Wavelet Transform Visualization, an Application on Financial Data

CSE6242 Project, Team 148

Bryan F. Gray
bgray44@gatech.edu

Daniel H. Nitiutomo
dnitiutomo3@gatech.edu

Ryan Plain
rplain3@gatech.edu

M. A. A. Shihab
mshihab6@gatech.edu

Anatoly Stolbov
astolbov3@gatech.edu

1 INTRODUCTION

In recent decades Wavelet Transform (WT) has gained popularity in both data analysis and applied science. Its list of applications includes, but is not limited to: denoising, feature extraction, time frequency analysis, prediction, and multi resolution analysis[17].

Wavelet Transform was originally inspired by Fourier transform (FT). It allows for the decomposition of a signal or time series into a set of frequency based components without suffering some of the common drawbacks of FT. One specific benefit of using WT is its ability to capture both time and frequency domains in a single representation. In addition to this, WT allows one to work with non-stationary data, which is critical for financial time series analysis. Please refer to Kim S. and In F. for a thorough introduction to the theory behind WT [7] or Crowley [6] for a less technical introduction to WT with multiple examples using economic data.

2 PROBLEM DEFINITION

The problem we attempted to address with this project is that we wanted to apply WT in a new and novel way. Specifically, we wanted to find out if WT could be leveraged to visually analyze and interpret financial market behavior over time. In addition to this, we wanted to build a sandbox where one could test and apply multiple wavelets on different financial time series. This is something that had not been done before. To recap, our application serves two main functions:

- Tool for seamless multi-resolution or frequency domain analysis of financial time series data (examples [5], [8])
- WT sandbox purpose built to better understand a relatively new approach in financial data analysis

3 SURVEY

Wavelet Transform has multiple applications in finance. It is often used to denoise or pre-process data before applying predictive algorithms [3, 9, 10, 13, 14]. For example, Luo [12] applied Discrete Wavelet Transform (DWT) to enhance the feature extraction component of their pattern recognition algorithm. On the other hand, Bogdanova [5] and Kilic [8] used a Continuous Wavelet Transform (CWT) power spectrum to visually analyze how market behavior changes over the time. We were more interested in the latter approach.

Continuous Wavelet Transform is the central component of our visualisation. It produces a color map where users can see the influence of different frequency components and how those interact with different time periods. For the purpose of visual analysis with CWT, most researchers often choose Morlet wavelet due to its support in SW packages and general popularity [2, 5, 8]. However Aguiar-Conraria [2] and Lilly [11] argued Morse wavelets demonstrated superior qualities over Morlet.

With regards to Discrete Wavelet Transform, Ortega [14] and Berger [4] argued for the Haar wavelet due to its suitability for prediction purposes. While Ozun [15] and Khaled [3] have successfully applied Daubechies wavelets for pre-processing before implementing prediction algorithms. Also, Razak [1] empirically showed that Daubechies wavelets outperform others in financial data denoising.

In our application we will include multiple wavelet functions to allow user compare functions and the results they produce using DWT and CWT.

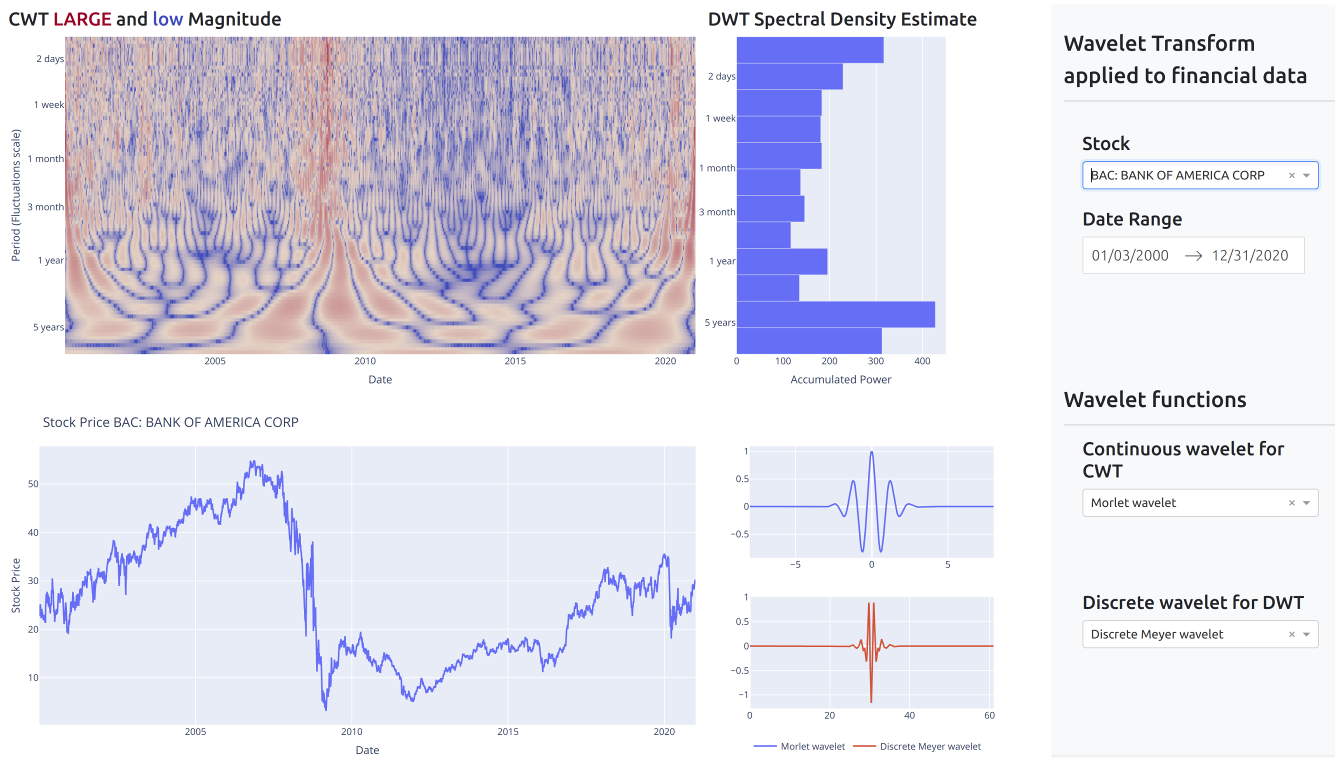

Figure 1: Visualization ui overview (see 4.1 for details)

4 PROPOSED METHOD

Below, we have detailed the methods in which we collected data, coded our algorithms, and ultimately built our user interface.

4.1 Overview

The application provides users with a playground where they can apply wavelet transformations on financial data. For an example of our application, please refer to Figure 1. The user interface includes following visualization parts:

- Contour plot of the CWT with date and period as the x and y axis respectively and associated density values as a color map
- Horizontal bar chart with a power spectrum estimation computed from DWT decomposition
- Line chart of stock price over time
- The respective discrete and continuous wavelet functions that applied to CWT and DWT

All fields for the user input are conveniently located in single panel on right side the screen. Inputs allow the end user to interact with the chart and choose:

- specific stocks, bonds, or economic data
- time interval for the analysis
- wavelet function for both CWT and DWT transforms

4.2 Wavelet transform

Analogous to a common power spectrum produced using Fourier Transform, the spectogram from WT provides insights about frequency content of the signal. While FT fits data with cosine and sine waves, WT fits data with small, limited in time waves, or wavelets. FT power spectrum answers the question: what is the amplitude of a particular wave frequency in the signal? WT spectograms, on the other hand, answer: how pronounced is the pattern of a selected wavelet over a particular period? In short, WT allows us to analyze fluctuations of a given frequency localized in time.

Continues Wavelet Transform is continuous and its coefficients are estimated for a wavelet function of any given period length at any point in time.

Discrete Wavelet Transform is discrete, because it can be calculated only for functions with scale equal to power of 2 (2, 4, 8...). DWT decomposes time series into a number components with defined frequency. The decomposition process is somewhat similar to filtering certain frequencies ranging from small to large.

The main application for DWT is denoising. Not too long ago it replaced Fourier transform in sounds applications and lead to a significant improvement in quality of cellphone calls. While its useful for other applications, it isn't popular in construction of a spectrum density charts. Our computation of DWT density spectrum was inspired by formula of CWT energy density in [16]:

$$W^2(\lambda, t)/\lambda^2$$

where W is value of wavelet coefficient in decomposition, λ is scaling factor or length of the wavelet period, t - time of the observation. Since DWT components can't be estimated at all given times, we simply compute a sum of all available observations. The accumulated power of a given periodic component stands for quadratic sum of its variance, normalized by the scale of applied wavelet in decomposition.

4.3 Visualization decisions

In order to outline the effort and care we put into the application design, we'd decided to share a few particular visualization decisions:

(1) The layout of the chart is optimized so that the similar axis of any two charts is aligned. The CWT and DWT chart share similar y axis period, which stands for scale or length of fluctuations. In addition to this, the CWT and price chart share exactly the same x axis timeline.
(2) We decided to add charts of wavelet functions so the end user could see the exact function they were applying to their analysis. Both charts are conveniently placed next to its selection drop downs.
(3) We then hid the color bar of the CWT because it disrupted visual flow between CWT and DWT charts and sometimes caused some of the other charts and widgets to become misaligned. Because of this edit, we then added color coding to HTML title to outline the ends of the color spectrum.
(4) We chose the log scale for the vertical axis of DWT and CWT chart due to the nature of the data as well as data availability for DWT.
(5) We replaced exact data values of the tick labels with approximate location of common time measures in weeks, month and years to further improve readability
(6) We limited all color shades on purpose to blue and red to improve aesthetics

4.4 Data Collection

Our aim was to collect different types of financial time series that could be relevant to investors. Those time series include daily stock data from SP 500 companies, daily global index price data, daily aggregate bond yield data, and various economic time series at varying frequencies such as commodity prices (daily), inflation indices (monthly), economic output data (monthly and quarterly), employment (weekly and monthly), and interest rates (daily and monthly). The range of the time series collected was from Jan-2000 to Dec-2020.

We sourced the data from 2 online data repositories, Wharton Research Data Services (WRDS) and FRED (Federal Reserve of Economic Data) from the Federal Reserve Bank of St. Louis. The security and index pricing data was sourced from the WRDS CRSP (Center for Research in Security Prices) data set. In order to retrieve that data, we navigated to the WRDS website and requested CSV files for each of the 500 securities in the SP 500 index. Next, we used Python scripts to manipulate the price time series within each of the CSV files, to adjust for stock splits, and to input them into the SQLlite database. For the economic time series data from FRED, we used the FREDAPI python package to query the relevant time series and store them.

The ending result of our data collection was 2 tables within the database: a lookup table with metadata about each time series(name, ticker, / category) and another with all of time series data. Each data set has a unique key or identifier value that joins the two tables together. The resulting database had around 2.3 million data points and consumed about 180MB in storage. Lastly, the reaspm we chose to store all this data was SQLite was because it was lightweight and easy to use.

4.5 Technology

Our platform was built using Dash which is a python framework created by Plotly for creating interactive web applications. It allowed us to self-contain everything within the python ecosystem while leveraging scientific, web application, and visualization libraries.

To collaborate on the development we stored our source code in a git repository that could then be cloned. For distribution purposes we then utilized published image in docker.

List of python packages: dash, dash-bootstrap-components, plotly, pandas, numpy, matplotlib, sqlalchemy, PyWavelets

5 EXPERIMENTS/EVALUATION

5.1 Questions we hope to answer

- Does CWT show any trends/insights in the data?
- How does WT align with pricing information/chart?
- How results from CWT align with DWT?
- How do results from different wavelet functions compare with each other?
- What WT algorithm works best for financial data?
- Can any predictions or investment decisions be made with this application?

5.2 Experiment Details and Evaluation

For the purposes of evaluating our project we really were focused on two goals:

(1) Does our application allow us to visually analyze and interpret financial market behavior over time for a given security? Are we able to identify any trends?
(2) Did we successfully build a sandbox where one could test and apply multiple wavelets on different financial time series?

With regards to evaluating question 1, we took an experimental approach. We tried to identify as many situations as possible where significant stock market events occurred and then leveraged our application to apply different wavelets on those situations. We then tried to make interpretations by coupling what our wavelet charts were showing us with what we already knew about the selected financial time series. We tested many examples including, but not limited to:

- BP Oil Spill
- Housing Crisis and Lehman Brothers Collapse
- Oil crash of 2014
- COVID Outbreak
- Quantitative Easing

Since our evaluation was based on analyzing the charts we produced we thought it would be useful to step through a specific example to show how we interpreted the charts. For example lets consider Bank of America during the years of 2006 to 2012. Bank of America was one of the securities that was negatively affected by the 2008 Housing Market crash. From December of 2006 to February of 2009, the stock price fell by 95% and has yet to reclaim its 2006 all time high. Given the fundamental backdrop, Bank of America was an ideal time series for analysis.

First, lets look at our Continuous Wavelet function. Below is what the Morlet Wavelet function looks like. This is what the Continuous Wavelet Transform will leverage to transform the pricing data.

Figure 2: Morlet Wavelet

Next, please see the Continuous Wavelet Transform chart below in addition to its corresponding price chart. The CWT chart shows the outcome of the CWT transform and its corresponding coefficients are illustrated with a color map. The x axis represents the timeline, while the y axis represents the length of applied wavelet function or the period. The color shows the magnitude of the wavelet of a given length for a given time. So when the chart is red, that means there is a significant fluctuation matching the shape of the wavelet function. The darker the shade of red, the larger the magnitude of the price fluctuation. In other words, the red color shows significant price movements.

In the case of Bank of America, we saw hotspots in 2007 during the 2 day periods. This showed us that there was a lot of volatility in the market and could signaling that a big event may be coming. In 2008, you can see the darkest part of the chart at the 2 day period when the housing bubble burst, and the market began to crash taking Bank of America along with it. Also, some other interesting cycles can be seen at the 1 month and 1 year periods. You can see that during those time periods there were some pretty significant price movements that that all occurred right after each other.

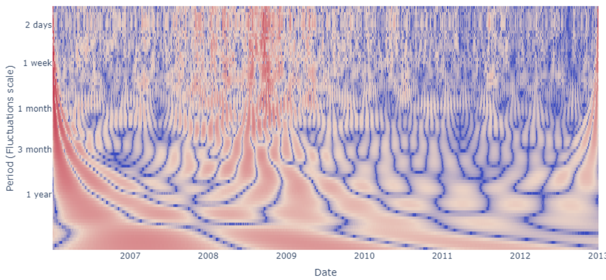

Figure 3: Continuous Wavelet Transform

Figure 4: Bank of America Price Chart

Below we performed Discrete Wavelet Transform. In this example we fit the Discrete Meyer wavelet again to the pricing data. The chart we produced using DWT is a little easier to interpret. It summarizes the significant frequencies or periods within the time horizon that was selected. The higher the value on the x axis in the DWT chart, the more significant the fluctuation of the given length. DWT was consistent with the analysis above and the 2 day period, 1 month, and a year period all were significant for this specific stock and given time horizon.

Given the above analysis, we think it is clear that both CWT and DWT have the ability to uncover insights in data that were not observable in a standard time series line chart. They can quickly identify periods of high volatility and what specific frequency components were the most pronounced. However, whether or not the wavelet transform has the ability to make predictions or find signals in new data is up for debate. It looks like WT could potentially detect new up or down cycles but the

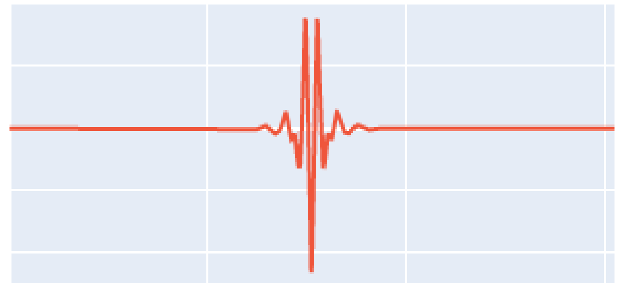

Figure 5: Discrete Meyer Wavelet

Figure 6: Discrete Wavelet Transform

signal may not come soon enough. This is something we could research further in future iterations of our application.

Question 2 was much easier to evaluate. We did successfully create an application where one could choose a security and apply multiple wavelets on different time series. The sandbox allows for individual research and exploration of wavelets without writing a single line of code. This is something that could be very useful if built upon further and be a useful tool for researchers, investors, etc.

Lastly, some other general observations we had are detailed below:

- The choice of wavelet can cause vastly different results. For example, the Haar function is much different than some of the others. We believe different wavelets might be more useful for different time series as well as different research applications.
- While WT is praised for handling non-stationary time series, it still has its limitations. In case of exponential trends the method seems to lose its informational value. In this case we recommend reducing time interval or transforming data to log-returns.
- WT's visual interpretation seems to be more of an art than a science. Different functions tell you different things. Some are functions are useful for some time series while they are not useful for others. WT requires more research but definitely provides a lot of value already.
- In the future, it would be interesting to iterate through the different wavelet functions and backtest which ones perform best multiple time series.
- It would be interesting to couple WT with other machine learning algorithms to see if they result in a better understanding of financial markets. Specifically, training neural networks with WT or other algorithms would be an interesting application to explore further.

6 WORKLOAD DISTRIBUTION

Our general workload distribution is detailed below. Each team member was assigned an aspect of the project in which they were the sub-project manager and were responsible for making sure it was completed in a timely fashion. With that being said, each team member was still be somewhat involved in all other aspects of the final project. The visualization and algorithm in particular, required a lot of resources and had input by most if not all of our team members.

- Daniel gathered all of the data, cleaned it, and maintained the database. He also took ownership of our poster document.
- Anatoly took ownership of the backend and programmed the algorithms. He also proposed visualization structure, and lead the overall application development. Lastly, he authored idea, wrote the proposal, and helped edit the final report.
- Ryan and Mahmoud took ownership of the web application and front end, integrating visualization structure from Anatoly to Plotly. Mahmouod was in a completely different timezone, and Ryan took responsibility of collaborating and updating Mahmoud. They assisted the team with communication, took care of the docker distribution, YouTube installation guide, and helped with the proposal presentation.
- Bryan edited and wrote proposal report, did the proposal presentation, progress report, and wrote the final report.

7 CONCLUSIONS AND DISCUSSION

In conclusion, our application has proven useful for both the interpretation of trends in financial data as well as a sandbox environment which could be beneficial in furthering research on Wavelet Transformation. It is particularly helpful when looking at financial data because it allows a user to decompose the time series into frequency based components and avoids drawbacks other methods like Fourier Transform have with non-stationary data. While Wavelet Transform, is still a relatively new topic, it already shows promise in the decomposition and understanding of financial data. However, in WT still leaves a lot of room for interpretation and human interaction. As research progresses, we expect Wavelet Transform based algorithms gain even more traction in the finance community and could even be used to make investment decisions in the future.

REFERENCES

[1] Noor 'Aina Abdul Razak, Rasimah Aripin, and Mohd Tahir Ismail. 2010. Denoising Malaysian time series data: A comparison using discrete and stationary wavelet transforms. In *2010 International Conference on Science and Social Research (CSSR 2010)*. IEEE, 412–415.

[2] Luís Aguiar-Conraria and Maria Joana Soares. 2014. THE CONTINUOUS WAVELET TRANSFORM: MOVING BEYOND UNI- AND BIVARIATE ANALYSIS. *Journal of economic surveys* 28, 2 (2014), 344–375.

[3] Khaled A. Althelaya, Salahadin A. Mohammed, and El-Sayed M. El-Alfy. 2021. Combining Deep Learning and Multiresolution Analysis for Stock Market Forecasting. *IEEE Access* 9 (2021), 13099–13111. https://doi.org/10.1109/ACCESS.2021.3051872

[4] Theo Berger. 2016. Forecasting Based on Decomposed Financial Return Series: A Wavelet Analysis. *Journal of forecasting* 35, 5 (2016), 419–433.

[5] Boryana Bogdanova and Bozhidar Nedev. 2019. Volatility breakdown of the momentum effect observed at the New York Stock Exchange. In *AIP conference proceedings*, Vol. 2172.

[6] Patrick M Crowley. 2007. A GUIDE TO WAVELETS FOR ECONOMISTS. *Journal of economic surveys* 21, 2 (2007), 207–267.

[7] F. H. In and S. Kim. 2012. *An Introduction to Wavelet Theory In Finance, A Wavelet Multiscale Approach*. World Scientific Publishing Company, New York.

[8] Deniz Kenan Kilic and Omur Ugur. 2018. Multiresolution analysis of SP500 time series. *Annals of operations research* 260, 1 (2018), 197–216.

[9] S LAHMIRI. 2014. IMPROVING FORECASTING ACCURACY OF THE SP500 INTRA-DAY PRICE DIRECTION USING BOTH WAVELET LOW AND HIGH FREQUENCY COEFFICIENTS. *Fluctuation and noise letters* 13, 1 (2014), 1450008.

[10] Zhaodi; Sun Liling; He Maowei; Chen Hanning Liang, Xiaodan; Ge. 2019. LSTM with Wavelet Transform Based Data Preprocessing for Stock Price Prediction. *Mathematical Problems in Engineering* 2019 (2019). https://doi.org/10.1155/2019/1340174

[11] J.M Lilly and S.C Olhede. 2009. Higher-Order Properties of Analytic Wavelets. *IEEE transactions on signal processing* 57, 1 (2009), 146–160.

[12] Ji Wei Luo. 2021. A Study on Stock Graph Recognition Based on Wavelet Denoising and DTW Algorithm. *Mathematical problems in engineering* 2021 (2021).

[13] João Nobre and Rui Ferreira Neves. 2019. Combining Principal Component Analysis, Discrete Wavelet Transform and XGBoost to trade in the financial markets. *Expert Systems with Applications* 125 (2019), 181–194. https://doi.org/10.1016/j.eswa.2019.01.083

[14] Luis Ortega and Khaldoun Khashanah. 2014. A Neuro-wavelet Model for the Short-Term Forecasting of High-Frequency Time Series of Stock Returns. *Journal of forecasting* 33, 2 (2014), 134–146.

[15] Alper Ozun and Atilla Cifter. 2008. Modeling long-term memory effect in stock prices: A comparative analysis with GPH test and Daubechies wavelets. *Studies in economics and finance (Charlotte, N.C.)* 25, 1 (2008), 38–48.

[16] Donald B. Percival and Andrew T. Walden. 2007. *Wavelet methods for time series analysis*. Cambridge Univ. Press, Cambridge [u.a. http://www.worldcat.org/title/wavelet-methods-for-time-series-analysis/oclc/254371438

[17] Manel Rhif, Ali Ben Abbes, Imed Farah, Beatriz Martínez, and Yanfang Sang. 2019. Wavelet Transform Application for/in Non-Stationary Time-Series Analysis: A Review. *Applied sciences* 9, 7 (2019), 1345.