J.N. Andrews
Honors Program
Andrews University

Proposal for Senior Honors Thesis

HONS 497 Senior Honors Thesis   **Credits** 2 (2 minimums required)

Directions: Please return signed proposal to the Honors Office **at least one week prior to your scheduled meeting with the Honors Council**. This proposal must be accepted by Honors Council the semester before presentation.

Student's Name: Lloyd Tumaneng Martinez Jr
Primary Advisor: Jerry Chi, MBA, Ph.D., Ph.D.
Secondary Advisor:
Thesis Title: The Effective Predictors of the SPX Index
Local phone: 719-235-8075   Email: martinezl@andrews.edu

Expected date of Graduation: May 2021

I. Provide goals and brief description of your project or research.

In the finance community, academics and professionals spend amount of effort trying to predict the price of the stock market. The stock markets a collection of markets and exchanges where trades are made daily with over millions of dollars being transacted. The incentive for most investors in the financial market is to make money. This makes modeling and forecasting financial asset returns within the market exchange crucial as it impacts key-decision making in areas such as portfolio and risk management. While there has been a denial of forecasting stock market prices, there have been studies that contend that price trends and correlations between economic figures allow the market to be predictable to a certain extent.

When examining the time-varying fluctuation of a financial asset, econometric models use point data for a given observation in time. Econometric models make use of statistical tools and economic theories in combination with economic variables to forecast an intended variable. The daily data on the asset's closing price has been the prevailing choice in filtering out intra-day volatility. This research project aims to use multiple linear regression to predict the closing price of the SPX index on a daily, weekly, and monthly basis using our independent variables: VIX (Volatility Index), IORR (Interest Rate on Required Reserves), TB (10-Year Treasury Constant Maturity Minus 2-Year Treasury Constant Maturity), CB (ICE BofA US High Yield Index Option-Adjusted Spread), GC (Gold Price), and TWUSDX (Trade Weighted U.S. Dollar Index: Broad, Goods and Services). The SPX is a stock index that is comprised of the 500 largest U.S. publicly traded companies by market capitalization. Some companies traded on this index include Microsoft and Apple. If markets are predictable, how accurate is our prediction of the actual market? We focus on using modern statistical techniques such as linear regression, highlighting the main challenges within financial markets, and achievements in index forecasting by using variables not commonly used for analysis and forecasting market price. Finally, we discuss potential challenges and possible directions for future research.

II. Outline your methodology. Please be specific. How does this achieve your goals and how reliable is it?

We gathered data from the Wall Street Journal's (WSJ) online database, Federal Reserve Economic Data (FRED) online database, and Chicago Board of Exchange (CBOE) database. We recorded historical daily prices from the following on an xlsx document: WSJ (SPX and GC), CBOE (Option Volume, and VIX), and FRED (IORR, TB, CB, and TWUSDX). Each website provided a tool to fetch the closing price of each variable from 7 July 2010 to 3 October 2019 eliminating the need to mine the data manually. The closing price is an important piece of information useful for swing traders and positional traders. It tells a lot about the thinking of big investors that allocate large amounts of money into the market for asset management purposes. Hence, the closing price has practical significance for portfolio managers and investors (Seethalakshmi, 2018).

The data was processed to fit into a regression model. First, dates were represented in string format 'MM-DD-YYYY' followed by a column ordered by 'Weeks #'. Next, we deleted any additional numbers for independent variables that did not correspond to the date of the dependent variable. Afterward, we analyzed each independent variable using descriptive analysis. Next, we used the mean value of the independent variable for any missing value when the independent variable did not contain a value, and the dependent variable contained value.

We attempt to use linear regression to fit the data utilizing past data rather than an actual prediction of future prices. We can demonstrate that the method can work in real-time to forecast future value by showing how well the model fits in a controlled environment. We considered several variables for the prediction of the SPX: VIX, IORR, TB, CB, GC, and TWUSDX. We used the closing prices of each of these variables for the data.

First, we utilized a correlation analysis to determine whether our factors relate to the SPX index. Second, we develop a predictive model using the data from the linear regression as a linear regression model tries to provide the best

possible fit. Third, we use a T-test to test the statistical significance of the model. Lastly, we check the model through the forecasting performance evaluation.

We measure the accuracy of our models through important metrics including mean absolute error (MAE), mean square error (MSE), root-mean-square error (RMSE), mean absolute percentage error (MAPE), mean percentage error (MPE), and adjusted r-square. We measure the accuracy of our models through these metrics because MAE measures the average size of error generated by the regression model, MSE penalizes our outliers. Much harder, RMSE represents the standard deviation of the residuals in our model, MAPE allows to see what percentage is our standard deviation in percent value rather than numerical value, MPE provides a comparison between positive and negative errors, and adjusted-r square because it takes into account the number of predictors we have in the model whereas r-square does not.

We can verify whether how close our predicted model compares to the actual model by using these forecasting error functions. Comparing the forecasting performance of competing models is an important aspect of any forecast model. Thus, we employ three different models in this report on a daily, weekly, and monthly basis.

The descriptive statistics, correlation test, linear regression analysis, and forecasting error tests were performed on Excel.

Equation for calculating correlation is given below.

$$Correlation = \frac{N \sum xy - (\sum x)(\sum y)}{\sqrt{[N \sum x^2 - (\sum (x)^2][N \sum (y)^2]}}$$

Equation for calculating linear regression is given below.

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \epsilon$$

Forecast Error Check:

$$\hat{\sigma}_t = forecasted\ value\ |\ \sigma = actual\ value$$

Equation for calculating Mean Absolute Error (MAE) is given below.

$$MAE = \frac{1}{N} \sum_{i=1}^{n} |(\hat{\sigma}_t - \sigma_t)^2|$$

Equation for calculating Mean Square Error (MSE) is given below.

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (\hat{\sigma}_t - \sigma_t)^2$$

Equation for calculating Root Mean Square Error (RMSE) is given below.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{\sigma}_t - \sigma_t)^2}$$

Equation for calculating Mean Absolute Percent Error (MAPE) is given below.

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{N} \frac{|\hat{\sigma}_t - \sigma_t|}{\sigma_t}$$

Equation for calculating Mean Percent Error (MPE) is given below.

$$MPE = \frac{100\%}{n} \sum_{i=1}^{N} \frac{(\hat{\sigma}_t - \sigma_t)}{\sigma_t}$$

III. **Explain in what sense your project is original, unique, or beyond normal senior expectations. How does it relate to current knowledge in the discipline?**

While nearly all faculty in the AU School of Business, such as my mentor Dr. Chi, involve themselves in research endeavors, the vast majority of students pursuing business degrees do not actively participate in research during their undergraduate studies. Outside of the J.N. Andrews Honors programs at any school, business students are not expected to have an undergraduate research study as most research techniques within the field are taught at the Master's degree level. At most colleges, being an honor student of their respective college requires them to undertake a senior thesis whereas normal seniors are not expected to undergo the task. Furthermore, business students are intended to learn more about their respective fields rather than delve into research as most people are first introduced to financial terms in their college years unlike other fields such as science. While business students are required to work on a paper or projects, most submitted papers do not require students to look at past works but rather give their perspective using terminology or ideas they

2

learned in class. Thus, there is hardly any cross-referencing or items learned in other business classes. In this case, this project would represent an effort that is 'above and beyond' of an expected senior student at their undergraduate level. While the techniques utilized in this project are used in other fields, there has been no study done at this school on predicting prices uses economic variables and sentiments within the financial markets.

Apart from Andrews University-related matters, my topic is unique in which some of the independent variables (TB, CB, and TWUSDX) chosen for this project are not specifically studied or looked at within the financial industry. Furthermore, the implementation of new ways of trading affects market sentiment ultimately impacting how markets move. Therefore, this project will add to current knowledge by providing more knowledge on how variables not usually analyzed in the financial industry can impact financial market prediction.

IV. **Include a substantive annotated bibliography of similar or related work.**

Abu-Mostafa, Y.S., Atiya, A.F. (1996). Introduction to Financial Forecasting. *Applied Intelligence,* 6, 205–213. doi: 10.1007/BF00126626

Abu-Mostafa and Atiya provide a brief introduction to forecasting in financial markets, describe basic approaches to forecasting, and discuss the noisy nature incorporated into financial data. This article concluded that EMH was either hard to prove and hard to disprove but noted that there were price trends and correlations in economic figures that proved EMH to not be entirely true. In my study, we are assuming the price does not fully reflect the market. Thus, in my thesis, while forecasting may cause a challenge due to its noisy and disordered nature, we are saying exploring whether it is predictable. While their paper approach forecasting differently, the underlying concepts of forecasting are the same.

Cao, H. (1999). The Effect of Derivative Assets on Information Acquisition and Price Behavior in a Rational Expectations Equilibrium. *The Review of Financial Studies, 12*(1), 131-163. Retrieved July 25, 2020, from www.jstor.org/stable/2645989

Cao's paper is a comprehensive report of how derivates markets have grown in recent years and the impact it has on underlying assets. This paper argues that derivatives assets contribute to excessive market volatility and it is important to understand how it affects stock price volatility, market liquidity, and price efficiency. He shows that the presence of the derivative assets causes "informed investors" to acquire more precise information in advance because they can take advantage of that information in the underlying asset market and the derivatives market. In my study, I use information from both the asset market and derivative market to see whether they are a predictive factor of the SPX index.

Chong, E., Han, C., & Park, F. (2017). Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert Systems With Applications, 83,* 187-205. https://doi.org/10.1016/j.eswa.2017.04.030

Chong and Park address a comprehensive review of the introduction to forecasting using various techniques from econometric or statistical methods to neural networks. They explain the various opinions on how efficient markets are and empirical studies of who supports which side. They address the different methodologies and periods that people utilize. In this study, I utilize their empirical study to place where my research's stands in terms of the divide among the efficiency of markets.

Conrad, J. (1989), The Price Effect of Option Introduction. The Journal of Finance, 44: 487-498. doi:10.1111/j.1540-6261.1989.tb05068.x

Conrad examined the introduction of options into the securities market and the lasting effect is holds within the securities market. They examined the introduction of options from 1974 to 1980 and how it ultimately affected the pricing of various stocks. They concluded that while the systematic risk of securities does not appear to be affected by options introduction, they suggest that there was inside trading denoted by the trading volume of certain stocks before the announcement. This meant that dealers or other traders-built inventory for hedging purposes in anticipation of trading volume in options. In this study, I will be using information built into options as a means to understand the direction they believe markets will move.

Dotsey, Michael, The Predictive Content of the Interest Rate Term Spread for Future Economic Growth (1998). FRB Richmond Economic Quarterly, vol. 84, no. 3, Summer 1998, pp. 31-51, Available at SSRN: https://ssrn.com/abstract=2126284

Dotsey examined whether the interest rate term spread contained any useful information on future economic activity and the probability of recession. They concluded that the spread of interest rates contained information, not within past economic growth and monetary policy. While the spread's forecasting performance has declined, there were still noticeable times in which the term spread outweighed other predictive factors. I use this information in my thesis to explain the usage of interest rates and the time it takes to impact the economy and securities market.

Fama, Eugene F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance, 25*(2), 383-417. doi:10.2307/2325486

Fama questions the idea of predicting stock price behavior and expanding on the works of the Efficient Market Hypothesis. The basic assumption of the technical theory is that history repeats itself making past patterns of price repeat in the future. The premise of questioning analyzing future behavior stems from "efficient market" and random-walk theorists where, at any point in time, actual prices of securities already reflect information based on events that already occurred. This article is

the opposite of how I explore my paper but a premise of understanding how there are different theories within the securities market.

Fama, Eugene F. (1995). Random walks in stock market prices. *Financial Analysts Journal* 51, 75–80.
Fama's selected paper defined the Efficient Market Hypothesis as: "That security prices at any point in time "fully reflect" all available information." This theory is concerned with whether prices at any point in time "fully reflect" available information. This paper is the basis of most finance major's understanding of the securities market and an opponent to my thesis paper.

Fernandes, M., Medeiros, M., & Scharth, M. (2014). Modeling and predicting the CBOE market volatility index. *Journal Of Banking & Finance, 40,* 1-10. https://doi.org/10.1016/j.jbankfin.2013.11.004
Marcel et. al. check whether the VIX index is dependent on other rates such as Fed Rates and credit spreads. They conclude that the VIX does not rely on such measures and holds a negative relationship with the S&P 500 index returns and a positive link to the S&P 500 volume change. In. my thesis paper, I use the VIX as a bias indicator of how steady the securities markets are.

Hentschel, L. (2003). Errors in Implied Volatility Estimation. *The Journal of Financial and Quantitative Analysis, 38*(4), 779-810. doi:10.2307/4126743
Hentschel reviews the usability of the VIX. They concluded it was biased to a certain degree but it did what it was meant to do, predict volatility. It is biased in that it is subject to errors including noise and option bias. While the VIX calculation has changed since this article was written, I use this article as an affirmation in my study that the factor is biased in some aspects.

M, H., E.A., G., Menon, V., & K.P., S. (2018). NSE Stock Market Prediction Using Deep-Learning Models. *Procedia Computer Science, 132,* 1351-1362. https://doi.org/10.1016/j.procs.2018.05.050
While this article focuses on different deep-learning models, it does explain the limitations and usages of linear and non-linear models. They emphasize the usage of the closing price of two different markets: the National Stock Exchange (NSE) of India and the New York Stock Exchange (NYSE). This article concludes that Neural Network outperforms ARIMA models because they fail to identify non-linearities within data. This is important because this study's model is based on a linear model. This allows me to use my current data and do future research using a Neural Network Model.

Poon, Ser-Huang, and Clive W.J. Graner (2003). Forecasting Volatility in Financial Markets: A Review. *Journal of Economic Literature.* Vol. XLI (June 2003) pp. 478-539.
In this review, Poon and Granger concentrate on whether volatility is forecastable. The reason it's important to see whether it truly is forecastable is that the VIX is a forecasting tool for future price movement. They concluded that financial market volatility was forecastable to certain extents. This study provides affirmation that there are factors, such as the VIX, in predicting the SPX underlying value.

Stephan, J.A. and Whaley, R.E. (1990), Intraday Price Change and Trading Volume Relations in the Stock and Stock Option Markets. The Journal of Finance, 45: 191-220. doi:10.1111/j.1540-6261.1990.tb05087.x
Stephan and Whaley explore the relationship between price changes and trading volume of options and stock. They concluded that prices change in the stock market lead to price changes in the options market. This is quite different from what I am trying to achieve but it's important to note for my limitations and the efficiency in stock prices versus the options market.

Whaley, Robert E., Understanding the VIX (November 6, 2008). https://doi.org/10.3905/JPM.2009.35.3.098, Available at SSRN: https://ssrn.com/abstract=1296743 or http://dx.doi.org/10.2139/ssrn.1296743
Whaley created a comprehensive review of understanding the VIX. It goes through the history and talks about its actual implications within the stock market. The author, who first introduced the VIX in 1993, talked about how the media uses the actual index incorrectly. In my thesis, I use his understanding of why the VIX was created and how it holds some predictive factor for how much the SPX index could move affirming the idea of stock market prediction.

Zhong, X., & Enke, D. (2017). Forecasting daily stock market return using dimensionality reduction. *Expert Systems With Applications, 67,* 126-139. https://doi.org/10.1016/j.eswa.2016.09.027
Zhong and Enke discuss forecasting the daily direction of the stock market (SPY) utilizing 60 financial and economic features. They utilize different factors because stock markets have many interrelated factors: economical, industrial, company-specific, psychological, and political variables. In my study, we use multiple economic variables including interest rates and exchange rates, and psychological variables as investor's expectations of markets and institutional investor's choices by utilizing the closing price of the market. Furthermore, some of the variables utilized in their study will also be used in this study including gold, treasury bills rates, and exchange rates between the US and other currencies.

V.   **Provide a statement of progress to date and list the research methods coursework completed.**
      As of August 2020, while there is no required research coursework for my Finance major, I have completed relevant coursework (Business Finance, International Finance, Macroeconomic theory, Macroeconomic theory, Economics

Money and Banking, Operation Management, and Statistics) that have an impact on my thesis. I also completed the required Honors credit (Research Pro-Seminar). I began my research at the end of the Fall 2019 semester and continued it during the Spring 2020 semester but have not registered the project until the Summer 2020 semester. Over these semesters, I have completed all parts of my thesis paper on the subject. Thus, I have a clear understanding of relevant past articles and the methodologies utilized for my research. My current step that I am taking is publishing my thesis in a journal. I have found a journal to publish in and only awaiting whether it gets accepted or denied and the next steps to further improve my paper.
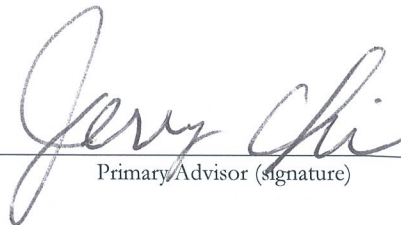
Department Chair Approval
- **This student's performance in his/her major field is acceptable.**
- **He/she has completed the requisite research methods coursework for the research to be pursued.**
- **I understand that he/she plans to graduate with Honors.**

_____
Department Chair (signature)

Research Advisor Approval

I have read and support this proposal:

_____
Primary Advisor (signature)

I have read and support this proposal:

_____
Secondary Advisor (signature)

If human subjects or if live vertebrate animals are involved, evidence of approval from the Institutional Review Board or an Animal Use Committee is needed through the campus scholarly research offices (Ext. 6361).