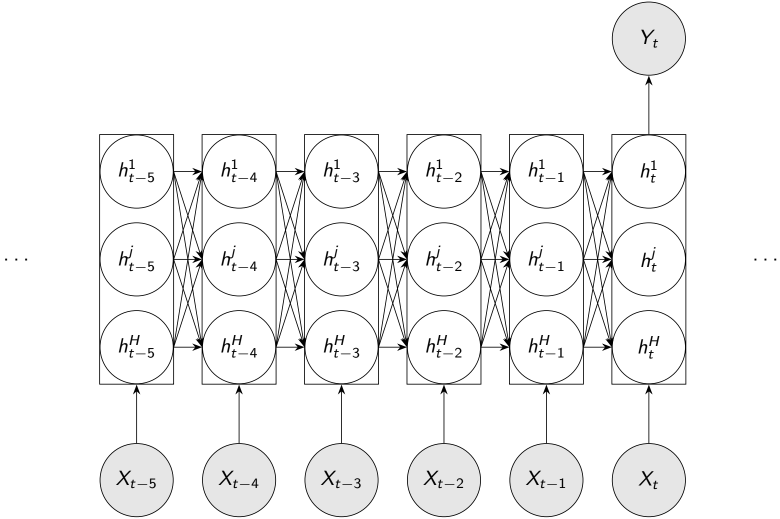
# Recurrent Neural Network Predictors

- Input-output pairs $\mathcal{D} = \{X_t, Y_t\}_{t=1}^{N}$ are auto-correlated observations of $X$ and $Y$ at times $t = 1, \ldots, N$
- Construct a nonlinear times series predictor, $\hat{Y}(\mathcal{X})$, of an output, $Y$, using a high dimensional input matrix of $T$ length sub-sequences $\mathcal{X}$:

$$y = F(\mathcal{X}) \text{ where } \mathcal{X}_t = seq_T(X_t) = (X_{t-T+1}, \ldots, X_t)$$

- $X_{t-j}$ is a $j^{th}$ lagged observation of $X_t$, $X_{t-j} = L^j[X_j]$, for $j = 0, \ldots, T-1$.

# Recurrent Neural Networks



# Recurrent Neural Networks

- For each time step $t = 1, \ldots, T$, a function $F_h$ generates a hidden state $h_t$:

$$h_t = F_h(X_t, h_{t-1}) := \sigma(W_h X_t + U_h h_{t-1} + b_h),\ W_h \in \mathbb{R}^{H\times P}, U_h \in \mathbb{R}^{H\times H}$$

- When the output is continuous, the model output from the final hidden state is given by:

$$Y = F_y(h_T) = W_y h_T + b_y,\ W_y \in \mathbb{R}^{M\times H}$$

- When the output is categorical, the output is given by

$$Y = F_y(h_T) = \text{softmax}(F_y(h_T))$$

- *Goal*: find the weight matrices $W = (W_h, U_h, W_y)$ and biases $b = (b_h, b_y)$.

# Limit Order Book Updates

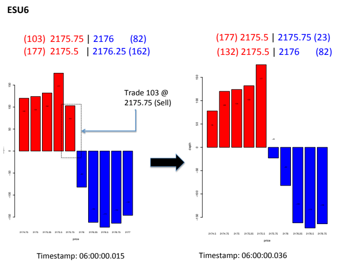



Figure: An exemplary sequence of limit order book updates in the ES futures market (ESU6) is shown before and after the arrival of a sell market order.

| time | $\mathcal{X}_t^1$ | $M_t^s$ | $L_t^{a,1}$ |
|---|---|---|---|
| $t_0$ | $(2175.75, 2176.0, 103, 82)$ | 0 | 0 |
| $t_1$ | $(2175.5, 2176.0, 177, 82)$ | 103 | 0 |
| $t_2$ | $(2175.5, 2175.75, 177, 23)$ | 0 | 23 |

Table: The limit order book of ESU6 before and after the arrival of the sell aggressor.

## Notation

- LOB state: $\mathcal{X}_t := (\mathbf{s}_t^b, \mathbf{s}_t^a, \mathbf{q}_t^b, \mathbf{q}_t^a)$
- Limit orders:
  $\mathbf{L}_t^b := (L_t^{b,1}, \ldots, L_t^{b,n})$,
  $\mathbf{L}_t^a := (L_t^{a,1}, \ldots, L_t^{a,n})$
- Market orders: $M_t^b$ and $M_t^s$ ('aggressors')
- Cancelations:
  $\mathbf{C}_t^b := (C_t^{b,1}, \ldots, C_t^{b,n})$,
  $\mathbf{C}_t^a := (C_t^{a,1}, \ldots, C_t^{a,n})$

# Limit Order Book Updates

| | | $\Omega_\tau^1$ | | | | | | |
|---|---|---|---|---|---|---|---|---|
| time | $\mathcal{X}_0^1$ | $M_\tau^s$ | $M_\tau^b$ | $C_\tau^{b,1}$ | $C_\tau^{a,1}$ | $L_\tau^{b,1}$ | $L_\tau^{a,1}$ | $\mathcal{X}_t^1$ |
| $t_0^-$ | $(2175.75, 2176.0, 102, 82)$ | $\{\}$ | $\{\}$ | $\{\}$ | $\{\}$ | $\{\}$ | $\{\}$ | $(2175.75, 2176.0, 102, 82)$ |
| $t_0$ | $(2175.75, 2176.0, 102, 82)$ | $\{\}$ | $\{\}$ | $\{\}$ | $\{\}$ | $\{1\}$ | $\{\}$ | $(2175.75, 2176.0, 103, 82)$ |
| $t_1$ | $(2175.75, 2176.0, 102, 82)$ | $\{103\}$ | $\{\}$ | $\{\}$ | $\{\}$ | $\{1\}$ | $\{\}$ | $(2175.5, 2176.0, 177, 82)$ |
| $t_2$ | $(2175.75, 2176.0, 102, 82)$ | $\{103\}$ | $\{\}$ | $\{\}$ | $\{\}$ | $\{1\}$ | $\{23\}$ | $(2175.5, 2175.75, 177, 23)$ |

Table: *The state of the top of the top-of-the-book $\mathcal{X}_t^1$ is updated by data $\mathcal{D}_\tau^1$.*

# Fill Ratios

In the event of a sell market order arriving at time $t$, the trade-to-book ratio of a level $j$ bid limit order, $L_0^{b,j}$, placed at time $t_0$ is:

## Trade-to-Book Ratio

$$\mathrm{R}_t(L_0^{b,j}; \mathcal{D}_\tau^{b,j}, \omega) = \frac{M_t^s}{Q_0^{b,j} - \left(\sum_{u \in \mathbf{t}^s} M_u^s + \omega \sum_{i=1}^{j} \sum_{u \in \mathbf{t}^{c,i}} C_u^{b,i} - \sum_{i=1}^{j} \sum_{t \in \mathbf{t}^{b,i}} 1_{\{\phi_{u,u} < \phi_{u,t_0}\}} L_u^{b,i}\right)}$$

- $Q_0^{b,j} := \sum_{i=1}^{j} q_{t_0}^{b,i}$ is the sum of the depths of the queue at time $t_0$ up to the $j^{th}$ bid level
- $\sum_{u \in \mathbf{t}^s} M_u^s$ are the sell market orders arriving at times $\mathbf{t}_s$;
- $\sum_{u \in \mathbf{t}^{c,i}} C_u^{b,i}$ are the level $i$ bid orders cancelled at times $\mathbf{t}^{c,i}$;
- $1_{\{\phi_{u,u} < \phi_{u,t_0}\}}$ is an indicator function returning unity if a subsequent limit order placed at time $u$ has higher queue priority than the time $t_0$ reference limit order; and
- $\omega \in [0,1]$ is an unknown cancellation parameter which denotes the proportion of cancellations of orders with higher queue priority than the reference limit order over the interval $\tau$.

# Example: FIFO market

1. Suppose at time $t_0^-$ the queue depth at the best bid is 50. The largest order has size 20.
2. The reference limit order to buy 50 contracts at the best bid level is received by the exchange at time $t_0$.
3. A market sell order of size 25 arrives in $(t_0, t]$.
4. The best bid for 20 is cancelled in $(t_0, t]$.
5. The queue position of the reference order consequently advances so that there are 5 contracts ahead of it.

If a new sell market order of size 10 arrives at time $t$ then its trade-to-book ratio, with respect to the reference limit order, has the value

$$\mathcal{R}_t(50; \mathcal{D}_\tau^1, 1) = \frac{10}{50 + 50 - (25 + 1 \cdot 20 + 0)} = 2/11 \quad \text{(partial fill)}$$

# Definition of an Execution Strategy

**Strategy**

A strategy is a n-vector function $\mathcal{L} : \mathbb{R}^+ \times \mathbb{Z} \cap [-m, m] \to \mathbb{Z}^n$ of the form $\mathcal{L}_t(\hat{Y}_t)$, where $t$ denotes the time that the trade is placed. Based on the predicted value of $\hat{Y}_t$, the strategy quotes on either the bid and ask at one or more price levels.

# Definition of a Market Making Strategy

**Market Making Strategy**

A market making strategy is the pair $\mathcal{L}_t := (\mathcal{L}_t^a, \mathcal{L}_t^b)$ representing the quoting of a bid and ask at time $t$.

$$\mathcal{L}^a(\hat{Y}_0) \begin{cases} \{0, L\}, & \hat{Y}_0 = 1, \\ \{L, 0\}, & \hat{Y}_0 = 0, \\ \{L, 0\}, & \hat{Y}_0 = -1. \end{cases} \qquad \mathcal{L}^b(\hat{Y}_0) \begin{cases} \{L, 0\}, & \hat{Y}_0 = 1, \\ \{L, 0\}, & \hat{Y}_0 = 0, \\ \{0, L\}, & \hat{Y}_0 = -1. \end{cases}$$

## Spread State

The state of the spread at time $t$ based on the market making strategy $\mathcal{L}_0$ is a function $Z : [-1, 1] \cap \mathbb{Z} \to [-1, 1] \cap \mathbb{Z}$ of the form

$$Z_t(\hat{Y}_0) = \begin{cases} 1, & A := \bigcup_{k=1}^{n}\{\mathcal{R}_t^{k,a} \geq 1\} \cap \bigcup_{k=1}^{n}\{\mathcal{R}_t^{k,b} \geq 1\} \neq \emptyset, \\ -1, & B := \bigcup_{k=1}^{n}\{\mathcal{R}_t^{k,a} < 1\} \cap \bigcup_{k=1}^{n}\{\mathcal{R}_t^{k,b} < 1\} \neq \emptyset, \\ 0, & (A \cup B)^c \neq \emptyset. \end{cases}$$

**Realized P&L**

Let $\Phi : [-1, 1] \cap \mathbb{Z} \to \mathbb{R}$ denote the realized P&L from capturing the spread or adverse selection, after including transactions costs $c$ :

$$\Phi(z) = \begin{cases} \mathcal{L}^a(\hat{Y}_t) \cdot \mathbf{s}_t^a - \mathcal{L}^b(\hat{Y}_t) \cdot \mathbf{s}_t^b - 2Lc, & z = 1, \\ \mathcal{L}^a(\hat{Y}_t) \cdot \mathbf{s}_t^a - \mathcal{L}^b(\hat{Y}_t) \cdot \mathbf{s}_t^b - L(\delta + 2c), & z = 0. \end{cases}$$

The size of the order on each side of the book is assumed to be the same $|\mathcal{L}^a| = |\mathcal{L}^b| = L$, $\delta$ is the spread and $c$ is the transaction cost per contract.

- The realized P&L from capturing the spread or adverse selection

$$\Phi(z) = \begin{cases} L\left(\delta n(\hat{Y}_0) - c'\right), & 1 \\ L\left(\delta(n(\hat{Y}_0) - 1) - c'\right), & 0 \end{cases}$$

- $c'$ is a round-trip transaction cost
- $n : [-1, 1] \cap \mathbb{Z} \to [1, 2] \cap \mathbb{Z}$ with $n(0) = 1$ and $n(-1) = n(1) = 2$.
- The cash flow at time $t$ from the strategy $\mathcal{L}_0$ as a function of the prediction $\hat{Y}_0$ is given by

$$V_t(\hat{Y}_0) = \sum_{z \in \{0,1\}} \mathbf{1}_{\{Z_t(\hat{Y}_0) = z\}} \Phi(z).$$

# Toy Example: Strategies

$MM_1 = (\{1,0\},\{1,0\})$ simply places a one lot bid at the inside market and does not use a prediction.

$MM_2 = (\mathcal{L}_0^a, \mathcal{L}_0^b)$ uses the prediction $\hat{Y}_0$:

$$\mathcal{L}^a(\hat{Y}_0) = \begin{cases} \{0,1\}, & \hat{Y}_0 = 1, \\ \{1,0\}, & \hat{Y}_0 = 0, \\ \{1,0\}, & \hat{Y}_0 = -1. \end{cases} \qquad \mathcal{L}^b(\hat{Y}_0) = \begin{cases} \{1,0\}, & \hat{Y}_0 = 1, \\ \{1,0\}, & \hat{Y}_0 = 0, \\ \{0,1\}, & \hat{Y}_0 = -1. \end{cases}$$

# Toy Example: Parametric Fill Probabilities

| | $y$ | | |
|---|---|---|---|
| | -1 | 0 | 1 |
| $P(Y_0 = y)$ | $a\,(1 - \exp(-\lambda_1(t - t_0)))$ | $\exp(-\lambda_1(t - t_0))$ | $(1 - a)\,(1 - \exp(-\lambda_1(t - t_0)))$ |
| $P(\mathcal{R}_t^{a,1} \geq 1 \mid Y_0 = y)$ | $1 - \exp(-\lambda_2 b(t - t_0))$ | $1 - \exp(-\lambda_2(t - t_0))$ | 1 |
| $P(\mathcal{R}_t^{b,1} \geq 1 \mid Y_0 = y)$ | 1 | $1 - \exp(-\lambda_2(t - t_0))$ | $1 - \exp(-\lambda_2 b(t - t_0))$ |
| $P(\mathcal{R}_t^{a,2} \geq 1 \mid Y_0 = y)$ | 0 | 0 | $1 - \exp(-\lambda_2(t - t_0))$ |
| $P(\mathcal{R}_t^{b,2} \geq 1 \mid Y_0 = y)$ | $1 - \exp(-\lambda_2(t - t_0))$ | 0 | 0 |

Table: *The conditional fill probabilities and the mid-price change probabilities. $\lambda_1$ is the decay parameter for the mid-price change probabilities. $\lambda_2$ is the decay parameter for the fill probabilities. The parameter $a > 0$ is used to control the asymmetry in the mid-price movements, where $a = 0.5$ is the symmetric case. The parameter $b$ is used to adjust the relative fill rates by level. It is assumed that quotes at outer prices levels are more difficult to be filled and hence $0 < b < 1$.*

# Toy Example with parametric fill probabilities

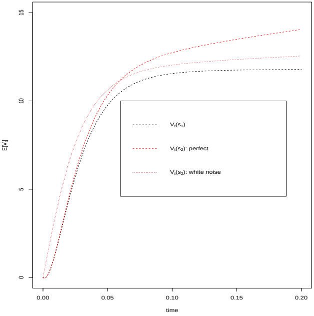



Figure: *The expected realized P&L of strategy $MM_2$ compared with strategy $MM_1$ for the following configuration $a = 0.5$, $b = 0.5$, $\lambda_1 = 1$. The spread $\delta = \$12.5$ and the round-trip transaction cost is $c = \$0.7$. Larger values of $\lambda_2$ correspond to increased growth rates of the fill probabilities. The figure also compares the effect of a white noise predictor (dotted) with a perfect predictor (dashed). The equilibrium level of the white noise predictor is lower than the perfect predictor but more profitable than strategy $MM_1$.*

**Confusion Matrix**

The confusion matrix is a function $C : \mathbb{R}^+ \rightarrow \mathbb{R}_+^{M \times M}$, $M = 2m + 1$ of the form

$$C_{ij}(t) := P(\hat{Y}_t = y_j \mid Y_t = y_i), \forall i, j \in \{1, \ldots, M\} \times \{1, \ldots, M\},$$

for a predicted state $\hat{Y}_t \in \mathbf{y} :=\in [-m, m] \cap \mathbb{Z}$ and a true state $Y_t \in \mathbf{y}$.

**Trade Information Matrix**

The trade information matrix is a function $T : \mathbb{R}^+ \to \mathbb{R}^{M\times M}$ given by

$$T_{ij}(t; \Omega_0^k, \mathcal{D}_\tau^k, \omega) := P(Y_0 = y_i)\mathbb{E}[V_t(\hat{Y}_0 = y_j)|Y_0 = y_i, \hat{Y}_0 = y_j]$$

which uses the triple $\Omega_0^k := (\mathcal{L}_0^k, \hat{Y}_0, Y_0)$, consisting of predictions $\hat{Y}_0$, the true state $Y_t$ and the $k^{th}$ level offer placed by a strategy $\mathcal{L}_0$ at time $t_0$, in addition to the order book events $\mathcal{D}_\tau^k$.

**Expected Cash Flow**

The expected cash flow from the triple $\Omega_0 := (\mathcal{L}_0, \hat{Y}_0, Y_0)$ is

$$\mathbb{E}[V_t] = tr(C(t_0)T'(t))$$

where $T'$ denotes the transpose of $T$.

| | $\hat{Y}_0 = y$ | | |
|---|---|---|---|
| | -1 | 0 | 1 |
| -1 | 1.062 | 1.062 | 1.062 |
| 0 | 9.660 | 9.660 | 9.660 |
| 1 | 1.062 | 1.062 | 1.062 |

| | $\hat{Y}_0 = y$ | | |
|---|---|---|---|
| | -1 | 0 | 1 |
| -1 | 2.195, | 1.062 | 1.0695 |
| 0 | 9.660 | 9.660 | 9.660 |
| 1 | 1.0695 | 1.062 | 2.195 |

Table: *Trade information matrices for the $MM_1$ (left) and $MM_2$ (right) strategies evaluated at elapsed time $t = 0.2$.*

# Spatio-Temporal Model

- The response is

$$Y_t = \Delta p^t_{t+h} \tag{1}$$

- $\Delta p^t_{t+h}$ is the forecast of discrete mid-price changes from time $t$ to $t+h$, given measurement of the predictors up to time $t$.
- The predictors are embedded

$$x = x^t = \text{vec}\begin{pmatrix} x_{1,t-k} & \cdots & x_{1,t} \\ \vdots & & \vdots \\ x_{n,t-k} & \cdots & x_{n,t} \end{pmatrix} \tag{2}$$

- $n$ is the number of quoted price levels, $k$ is the number of lagged observations, and $x_{i,t} \in [0,1]$ is the relative depth, representing liquidity imbalance, at quote level $i$:

$$x_{i,t} = \frac{q_t^{a,i}}{q_t^{a,i} + q_t^{b,i}}. \tag{3}$$

# Spatial-Temporal Representation

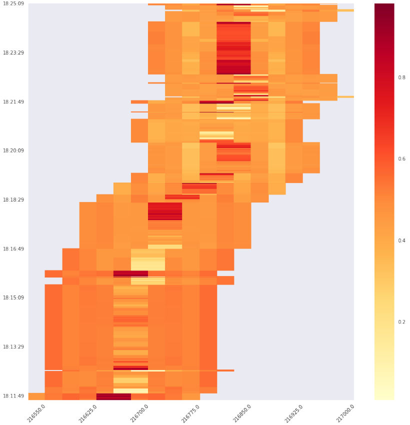



Figure: A space-time diagram showing the limit order book. The contemporaneous depths imbalances at each price level, $x_{i,t}$, are represented by the color scale: red denotes a high value of the depth imbalance and yellow the converse. The limit order book are observed to polarize prior to a price movement.

# Historical Data

- At any point in time, the amount of liquidity in the market can be characterized by the cross-section of book depths.
- We build a mid-price forecasting model based on the cross-section of book depths.

| Timestamp | $s_t^{b,1}$ | $s_t^{b,2}$ | ... | $q_t^{b,1}$ | $q_t^{b,2}$ | ... | $s_t^{a,1}$ | $s_t^{a,2}$ | ... | $q_t^{a,1}$ | $q_t^{a,2}$ | ... | $Y_t$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 06:00:00.015 | 2175.75 | 2175.5 | ... | 103 | 177 | ... | 2176 | 2176.25 | ... | 82 | 162 | ... | -1 |
| 06:00:00.036 | 2175.5 | 2175.25 | ... | 177 | 132 | ... | 2175.75 | 2176 | ... | 23 | 82 | ... | 0 |

Table: *The limit order book of ESU6 before and after the arrival of the sell aggressor. Here, the response is the mid-price movement over the subsequent interval, in units of ticks. $s_t^{b,i}$ and $q_t^{b,i}$ denote the level i quoted bid price and depth of the limit order book at time t. $s_t^{a,i}$ and $q_t^{a,i}$ denote the corresponding level i quoted ask price and depth.*

# The Price Impact of Order Flow

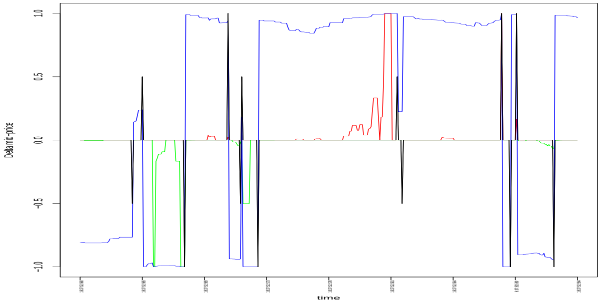
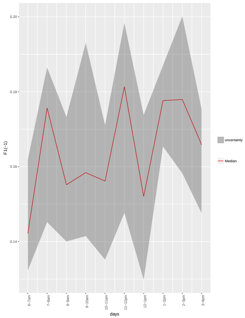



Figure: The black line represents the observed change in mid-price over a 34 milli-second period from 16:37:52.560 to 16:37:52.594. The liquidity imbalance (blue), scaled here to the $[-1, 1]$ interval, although useful in predicting the direction of the next occurring price change, is generally a poor choice for predicting when the price change will occur. The order flow is a better predictor of next-event price movement, although is difficult to interpret when either of the buy (red) and sell order flows (green) are small.

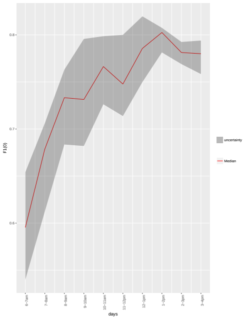
# Predictive Performance Comparisons

| Features | Method | $\hat{Y}=-1$ precision | $\hat{Y}=-1$ recall | $\hat{Y}=-1$ f1 | $\hat{Y}=0$ precision | $\hat{Y}=0$ recall | $\hat{Y}=0$ f1 | $\hat{Y}=1$ precision | $\hat{Y}=1$ recall | $\hat{Y}=1$ f1 |
|---|---|---|---|---|---|---|---|---|---|---|
| Liquidity Imbalance | Logistic | 0.010 | 0.603 | 0.019 | 0.995 | 0.620 | 0.764 | 0.013 | 0.588 | 0.025 |
| | Kalman Filter | 0.051 | 0.540 | 0.093 | 0.998 | 0.682 | 0.810 | 0.055 | 0.557 | 0.100 |
| | RNN | 0.037 | 0.636 | 0.070 | 0.996 | 0.673 | 0.803 | 0.040 | 0.613 | 0.075 |
| Order Flow | Logistic | 0.042 | 0.711 | 0.079 | 0.991 | 0.590 | 0.740 | 0.047 | 0.688 | 0.088 |
| | Kalman Filter | 0.068 | 0.594 | 0.122 | 0.996 | 0.615 | 0.751 | 0.071 | 0.661 | 0.128 |
| | RNN | 0.064 | 0.739 | 0.118 | 0.995 | 0.701 | 0.823 | 0.066 | 0.728 | 0.121 |
| Spatio-temporal | Elastic Net | 0.063 | 0.754 | 0.116 | 0.986 | 0.483 | 0.649 | 0.058 | 0.815 | 0.108 |
| | RNN | 0.084 | 0.788 | 0.153 | 0.999 | 0.729 | 0.843 | 0.075 | 0.818 | 0.137 |
| | FFWD NN | 0.066 | 0.758 | 0.121 | 0.999 | 0.657 | 0.795 | 0.065 | 0.796 | 0.120 |
| | White Noise | 0.004 | 0.333 | 0.007 | 0.993 | 0.333 | 0.499 | 0.003 | 0.333 | 0.007 |

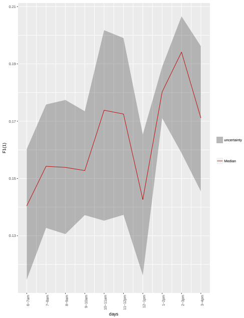
# Intra-day Predictive Performance




(a) F1 score of $\hat{Y} = -1$



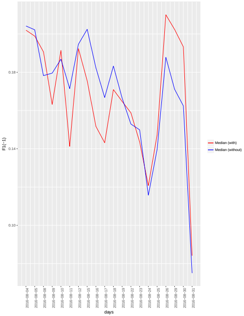

(b) F1 score of $\hat{Y} = 0$



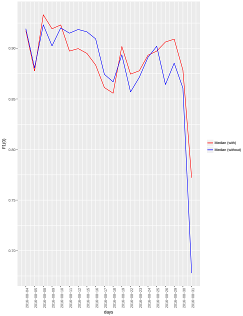

(c) F1 score of $\hat{Y} = 1$.

Figure: *The intra-day F1 scores are shown for (left) downward, (middle) neutral, or (right) upward next price movement prediction.*

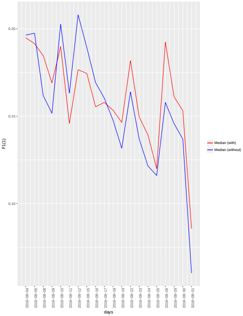
# Daily Retraining




(a) F1 score of $\hat{Y} = -1$




(b) F1 score of $\hat{Y} = 0$




(c) F1 score of $\hat{Y} = 1$.

Figure: *The F1 scores over the calendar month, with (red) and without (blue) daily retraining of the RNN, are shown for (left) downward, (middle) neutral, or (right) upward next price movement prediction.*

# ROC

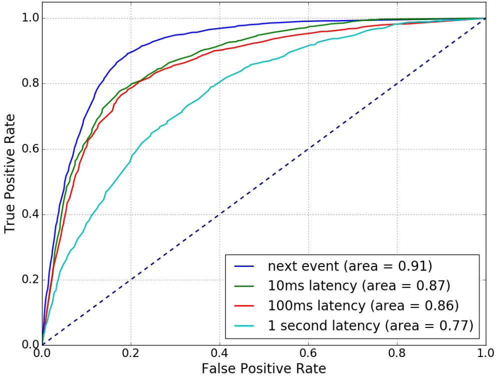



Figure: *The Receiver Operator Characteristic (ROC) curves of a binary RNN classifier over varying prediction horizons. In practice, the prediction horizon should be chosen to adequately account for latency between the trade execution platform and the exchange.*

# Strategy Comparison: Cumulative P&L and Positions

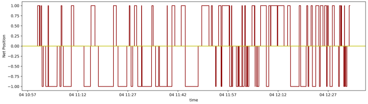
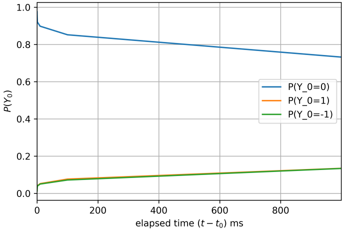




Figure: *(top) The top-of-the book prices together with the cumulative P&L of the market making strategies are shown over the period 11am to 12:30pm (CST) on August the 4th, 2016. (bottom) The position in ESU6 resulting from the MM2 strategy with a unit position limit.*

# Empirical Probabilities of Flips and Fills



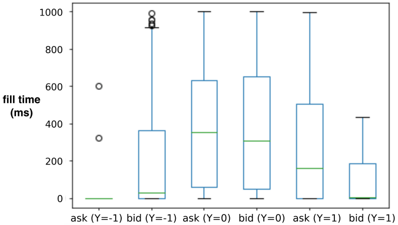




Figure: *(left) The empirical probabilities of price-flips are shown against prediction horizon h. The boxplot (right) shows the empirical distribution of fill times on August 4th 2016. Empirical fill times are separately estimated for best bid and ask quotes conditioned on* $Y_0$.

# Comparison of Empirical Probabilities of Flips and Fills

| MM1 | $y$ | | |
|---|---|---|---|
| | -1 | 0 | 1 |
| $P(Y_0 = y)$ | 0.134 | 0.732 | 0.134 |
| $P(R_t^{a,0} \geq 1 \| Y_0 = y)$ | 0.074 | 0.67 | 0.581 |
| $P(R_t^{b,0} \geq 1 \| Y_0 = y)$ | 0.563 | 0.615 | 0.107 |
| $P(Z = 1 \| Y_0 = y)$ | 0.042 | 0.412 | 0.062 |
| $P(Z = 0 \| Y_0 = y)$ | 0.554 | 0.461 | 0.563 |

| MM2 | $y$ | | |
|---|---|---|---|
| | -1 | 0 | 1 |
| $P(Y_0 = y)$ | 0.134 | 0.732 | 0.134 |
| $P(R_t^{a,1} \geq 1 \| Y_0 = y)$ | 0.007 | 0.511 | 0 |
| $P(R_t^{a,2} \geq 1 \| Y_0 = y)$ | 0 | 0 | 0.421 |
| $P(R_t^{b,1} \geq 1 \| Y_0 = y)$ | 0 | 0.504 | 0.011 |
| $P(R_t^{b,2} \geq 1 \| Y_0 = y)$ | 0.403 | 0 | 0 |

Table: The estimated empirical price movement probabilities, quote fill probabilities and spread fill probabilities conditioned on the movement of the true state over a forecasting horizon of $t = h = 1s$. Each column shows the corresponding conditional probabilities for each value of $Y_0$.

# MM2 Trade Information Matrices

| Level 1 | $\hat{Y}_0 = y$ | | |
|---|---|---|---|
| | -1 | 0 | 1 |
| -1 | 0.014 | 0.014 | 0.014 |
| 0 | 3.324 | 3.324 | 3.324 |
| 1 | 0.045 | 0.045 | 0.045 |

| Level 2 | $\hat{Y}_0 = y$ | | |
|---|---|---|---|
| | -1 | 0 | 1 |
| -1 | 0.604 | 0 | 0 |
| 0 | 2.457 | 2.225 | 4.059 |
| 1 | 0 | 0 | 0.640 |

Table: *The trade information matrix for all quotes placed at the inside market (top) and at the next price level away from the inside market (bottom).*

# MM2 Spread Fill Probabilities

| $P(Z_t = 1 \| Y_0 = y_i, \hat{Y}_0 = \hat{y}_j)$ | | | |
|---|---|---|---|
| | $\hat{y}$ | | |
| $y$ | -1 | 0 | 1 |
| -1 | 0.003 | 0 | 0 |
| 0 | 0 | 0.258 | 0 |
| 1 | 0 | 0 | 0.005 |

| $P(Z_t = 0 \| Y_0 = y_i, \hat{Y}_0 = \hat{y}_j)$ | | | |
|---|---|---|---|
| | $y$ | | |
| $y$ | -1 | 0 | 1 |
| -1 | 0.404, | 0.003 | 0 |
| 0 | 0.305 | 0.5 | 0.504 |
| 1 | 0 | 0.011 | 0.423 |

Table: *The probability that the spread is filled (left) and the probability of adverse selection (right) using the MM2 strategy conditioned on the true movement $Y_0$ and the prediction $\hat{Y}_0$.*

# Using Predictions for Market Making

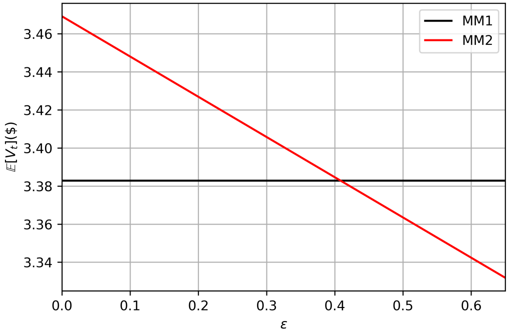



Figure: *This figure compares the expected P&L of the two market market strategies as a function of error $\epsilon$ in the confusion matrix. It is observed that the expected realized profit from MM2 (red) linearly decays with $\epsilon$, to the extent that it can become less profitable than the baseline strategy MM1 (black).*

# Conclusion

- Use a high frequency trade execution model to backtest the economic impact of supervised learners
- The model requires
  - (i) estimation of the fill probabilities conditioned on price movement
  - (ii) a trade information matrix to attribute the expected profit and loss of tick level predictive classifiers
- Identify the amount of error tolerance for a prediction based market making strategy versus a non-predictive strategy
- We find that RNNs outperform out-perform other ML and time series filtering techniques but are sensitive to the prediction lead-time

# Overfitting Assessment

| | In Sample | | |
|---|---|---|---|
| | -1 | 0 | 1 |
| -1 | 0.93 (0.03) | 0.03 (0.01) | 0.04 (0.03) |
| 0 | 0.10 (0.02) | 0.81 (0.03) | 0.09 (0.02) |
| 1 | 0.05 (0.02) | 0.04 (0.02) | 0.91 (0.02) |
| | Out-of-Sample | | |
| | -1 | 0 | 1 |
| -1 | 0.84 (0.06) | 0.06 (0.05) | 0.10 (0.04) |
| 0 | 0.10 (0.03) | 0.78 (0.06) | 0.12 (0.03) |
| 1 | 0.11 (0.04) | 0.06 (0.04) | 0.83 (0.04) |

Table: *The confusion matrices for the (left) in-sample performance of the RNN classifier with the (right) out-of-sample performance of the RNN classifier using a prediction horizon $h = 1ms$. Each confusion matrix is formed by averaging the daily confusion matrices over the month of August. The parentheses show the standard deviation.*

# Prediction Example

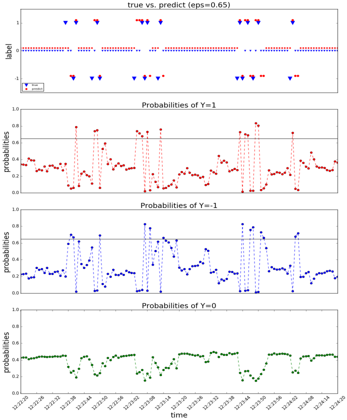



Figure: The (top) comparison of the observed ESU6 mid-price movements with the deep learner forecasted price movements over one milli-second intervals between 12:22:20 and 12:24:20 CST. The bottom three panels show the corresponding probabilities of predicting each class. A probability threshold of 0.65 for the up-tick or down-tick classification is chosen here for illustrative purposes.