Financial Time-Series Classification with Wavelets

Raphael Haddad
Associate Professor Philip Leong
School of Electrical and Information Engineering
FACULTY OF ENGINEERING & INFORMATION TECHNOLOGIES
THE UNIVERSITY OF SYDNEY

Introduction

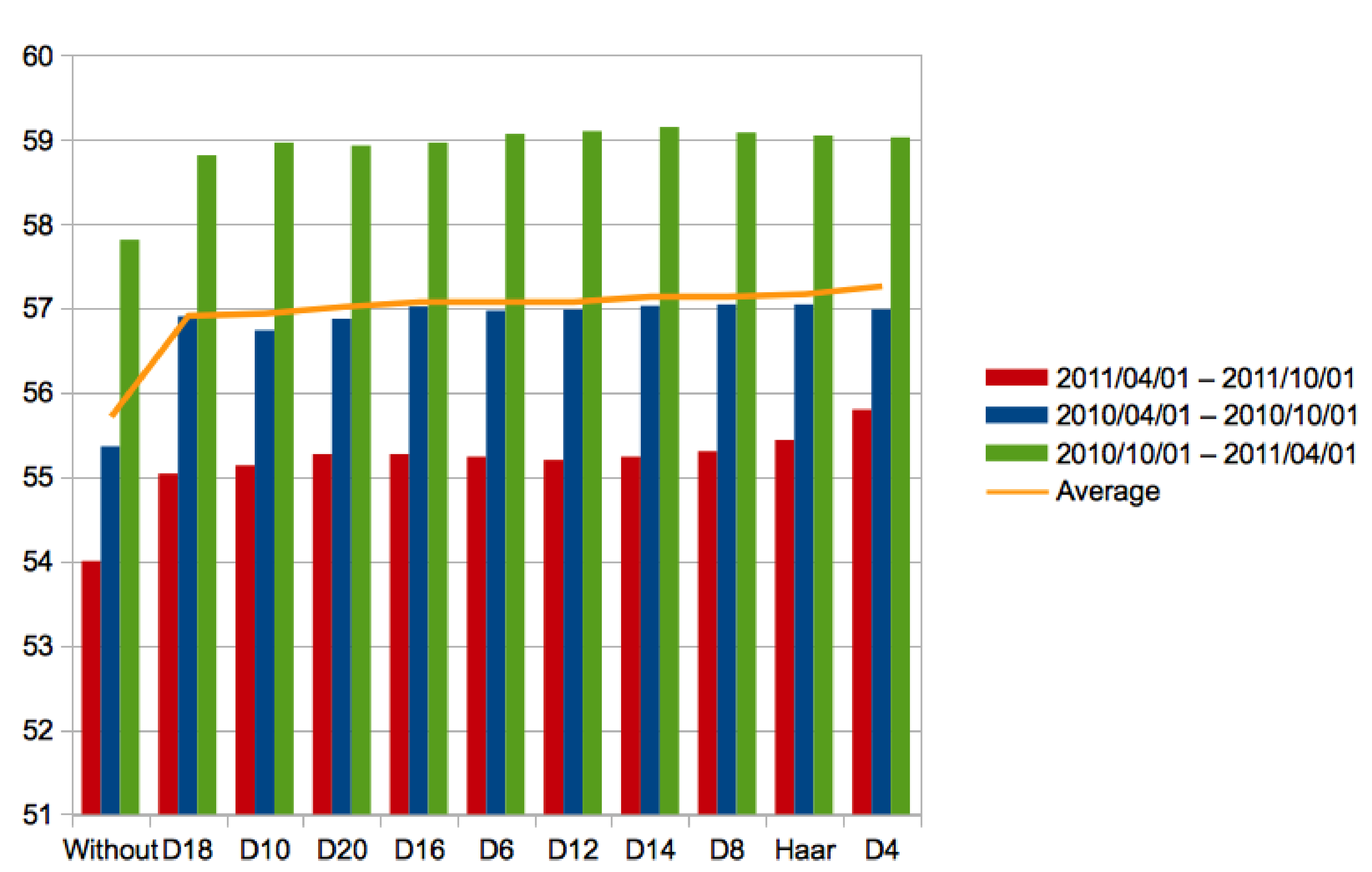
- Currency held in external country.
- Fluctuations in exchange rates.
  - Political.
  - Social.
  - Environmental.
- Small fluctuations can potentially cause large quantities of financial loss.

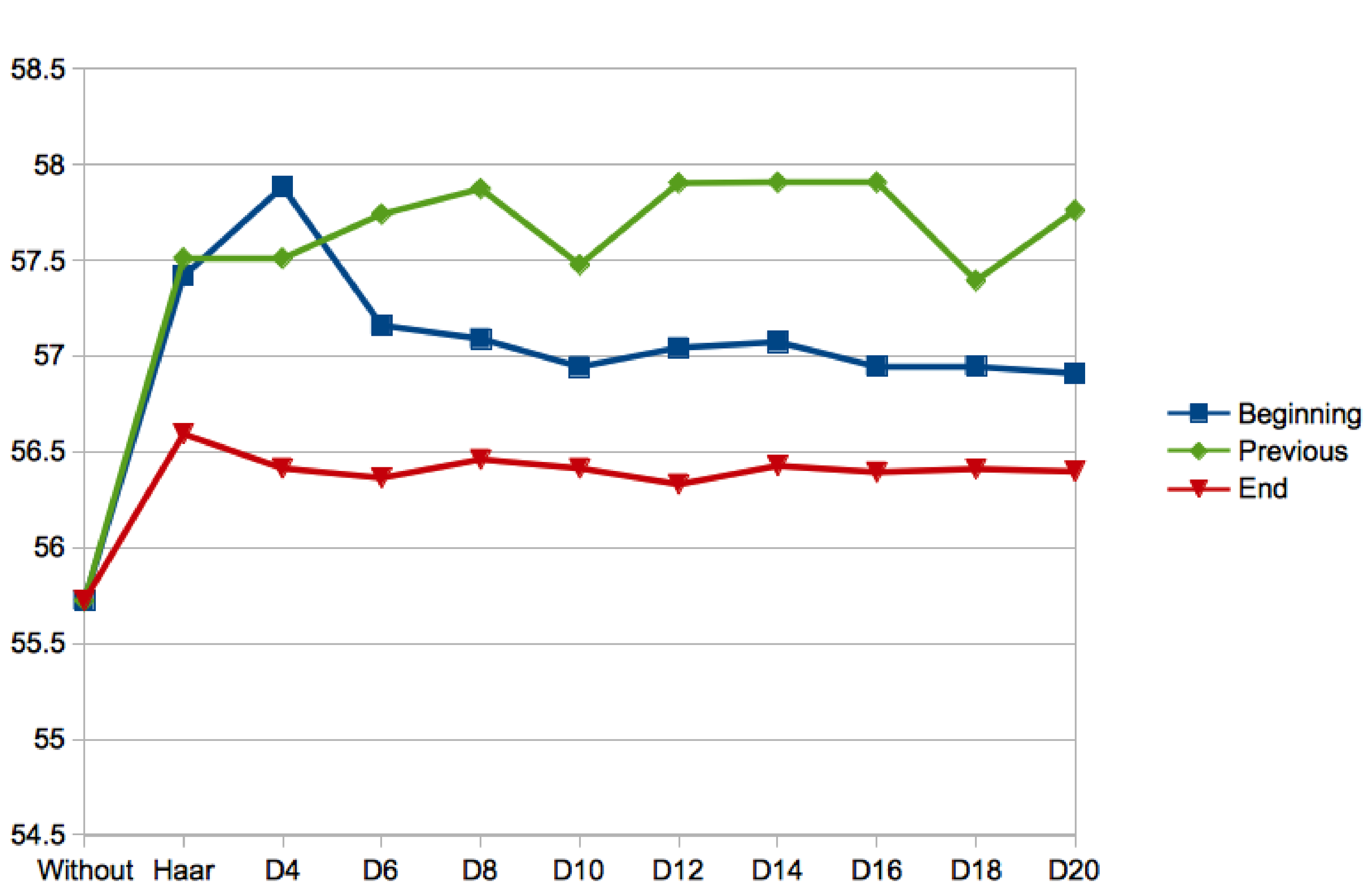
Hedging

- Financial operation.
- Dealing with potential financial loss.

Methodology

- Classifications are generated for a specific date, then compared to the actual occurrence.
  - The classifications themselves is the *volume of buys and sells of a specific currency.*
    - The quantity traded at a specific time.
  - The classifications will be generated using SVM learning algorithms.
    - The learning machine is trained two-weeks at a time.
      - the algorithm learns for two-weeks then classifications for the next two weeks.
  - Done across the time-line to eventually calculate a success percentage.
- This was done with a time-line where wavelet coefficients where attached. As well as a time-series where wavelets were not attached.
  - Used Daubachies (2-20) set of wavelets.
  - Then compared each wavelet to each other and the original classifications without wavelets
- Divided time-line using three different methods and compared to each other.
  - Generating coefficients between the beginning to current feature
  - Dividing time-series between the current feature and the end.
  - Dividing time-series between the current and the feature directly before the current feature (previous feature)





Wavelets

- Mathematical concept like Fourier Analysis. But better able to zoom into High-Frequency Signals.
- Creates a transform of a signal which can be reconstructed.

Features

- Starts at Zero and must end at Zero
- $T_{(m,n)} = \int_{-\infty}^{\infty} x(t) \frac{1}{a_0^{m/2}} \varphi(a_0^{-m} t - n b_0)\, dt$
  - a_0 is a specific fixed dilation step parameter.
  - b_0 is the translation parameter.
  - m and n control the wavelet dilation and translation.
  - φ is the wavelet function.
  - $T_{(m,n)}$ is the wavelet (detail) coefficient.
  - When a_0 and b_0 are chosen to be 2 and 1 respectively the wavelet function becomes known as the *dyadic grid arrangement*. In this case the wavelets are chosen to be orthogonal.
    - $T_{(m,n)} = \int_{-\infty}^{\infty} x(t) \vartheta_{m,n}(t) dt$

Multiresolution Analysis

- This was what was used.
- Creates an orthonormal basis.
- Scale independent representation.
- Has two different functions.
  - A mother wavelet (shown above).
  - A father (scaling) wavelet.
    - The father wavelet has the property
      - $\int_{-\infty}^{\infty} \delta_{(0,0)}(t) dt = 1$
      - To generate approximate coefficients
        - $S_{(m,n)} = \int_{-\infty}^{\infty} x(t) \delta_{m,n}(t) dt$
- Used Multiresolution Analysis for financial-time series.
  - Used by Mallat (1988) for geographical analysis.
  - Both scaling and detail coefficient were used.

Daubachies

- Standard for wavelet analysis.
- D2-D20 (goes up by even numbers).
  - The index is the number of coefficients.
  - The number of vanishing moments is the (index/2).

Results

- All 90 executions with Wavelets performed better than without wavelets.

Division

- Generating wavelets from the current hour until its previous hour preforms the best.
- Generating wavelets from the current hour until the end of the time-line performs the worst.

Wavelets

- All Wavelets increased accuracy of classifications.
- D4 wavelet performed the best on average.
- However the best results where D12, D14 and D16 Wavelets.

Conclusion

- Wavelets are able to increase the rate of classification.
- A wavelet library for machine learning applications was developed
- The library was applied to improve the classification of the time-series.
- The library is open source and can be used by industry for time-series prediction applications