Distance-Based High-Frequency Trading

Travis Felker
Quantica Trading
Kitchener, Canada
travis@quanticatrading.com

Vadim Mazalov   Stephen M. Watt
University of Western Ontario
London, Canada
Stephen.Watt@uwo.ca

PCA 2015: 17 April 2015
Euler Insternational Mathematics Institute , St Petersburg, Russia

The Problem

- Accurate and efficient short term prediction of *one* change in the price of an asset.

- A number of strategies developed over time, from simple and fast to sophisticated models. These include methods based on time series analysis, support vector machines, hidden Markov models, nearest neighbor classifiers, etc.

The Approach

- Observations as points in a multi-dimensional space of numerical technical indicators.
- **Clusters of points** representing **price movements up** and **down.**
- Points weighted by distance to cluster centroids.
- Predictions made when classification confidence is high enough.
- Groups of new points added as events occur, updating clusters.

The Stock Market

- US equity markets, e.g. NYSE, NASDAQ, AMEX, offer several venues where the same product may be bought and sold.
  (Alternative Trading Systems, ATS)
- Having multiple venues offer challenges and opportunities.
- Asynchronous purchase and sale at National Best Bid/Best Offer (NBBO) price.
- Offer routing may lead to market lock in composite exchange.

High Frequency Trading

- Consistent trading activity in a brief time span.
- Typically performed algorithmically by computers close to exchanges.
- Affected by factors including network infrastructure and latency, clearing fee structure, software optimization.

Technical Indicators

- Choose a set of numerical indicators.
- Too few => loss of precision.
  Too many => too sparse (since high dimension)
- We examine only quotes ate current best bid and ask, independently of complementary and supplementary securities.
- Divide into those common to all exchanges and those particular to one exchange.

Technical Indicators

- $ROC_i = w \frac{\frac{b_i}{o_i} - \frac{b_{i-1}}{o_{i-1}}}{t_i - t_{i-1}} + (1-w) ROC_{i-1}$

  b_i bid depth o_i offer depth
- n_{cb} (n_{co}) number of times an exchange/ATS locked the market on the bid (offer)
- n_{lb} (n_{lo}) number of times an exchange/ATS left the NBBO on the bid (offer)
- s_b (s_o) sum of weights of venues with bid (offer) equal to the NBBO bid (offer)

Form a $5E + 2$ dimensional space, with E exchanges.

Outliers and Normalization

- All market events used in the computation of features, though only some give price changes.
- Periodically prune values farther than 3 σ from cluster centers.
- To make features comparable, they are normalized

$$x_i' = \frac{x_i - x_i^{\min}}{x_i^{\max} - x_i^{\min}}$$

The Classifier

- In high-frequency setting the classifier should be efficient.
- We compute the feature-weighted distance from a test sample to the centroid of a cluster, since this is one of the least expensive techniques in artificial intelligence.
- Training is performed by collecting points and finding their centroid.
- Classification is done by finding the squared weighted Euclidean distance to the centroid.

Classification vs Prediction

- Classification happens with each quote received – a feature vector is formed and the distances to centroids are evaluated.
- A prediction is made only if the distances between the sample and the centroids satisfy certain criteria, i.e. if the feature point is relatively close to one of the two centroids.

Benefits of the Classifier

- ***Adaptability*****:** If values of some of the indicators change, the centroid will slowly move in the direction of change. The coordinates of a centroid can be updated in constant time with each new point or a group of points.
- ***Transparency*****:** The method facilitates control of the impact that certain indicators or weights have on the distance. The values of features and their weights can be easily analyzed by human experts to validate the model.
- ***Presence of a confidence measure*****:** The classification confidence can be derived intuitively from the distances and their ratios to regulate the accuracy and the number of predictions

Complexity

- Indicators are computed in *constant time* on each quote.
- Outlier test, given that the centroid and the standard deviation of the cluster have already been computed, takes $O(D)$ time, D being the dimension of the feature space.
- Normalization, computation of feature weights, and classification are done in $O(D)$ time.

Experimental Setting

- We collected MSFT (Microsoft) securities, using quotes from 10 leading exchanges/ ATSs.
- The recorded events: change in bid/offer prices and bid/offer depth.
- We recorded several days in December, 2011 with the total of 9,389,993 quotes and 4,658 price changes.

Experimental Setting, cont.

- Training was performed until both clusters had at least 10 points. The value of the weight in computation of the
- ROC was taken as 0.6.
- After 5 changes in price, parameters of a cluster were recomputed.

On-Change Accuracy

- The on-change distance was counted as correct if the distance to the centroid of the cluster in the direction of the price change was smaller than the distance to the other cluster.

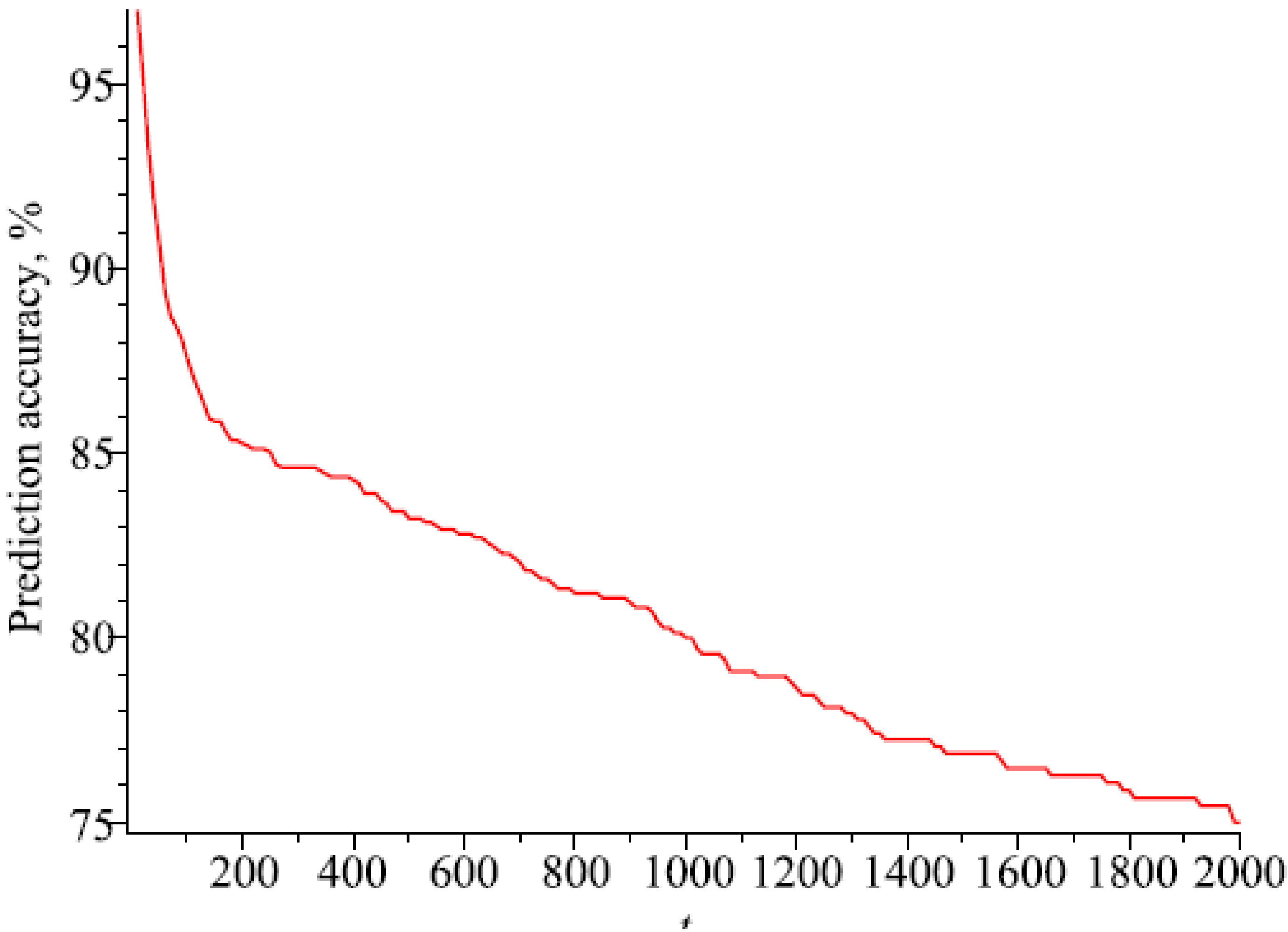
Prediction Accuracy

- If the prediction was in the direction of the price change, and the interval between a prediction and the actual change was greater than *t*, the count of correct predictions was incremented. If the interval was less than t, the count was not changed. Otherwise, the count of wrong predictions was incremented.
- This measure aimed to simulate real-life trading, when execution of a transaction takes a certain amount of time, depending on infrastructure.

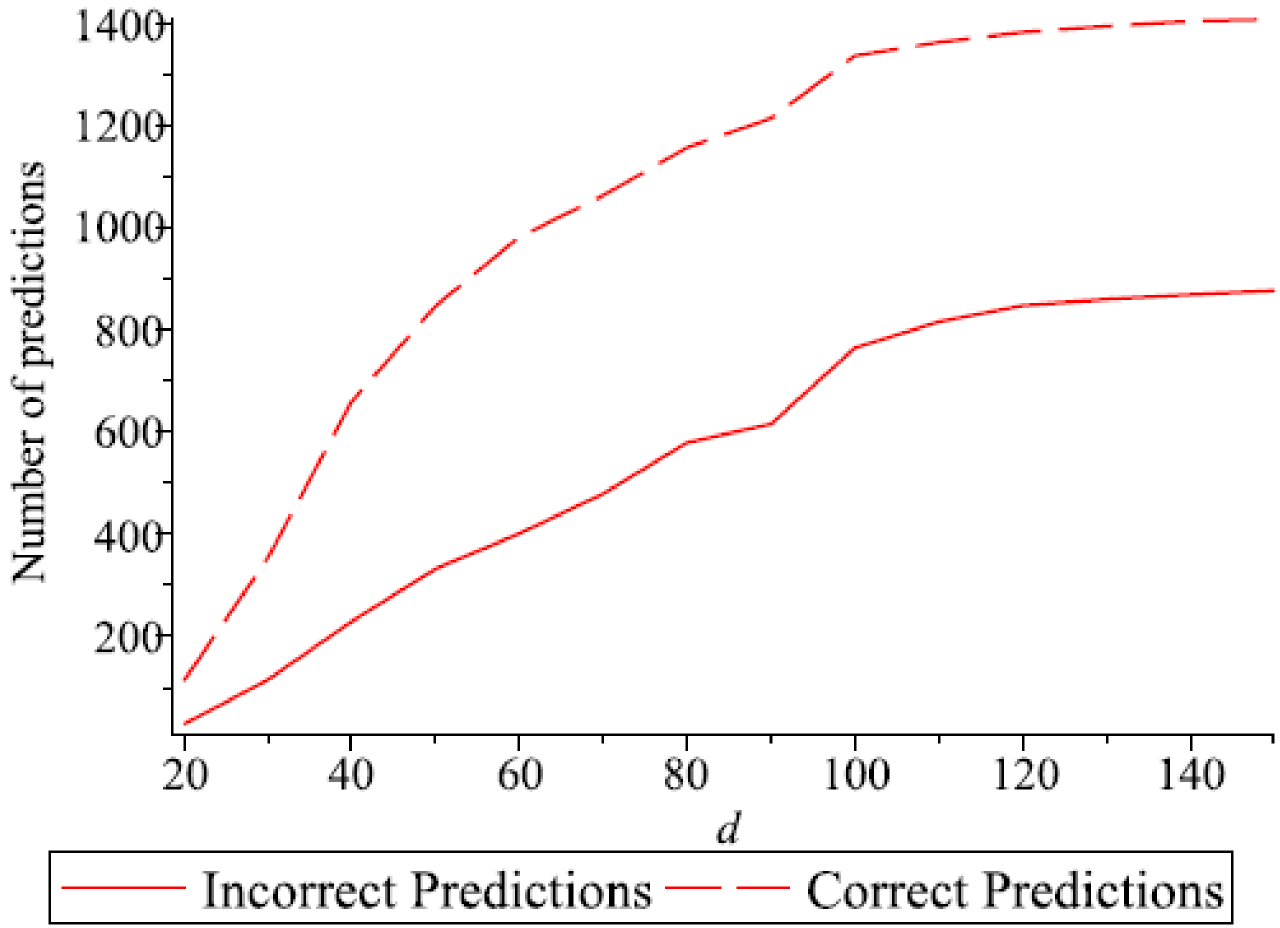
Experimental Results

- The on-change accuracy of the model on the recorded data was 96.25%.

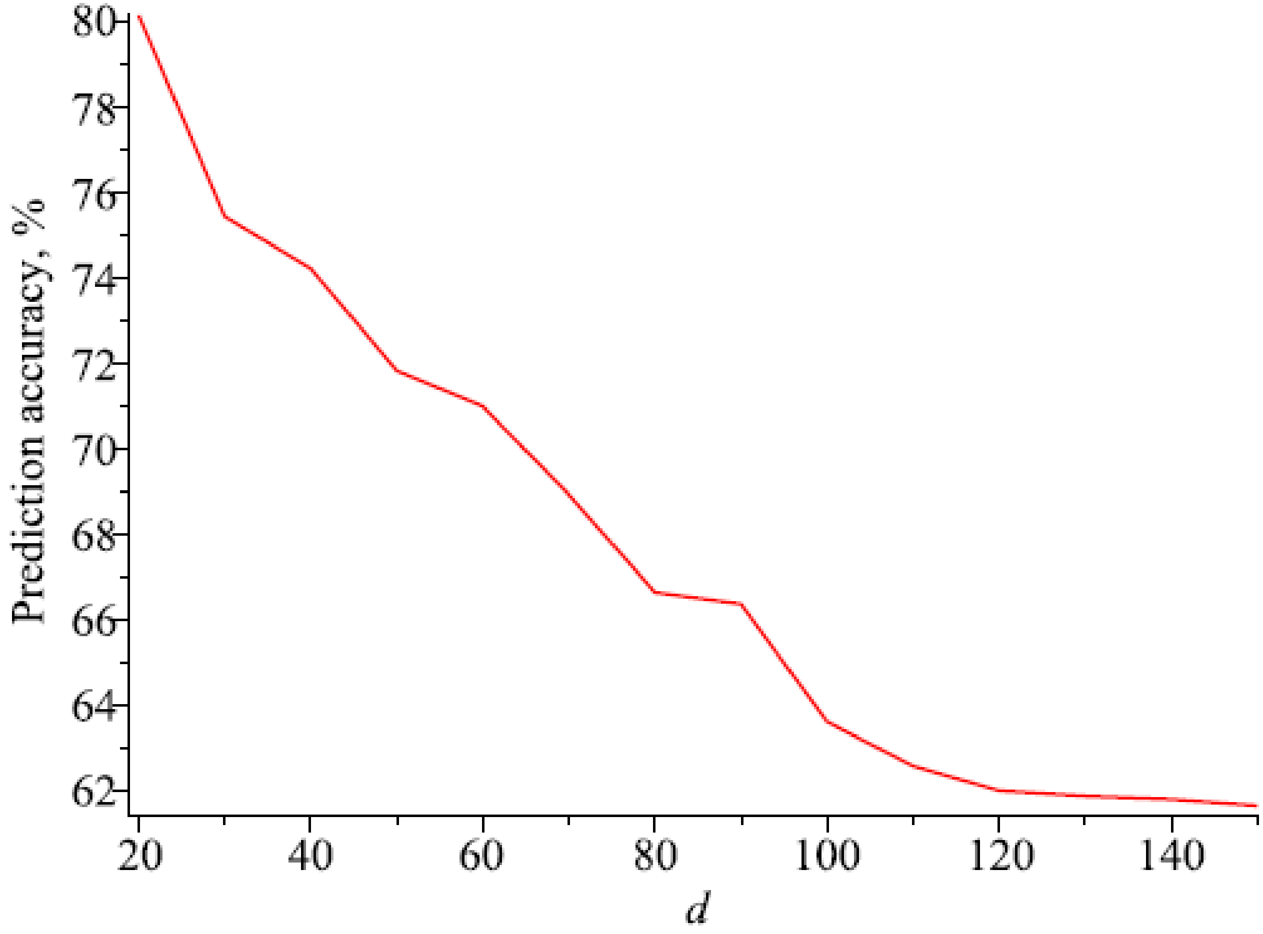
The prediction accuracy (wrt *t*)



Number of predictions made

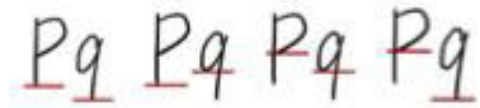


The prediction accuracy (wrt *d*)

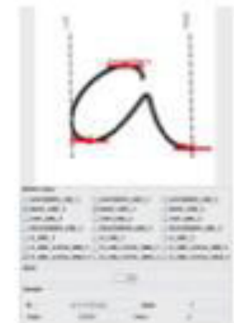


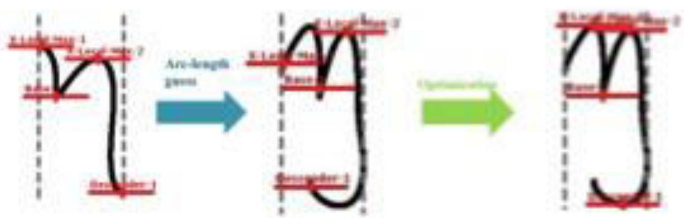
Comparison to HWR

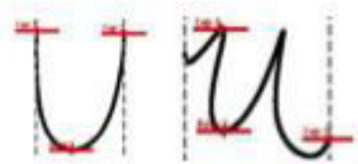
$e^x = \int e^x dx = \sum_{i=0}^{\infty} \frac{x^i}{i!}$



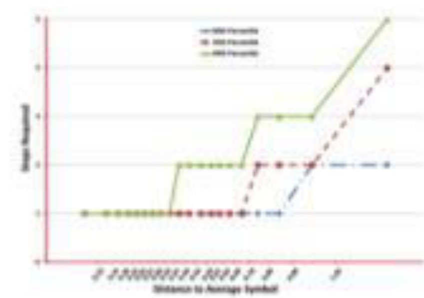
$$f(s) \approx \sum_{i=0}^{d} c_i P_i(s)$$



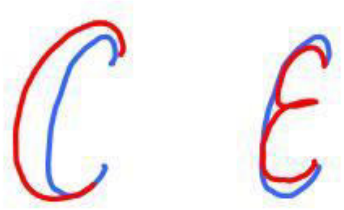




$$C(t) = (1-t)\bar{C} + tC_{targ}$$



HWR: Shape vs Variation



- The corners are not in the right places.
- Work in a jet space to force coords & derivatives close.
- Use a Legendre-Sobolev inner product

$$\langle f, g\rangle = \int_a^b f(t)g(t)dt + \mu_1 \int_a^b f'(t)g'(t)dt + \mu_2 \int_a^b f''(t)g''(t)dt + \cdots$$

- 1st jet space ⇒ set $\mu_i = 0$ for $i > 1$.
  - Choose μ_1 experimentally to maximize reco rate.
  - Can be also done on-line.

    [Golubitsky + SMW 2008, 2009]

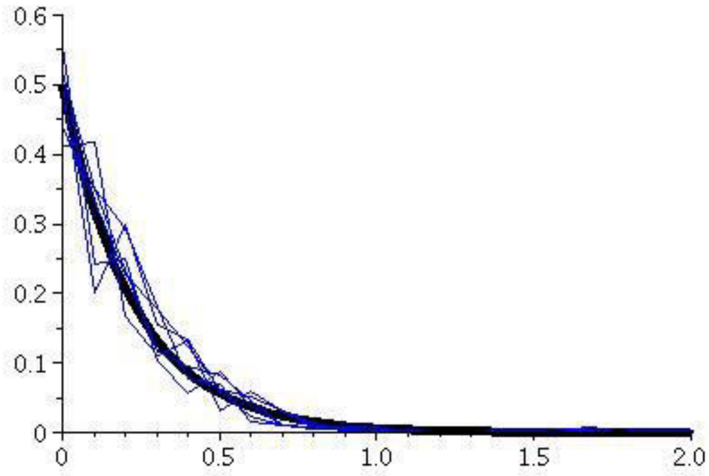
HWR: Distance Between Curves

$$\bar{x}(t) = x(t) + \xi(t) \qquad \xi(t) = \sum_{i=0}^{\infty} \xi_i \phi_i(t), \qquad \phi_i \text{ ortho on } [a, b] \text{ with } w(t) = 1.$$

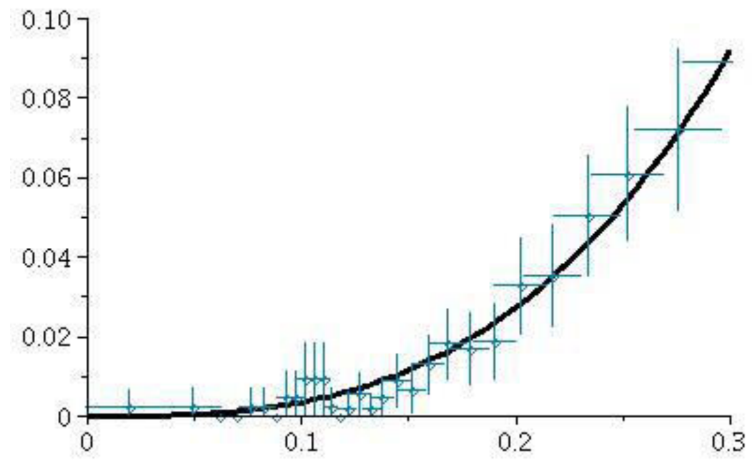
$$\bar{y}(t) = y(t) + \eta(t) \qquad \eta(t) = \sum_{i=0}^{\infty} \eta_i \phi_i(t)$$

$$\begin{aligned}
\rho^2(C, \bar{C}) &= \int_a^b \left[\left(x(t) - \bar{x}(t) \right)^2 + \left(y(t) - \bar{y}(t) \right)^2 \right] dt \\
&= \int_a^b [\xi(t)^2 + \eta(t)^2] dt \\
&\approx \int_a^b \left[\sum_{i=0}^{d} \xi_i^2 \phi_i^2(t) + \text{cross terms} + \sum_{i=0}^{d} \eta_i^2 \phi_i^2(t) + \text{cross terms} \right] dt \\
&= \sum_{i=0}^{d} \xi_i^2 + \sum_{i=0}^{d} \eta_i^2
\end{aligned}$$

HWR: Error Rates as Fn of Distance



SVM



Convex Hull

- Error rate as fn of distance gives confidence measure for classifiers [MKM – Golubitsky + SMW 2009]

Conclusion

- The model was determined to perform reasonably well, even with the simple financial indicators.
- To improve the model, indicators from supplementary and complementary products can be considered.
- Other directions for improvement: study of dynamics of distances during a flip and subclustering training sets.