The Search to Improve Status Quo Rationale for Stock Market Investments in Order to Help Middle-income Households

Ishan Gaur

Palos Verdes Peninsula High School, Rancho Palos Verdes, CA

Teacher

Mr. Peter Starodub

Table of Contents

1. PURPOSE **1**

1.1 Problems with Current Investing Practice 1

2. HYPOTHESIS **2**

3. BACKGROUND LITERATURE REVIEW **3**

3. 1 Available Research in Asset Prices 3

3.2 Review of Machine Learning Techniques 6

3.3 Support Vector Machines 6

3.4 Artificial Neural Networks 8

3.5 Deep Neural Networks 9

4. METHODOLOGY **9**

4.1 Portfolio Creation Dataset 10

4.2 Portfolio Creation and Testing 11

4.3 Machine Learning Dataset 14

4.4 Nu Support Vector Regression Model Construction 15

4.5 Deep Neural Net Construction 15

4.6 Linear Regression Modeling 16

5. RESULTS **17**

5.1 Investment Management Using CAPE and Less Frequent Reallocation 17

5.2 Nu Support Vector Regression 19

5.3 Deep Neural Net 20

5.4 Linear Regression 21

6. CONCLUSION **21**

1. Purpose

1.1 The Great Recessions

The Great Recession was the longest economic downturn since WWII and resulted in an aggregate drop of 16 trillion dollars in household net worth over one year (Elwell, 2013). Home prices dropped 30% and the S&P 500 fell 57% by mid-2009 (Rich and Federal Reserve Bank of New York, 2013). The lower 80% of households lost 2 decades of wealth (Vornovitsky, M., Gottschalck, A. and Smith, A., 2012) and economic recovery was greatly slowed down (Tanner and Abdih, 2009) by decreases in declines in consumer spending caused by decreases in the housing and financial markets (Case, Quigley, and Shiller, 2013). Over two years, the Great Recession caused "the largest sustained job loss in seven decades", leaving 8.7 million Americans unemployed (The Economic Policy Institute, 2012).

Even worse, post-crash stock valuations were very attractive, but the lack of trust meant people were afraid to exploit the newfound growth in financial markets – to ride out the "financial storm" (Guiso, Sapienza, and Singales, 2008). Over the next three years, by 2011, financial assets (stocks, bonds, etc.) had generated 96% of household wealth but households only increased their investment in financial assets by 3% on average, most of which was concentrated in the upper echelons of society (Elwell 2013).

1.2 Problems with Current Investing Practice

The public receives a poor education about investing and the stock market (van Rooji, Lusardi, and Alessie, 2011). Individual investors often sell their stocks en masse after a crash and then buy large quantities of stock when prices have been rising for some time, which results in their portfolio underperforming (Wang, 2011).

Furthermore, much research indicates that professional financial advice given to middle-income families is based on a skewed assessment of that client's risk tolerance leading to higher risk taking by the advisor (Guillemette, Finke, and Gilliam, 2012). This can be attributed to the common use of risk tolerance questionnaires, which have been shown to over estimate client tolerance (Guillemette, Finke, and Gilliam, 2012). Based on the results, financial advisors use a "constant mix" method where they maintain a steady split between stocks and bonds (Macey, 2010) under the misconception that a long-term buy-and-hold strategy will consistently bring gains (Grennon, 2016).

For these reasons, this project sought an alternative investment model that financial advisors can use to better manage investments in the stock market and achieve higher rewards at a lower risk than the conventional constant mix and buy-and-hold portfolios.

2. Hypothesis

The hypothesis is that current constant mix and buy-and-hold portfolio allocations can be improved to fulfill the long-term financial goals of middle-income households. The improvements will reduce losses during market crashes and be implemented based on data, instead of market sentiments. Second, a machine learning algorithm trained over economic data will be able to provide medium-term stock market predictions and validate the actions taken by the portfolios. The null hypothesis is that the current model used by financial advisors is the best that we can do for now and that machine learning algorithms cannot predict medium-term changes in financial markets.

3. Background Literature Review

This literature review is divided into two sections: first is a summary of available research in asset prices and second is a review of machine learning algorithms that may supplement the pattern recognition and forecasting of asset prices.

3. 1 Available Research in Asset Prices

In 2013, Eugene Fama, Lars Peter Hansen and Robert Shiller won the Nobel Prize in Economics for developing new methods to identify trends in asset prices (Englund and The Royal Swedish Academy of Science, 2013). Their work demonstrated 3 key ideas about financial markets:

1. Markets are efficient, reflecting arrival of news into changes in the stock prices very quickly, resulting in a random walk in the short term (Fama, 1965).
2. Stock prices fluctuate more than their long-term value: their prices may fluctuate but they follow a predictable long term path (Shiller, 2000).
3. The Generalized Methods of Moments (GMM), the statistical model created by Hansen, disproved the Consumption Capital Assets Pricing Model (CCAPM), which connected asset prices to the decisions of rational individuals (Hansen, 2001). His model also accounted for Fama's results but was unable to explain Shiller's conclusions (Nobel Media AB, 2014).

Nonetheless, Shiller's empirical work showed that over/undervaluation of stock markets could be measured to forecast market changes in the long term (5–10 years) (Shiller, 2000). Shiller popularized the cyclically adjusted price-to-earnings ratio (CAPE, conceived by Benjamin Graham, and David Dodd), which is defined as the price divided by the ten-year trailing average of inflation-adjusted earnings (Lahart, 2016). He demonstrated that CAPE shows an inverse relationship between stock valuations and returns (Shiller, 2000). This is because human irrationality in stock valuation causes over/undervaluation in the long term (Shiller, 2000).

This relationship can be used for asset allocation (Grennon, 2016). However, the method implemented by Grennon used 20/20 vision in hindsight to allocate the money left over from stocks, which skewed the evaluation of CAPE for allocation. Other attempts to use have also CAPE failed to show any statistically significant gains over various time periods (Damodaran, 2016b). In addition, the ratio was found to be non-mean reverting, which means it may be stuck at very high or very low values for extended periods of time (Kantor and Holdsworth, 2014).

CAPE uses the earnings of S&P 500 companies, which are subject to changes in accounting principles over time and lead over/underestimating some market valuations (Lahart, 2016). The Wall Street Journal (WSJ) uses a similar methodology, but it uses Federal Reserve data on the total value of the U.S. stocks instead (Lahart, 2016).

Going back to Fama's work, stock picking by active fund managers does not yield additional returns. Similarly, success stories of active investors can be very misleading because luck can largely skew the assessment of advisors and mutual fund managers (Knutzen, 2012) and consistently skilled fund managers are exceedingly rare in the first place (Damodaran, 2016a). This can make going to an active fund manager a huge gamble for the average person. It is also mathematically impossible for all active investors to beat passive investors or even beat the market. This is simply because for some to win in active investing, others have to lose: the average performance of active investors as a group must be the same as the market average and active investors pay a larger amount in transaction costs (Damodaran, 2016a). Even if active investing can help a single family in isolation, across all families and individuals engaging the market, at least more than half will lose their financial battle (Damodaran, 2016a).

The alternative, passive investing, solves this problem by having everyone follow the market and make money by choosing when and how much to invest. Passive investing is the idea of investing to track a market-weighted index (Asness et al., 2015). In efficient markets, like the stock market (Fama, 1965), passive investment is lower risk and achieves higher average returns than active investment (Knutzen, 2012). However, passive investing is still susceptible to market crashes.

A study of popular advice for portfolio allocation (among cash, bonds and stocks) has found that most recommend a lower ratio of bonds to stocks for an aggressive investor as compared to conservative one (Canner, Mankiw, and Neil, 1997). Financial advisors also make assumptions about inflation, medical costs, and life expectancy that distort their clients' financial goals and hurt the advisor's understanding of the household's portfolio risk (Kotlikoff, 2006). However, separation theorem suggests that all portfolios should contain the same distribution of assets (Kotlikoff, 2006). Furthermore, popular portfolios were significantly different from an optimal portfolio (Kotlikoff, 2006).

Different asset allocation strategies find differing levels of success based on the market trajectory and volatility (Macey, 2010). As such, certain dynamic asset allocation strategies, like ones to increase the amount of stocks or bonds in a certain portfolio as they perform better, work well in prolonged bull or bear markets. However, they lose out to a constant mix strategy in other instances (Macey, 2010). Furthermore, strategies that incorporate both a buy-and-hold strategy and a dynamic hedging strategy are effected less by sharp changes in market behavior like crashes (Liu, Longstaff, and Pan, 01).

This all presents an opportunity to improve the traditional investment model. If it is possible to make accurate long term market predictions based on current under/over-valuation, investors could better balance their assets to minimize the damage of economic downturn and to maximize the returns from rising.

3.2 Review of Machine Learning Techniques

Machine learning (ML) attempts to simulate elements of human thinking to allow researchers to analyze large sets of data and make predictions (Byrnes, 2016). Even though current economic research can give explanations for the short term (Fama) and long-term (Shiller) changes in the market, the medium-term is largely unexplained. Some answers may lie in trends hidden deep in historical data. This section reviews machine learning algorithms that may help to identify medium-term trends in asset prices.

3.3 Support Vector Machines

Support Vector Machines (SVMs) have successfully classified genetic mutations (Sehgal, Gondal, and Dooley, 2004), tracked moving objects (Tian et al., 2007), and improved training methods for image classification (Foody and Mathur, 2004). SVMs are another machine learning algorithm that can perform both regression and classification (Cortes and Vapnik, 1995). They are generally supervised learning algorithms, however, they can also perform unsupervised learning (Ben-Hur, 2008). An SVM creates a hyperplane with the largest margin between two classes in a non-linear space (Cortes and Vapnik, 1995).

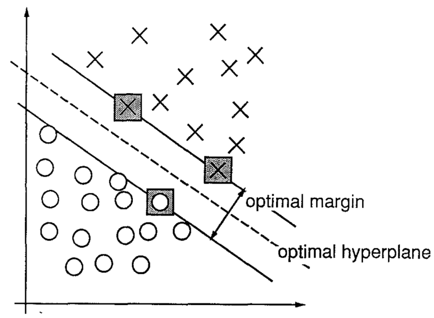

Fig. 1: Example SVM Classification (Cortes and Vapnik, 1995)

Fig. 1 (above), is an example of an SVM solving a 2 dimensional classification problem. Here, each input vector is an O or an X, depending on its classification. The boxed data points are support vectors (Cortes and Vapnik, 1995): support vectors are data points that lie on the edge of the division between classes (Cortes and Vapnik, 1995).

Unfortunately, this idea breaks if the data is not linearly separable, like the XOR logical function (Winston, 2014). This is where the vector function comes into play - it transforms the data into a higher-dimensional space where it is separable by a single hyperplane (Cortes and Vapnik, 1995).

In an SVM, this hyperplane minimizes the expression on top for values constrained by the inequality on the second line in fig. 2 (Cortes and Vapnik, 1995).

$$min_{w,b} \quad \frac{1}{2}\|w\|^2$$

$$where \quad y_i(w \cdot x_i) - b \geq 1$$

and y_i is of the same sign as x_i

Fig. 2: Constraints on an SVM

In *fig. 2*, w is a vector normal to the decision boundary, x_i is a member i of one of the classes, and b is some scalar (Bennet and Campbell, 2000). $\frac{1}{2}\|w\|^2$ is minimized because the combined width of the margins is inversely proportional to $\|w\|$. Thus, maximizing it yields the optimal decision boundary (Winston, 2014). Putting $\|w\|$ into a quadratic form is a mathematical convenience as it sets up a convex learning space (Winston, 2014).

After using Lagrangian multipliers to reconcile the expression to be minimized and its constraints, the maximization of the margin's width depends on the dot product of the support vectors (Winston, 2014). Kernel functions solve the dimensionality problem for this operation by describing the dot product in terms of the original space (Winston, 2014). Thus, maximizing the kernel function maximizes the margins between the support vectors (Steinwart and Christmann, 2008), which make SVMs very powerful tools and allow the efficient use of conventional optimization algorithms for learning.

3.4 Artificial Neural Networks

Artificial neural networks (neural nets) are used to find complex relationships in large datasets (Jain et al., 1996). They have seen success in many applications, such as image and voice recognition, robot control systems, and medical diagnosis (Basu et al., 2010).

Neural nets are analogous to the structure of neurons in the brain (Roberts, 2000): like neurons receive a signal, transform it, and pass it to the next neuron in the chain, each neuron in the net takes the sum of its inputs, applies an activation function to it, and passes the output along synapses (Nielsen, 2015). These synapses have weights that change as the neural net learns (Nielsen, 2015).

The input to a neuron is a collection of all of the "pre-activation" values. Pre-activation values are the products of the weights of the synapse and its input (Ng, no date). The sum of these pre-activation values is passed to an activation function, $f(z)$, which produces the output of the neuron (a).

However, this architecture still needs some mechanism that allows it to “learn”. This is done using an optimization algorithm, such as gradient descent with backward propagation of errors (backprop) (Nielsen, 2015).

3.5 Deep Neural Networks

Deep Neural Networks (DNNs) are feed forward neural nets with multiple hidden layers (Hinton et al., 2012). DNNs are able to use many hidden layers with non-linear activations to more effectively learn abstract relationships among the input features (Hinton et al., 2012). Just like neural nets, DNNs are trained using backprop with relation to different loss functions depending on the learning problem; common loss functions are an l2-loss (for regression) and cross-entropy (for classification) (Hinton et al. 2012).

Like normal neural nets, DNNs have seen much success in handwritten digit recognition (Cireşan et al., 2010), breast cancer diagnosis (Litgens et al., 2016), image classification (Cireşan et al., 2012), and speech recognition (Hinton et al. 2012).

4. Methodology

The methodology is split into two parts that answer three questions for the hypothesis:

1) Using CAPE with a new process to beat popular financial advice:
    a. Risk reduction and returns: Will deviating from a constant mix at high CAPE values increase returns?

b. Addressing CAPE's Non-reversion Problem: If reallocation is done less than annually, delaying the sale of stocks for a rising CAPE and delaying the accumulation of stocks for a dropping CAPE, will returns increase due to the stickiness of the ratio?

2) Can ML fill the gap left between Fama and Shiller's theories by predicting markets in the medium-term (1-2 years)?

4.1 Portfolio Creation Dataset

The data used in the investment model was downloaded from Shiller's website, which has data from 1881 to 2016 (Shiller 2016); more data was generated using these values.

Data downloaded from Professor Shiller's website		
S. No.	Data	Significance/Comments
1	Year	Between 1881 and 2016 (inclusive until December)
2	S&P Stock Composite Stock Market Index	Reflects stock prices movement of 500 large US companies
3	Dividend accruing to the index	Total dividends accrued to the index that year
4	One-year interest rate	One-year treasury bill discount rate. This was taken as a risk free interest rate.
5	CAPE P/E10	Cyclically Adjust Price Earning Ratio: the S&P Index price divided by the average inflation-adjusted earnings for the companies in the index over the last 10 years.
Calculated Data from above data		
6	Stocks Return Rate	The annual rate of return was calculated using the change in the S&P index and the dividends received during the year.
7	CAPE bands	Bands named 0, 1, 2,, etc. and 0,- 1, -2, etc. were formed at each half of a standard deviation as labels for the CAPE values. Half of a standard deviation was used to make the bands more sensitive to changes in the CAPE ratio and its impact on portfolio performance.

Fig. 3: Data Used for Experimentation

Data from Shiller's website and calcualated data							CAPE Bands	
							CAPE Mean	CAPE Std Dev
							16.47	6.48
							Years	136.00
	Beginning	During	End	CAPE 10 Yr	1 Year Interest Rate	S&P Return	No of 1/2 Std Dev	Band No
1919	7.85	0.53	8.83	6.15	5.56	19.24%	-3.18	-3.00
1920	8.83	0.51	7.11	6.14	7.3	-13.70%	-3.19	-3.00
1921	7.11	0.46	7.3	5.29	7.44	9.14%	-3.45	-3.00
1922	7.3	0.51	8.9	6.54	4.58	28.90%	-3.06	-3.00
1923	8.9	0.53	8.83	8.41	4.96	5.17%	-2.49	-2.00
1924	8.83	0.55	10.58	8.19	4.34	26.05%	-2.56	-2.00
1925	10.58	0.6	12.65	9.77	3.87	25.24%	-2.07	-2.00
1926	12.65	0.69	13.4	11.57	4.28	11.38%	-1.51	-1.00
1927	13.4	0.77	17.53	13.90	4.26	36.57%	-0.79	0.00
1928	17.53	0.85	24.86	19.35	4.64	46.66%	0.89	0.00
1929	24.86	0.97	21.71	26.70	6.01	-8.77%	3.16	3.00
1930	21.71	0.98	15.98	21.50	4.15	-21.88%	1.55	1.00
1931	15.98	0.82	8.3	16.49	2.43	-42.93%	0.01	0.00
1932	8.3	0.5	7.09	9.14	3.36	-8.55%	-2.26	-2.00
1933	7.09	0.44	10.54	8.79	1.46	54.87%	-2.37	-2.00
1934	10.54	0.45	9.26	13.26	1.01	-7.87%	-0.99	0.00
1935	9.26	0.47	13.76	11.65	0.75	53.67%	-1.49	-1.00
1936	13.76	0.72	17.59	17.50	0.75	33.07%	0.32	0.00
1937	17.59	0.8	11.31	21.84	0.88	-31.15%	1.66	1.00
1938	11.31	0.51	12.5	13.58	0.88	15.03%	-0.89	0.00
1939	12.5	0.62	12.3	16.18	0.56	3.36%	-0.09	0.00

Fig. 4: Data in the Excel Sheet

4.2 Portfolio Creation and Testing

Various portfolio allocations were tried based on the CAPE band. The portfolio was allocated between the S&P Index and one-year treasury bills. The treasury bills were taken as risk free securities to isolate the impact of changing stock allocations on the portfolio. Long-term bonds were not used because their returns can swing based on changes in interest rates and duration to maturity. In total, 30 scenarios were tried and the following scenarios were found to be worth noting. It was assumed that \$100 were invested in 1881 using above mentioned scenarios and rate of annual return was calculated based on the portfolio value in December 2016.

No.	Risk Tolerance	Strategy
1	Aggressive	Constant Mix: 100% S&P Index, 0% Risk free debt
2	Aggressive	Reduce stocks at high CAPE, review allocation annually: 100% S&P Index, 0% Risk free debt: below CAPE threshold* Decrease percentage of stocks in increments, while increase investments in risk free debt as CAPE increases
3	Aggressive	Reducing stocks at high CAPE, review allocation every *3 years*: 100% S&P Index, 0% Risk free debt: below CAPE threshold* 0% S&P Index, 100% Risk free debt: above CAPE threshold
4	Moderate	Constant Mix: 60% S&P Index, 40% Risk free debt
5	Moderate	Exit stocks at high CAPE, review allocation annually: 60% S&P Index, 40% Risk free debt: below CAPE threshold* 0% S&P Index, 100% Risk free debt: above CAPE threshold
6	Moderate	Exit stocks at high CAPE, review allocation *every 3 years*: 60% S&P Index, 0% Risk free debt: below CAPE threshold* 0% S&P Index, 100% Risk free debt: above CAPE threshold
7	Variable Mix	Increase stock allocation from high to low based on low to high CAPE, review allocation annually
8	Variable	Increase stock allocation from high to low based on low to high CAPE, review allocation every 3 years

Fig. 5: Key Portfolios Created during Experimentation

---

* CAPE threshold to be found during analysis

Robo Allocation 60-40

Band No	Allocation
-5	60%
-4	60%
-3	60%
-2	60%
-1	60%
0	60%
1	60%
2	60%
3	60%
4	60%
5	60%
6	60%
7	60%
8	60%
9	60%
Reallocation Frequency	1
Initial inv.	100
Final Value	1,749,954
Rate of return	7.45%
Rate of return diff	0

Stocks allocation	Beginning Stocks	Beginning Debt	Ending Stocks	Ending Debt	Total
60%					100.00
60%	60.00	40.00	60.48	41.92	102.40
60%	61.44	40.96	63.62	43.11	106.73
60%	64.04	42.69	60.73	44.98	105.71
60%	63.43	42.28	55.71	44.67	100.39
60%	60.23	40.15	77.28	41.85	119.13
60%	71.48	47.65	79.72	49.68	129.41
60%	77.64	51.76	77.37	54.92	132.29
60%	79.37	52.92	81.77	55.57	137.34

Fig. 6: Investment Model Data Sheet

Excel was used to create a table that directly referenced the data and applied the investment rule for each strategy to return the final balance and rate of return. As seen in red in fig. 6 (above), the investment rule was created by assigning a percent investment in stocks for a given band number. The frequency of reallocation or the re-application of the investment rule was done using the cell highlighted in yellow. Rebalancing based on the allocation percentage was done annually in every scenario. Rebalancing is the readjustment of money in stocks and treasury bills to match the desired percentages.

Next is the process of determining the final balance and rate of return.

First, the stocks allocation was found based on the CAPE band. Then, the beginning investment into stocks was calculated by multiplying amount from Total column (ending value of previous year) by the stocks allocation and populating that value in the Beginning Stocks column. The remaining amount was populated into Beginning Debt column, which tracked the money invested in treasury bills. The Ending Stocks column was populated by applying the S&P Return rate for the year and the Ending Debt column was similarly populated by applying interest rate for the year using 1 Year Interest Rate column. The final balance of the portfolio was found by adding the Ending Stocks and Ending Debt and was populated into Total column. The final balance was found by taking the sum of the total column through the time period.

The balance is then used to find the annual rate of return using Excel's Seek function by setting a goal of zero for the difference between final portfolio value and the guess return rate created by using a guess return rate.

4.3 Machine Learning Dataset

The data for the three machine learning algorithms was selected from the large set of economic indicators in the St. Louis Federal Bank Federal Research Economic Dataset (FRED). These indicators included data such as the treasury constant maturity rates, the number of employees in different economic sectors like farming and manufacturing, housing prices, and consumer purchasing power (there were a total of 145 features used from the dataset). The data also provided the NYSE closing price and was updated for each day since 1965. The target label for the data asked the model to predict the closing price of the NYSE 365 days in advance.

As part of preprocessing, all NaN values were converted to 0 and the data was normalized. Then it was put through Sci Kit Learn's MinMax preprocessing algorithm and shuffled before given to the algorithms for training and testing. The models were tested over less predictable examples like the crash after 9/11 and the Great Recession. In addition, it was attempted to use data that had been expanded to include quadratic terms using the product of different combinations of the variables in the original dataset. However, that dataset reduced the models' accuracy and was thus dropped.

4.4 Nu Support Vector Regression Model Construction

The support vector machine used in this project was implemented in SciKit Learn (SKLearn). SKLearn is a machine learning API that facilitates the implementation of machine learning algorithms. Out of the three Support Vector Regression (SVR) techniques supported by SKLearn, Nu Support Vector Regression (NuSVR) was chosen as it performed best in initial testing.

For the NuSVR model, the final (6^{th} out of 7) version used the "poly" kernel function. The other options were also tried, but they did not perform as well in testing. The value for gamma, which influences how much the model overfits (Cortes and Vapnik, 1995), was also changed and proved to have a significant impact on the model.

NuSVR used a hinge loss and nu regression instead of epsilon regression. Epsilon regression uses ε as the size of the margin within which all data points should be contained to construct the loss function (Smola and Schölkopf, 2004). Instead, nu regression uses ν as an upper bound on the fraction of training examples outside the margin and ν as a lower bound on the number of support vectors (Chang and Lin, 2009).

4.5 Deep Neural Net Construction

The first DNN had 8 hidden layers with 30 units per layer and an exponential linear unit (ELU) activation function for all of the layers. ELU alleviates the bias shift problem of the more popular ReLU (rectified linear unit) while retaining ReLU's ability to avoid the vanishing gradient problem (Clevert, Unterthiner, and Hochreiter, 2016). In addition, ELU is a zero-centered function, which has been shown to speed up learning in neural nets (LeCun, Kanter, and Solla, 1991).

Unfortunately, this setup did not seem to be learning well. Thus, over the next ten variations of the network, the neural net developed an architecture with 10 hidden layers and used combinations of ELU, ReLU, TanH, Softmax, Softplus, and Sigmoid as the activation functions in the network. Lastly, the final neural net used gradient descent to optimize the parameters. Even though gradient descent is slower, as discussed earlier, it was more accurate than Adagrad and Adam.

The varied mix of activations was not expected to change performance very much, however, it improved training accuracy by nearly 30%. This is an area for further investigation, but it is speculated that the introduction of more non-linearities encouraged learning of different abstractions from the data – akin to how convolutional neural nets extract new levels of abstract data from the input features in each layer (Montavon, Braun, and Müller, 2011). Tensorflow, which is an open source machine learning library released by Google, was used to construct and train the neural net as it allows for fast and highly parallelizable execution of machine learning tasks (Abadi et al., 2015).

4.6 Linear Regression Modeling

The linear regression model was of the form of $y = x \cdot m + b$. As x is a vector quantity, m was in the form of a matrix with dimensions $[f, o]$ where f is the number of features in x and o is the number of outputs contained in y (1 in this case). The weight matrix m and the bias value b were optimized using gradient descent and an l2 loss.

5. Results

5.1 Investment Management Using CAPE and Less Frequent Reallocation

It was determined that constant mix portfolios can decrease risk and improve returns by dynamically adjusting their exposure to stocks. Lowering the exposure when the CAPE value is between 1 and 2 standard deviations above the mean and dropping the exposure to zero past 2 standard deviations above the mean can accomplish this. Reevaluating the stocks allocation every 3 years further improves the results in the

CAPE Std Dev. Band	Aggressive - Baseline	Aggressive with Downside Management using CAPE	Aggressive with Downside Management using CAPE and Rebalancing every 3 years	Robo Allocation 60-40	Robo Allocation 60-40 With Downside Management Using CAPE	Robo Allocation 60-40 With Rebalancing every 3 years	Fluid Allocation Based on CAPE	Fluid Allocation Based on CAPE with 3 year Rebalance
-5	100%	100%	100%	60%	60%	60%	100%	100%
-4	100%	100%	100%	60%	60%	60%	100%	100%
-3	100%	100%	100%	60%	60%	60%	100%	100%
-2	100%	100%	100%	60%	60%	60%	100%	100%
-1	100%	100%	100%	60%	60%	60%	80%	80%
0	100%	100%	100%	60%	60%	60%	80%	80%
1	100%	100%	100%	60%	60%	60%	60%	60%
2	100%	60%	100%	60%	60%	60%	60%	60%
3	100%	20%	0%	60%	0%	0%	40%	40%
4	100%	20%	0%	60%	0%	0%	40%	40%
5	100%	0%	0%	60%	0%	0%	20%	20%
6	100%	0%	0%	60%	0%	0%	20%	20%
7	100%	0%	0%	60%	0%	0%	0%	0%
8	100%	0%	0%	60%	0%	0%	0%	0%
9	100%	0%	0%	60%	0%	0%	0%	0%
Rebalance Freq.	1	1	3	1	1	3	1	3
Initial inv.	100	100	100	100	100	100	100	100
Final Value	9,781,214	13,031,292	29,525,719	1,749,954	1,986,144	3,016,129	6,495,777	5,913,730
Rate of return	8.82%	9.05%	9.70%	7.45%	7.55%	7.88%	8.49%	8.41%

Fig. 7: Results of Passive Investing Guided by CAPE

aggressive portfolio and the moderate portfolio. This indicates that delaying reallocation based on CAPE delays selling early in low markets (high CAPE) and buying early in high markets (low CAPE). Fig. 7 (previous page) is a table of the 8 scenarios, their final balances, and rates of return.

Adding to this, it makes sense that increasing reallocation time does not work in the fluid case. There, the allocation is conservatively adapted to the CAPE ratio anyway: the higher offset of changing the stocks allocation to lower/higher values automatically delays the portfolio's ability to adapt to CAPE. That means it actually adapts to the market at the right time anyways. When three extra years are added, it results in the allocation changing too late, which means stocks are still bought and sold at non-optimal times. In other words, stocks are now sold after the top of growth and bought at after the

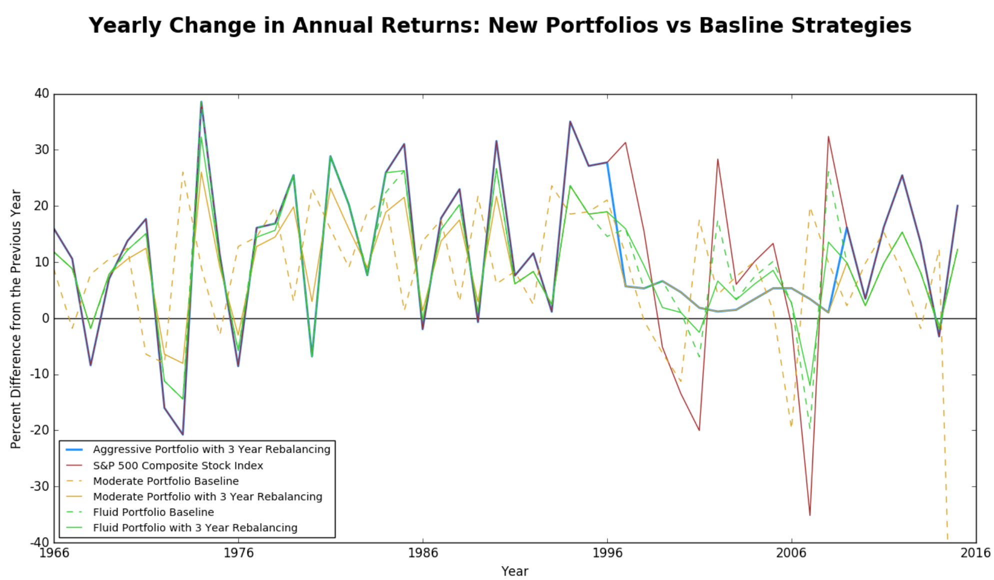

Fig. 8: Comparison of Aggressive, Moderate, and Fluid 3-year portfolios and their respective baselines.

bottom of a dip. Above, in fig. 8, notice how the moderate baseline trails behind the rest and how the fluid 3-year isn't able to adjust the magnitude of its investment fast enough to avoid dips completely and to maximize on returns from upticks. This proves what was inferred from the numerical data above.

5.2 Nu Support Vector Regression

NuSVR was the most successful ML algorithm. It originally did not scale well to the testing data, where it experienced sharp drops in R^2 values. However, since Fama's work indicates that markets are a random walk in the short term, this outcome may have been due to high variance in the market price. Changing the value for gamma in the model solved the problem. Paradoxically, increasing the bias of the model allowed it to better account for the variance encountered when predicting over the testing set. Setting gamma to 0.0523125 balanced the percent accuracy of the model with the R^2 values best.

		TRAINING					TESTING					
		Explained Variance	Mean Absolute Error	MSE	Median Absolute Error	R^2	Explained Variance	Mean Absolute Error	MSE	Median Absolute Error	R^2	AVG % Error
NEXT YEAR	Gamma = 0.1	0.995529	0.010477	0	0.007171	0.995520	0.507843	0.119815	0	0.085566	0.481901	0.210520
	Gamma = 0.055	0.993635	0.012832	0	0.008907	0.993625	0.594467	0.109324	0	0.081508	0.582926	0.192087
	Gamma = 0.0525	0.993418	0.013080	0	0.993408	0.993408	0.597303	0.109123	0	0.081193	0.586314	0.191733
	Gamma = 0.0523125	0.993400	0.013100	0	0.009138	0.993391	0.597290	0.081391	0	0.081391	0.586520	0.191810
	Gamma = 0.051875	0.993360	0.013154	0	0.009151	0.993351	0.597310	0.109238	0	0.081489	0.586747	0.191934
	Gamma = 0.05125	0.993293	0.013230	0	0.009207	0.993284	0.596513	0.109376	0	0.081687	0.585956	0.192176
	Gamma = 0.05	0.993160	0.013364	0	0.009306	0.993151	0.595972	0.109635	0	0.082535	0.585755	0.192632
	Gamma = 0.045	0.992498	0.014022	0	0.009662	0.992492	0.601968	0.109684	0	0.085117	0.594129	0.192719
	Gamma = 0.04125	0.991844	0.014602	0	0.009986	0.991842	0.602633	0.110452	0	0.085766	0.596423	0.194068
	Gamma = 0.04	0.991593	0.014812	0	0.010101	0.991591	0.601553	0.110557	0	0.085702	0.594979	0.194253
	Gamma = 0.007 (Default)	0.950572	0.038187	0	0.027588	0.950570	0.444131	0.150017	0	0.087782	0.145434	0.263585

Fig. 9: Training and Testing Performance Metrics

Fig. 10: Normalized NYSE Closing Price predictions for the recession. Smoothed using a 60-day moving average.

However, upon closer inspection, it was found that such close similarity was found only in the normalized graph but not on the rescaled graph. This seems to be because the SVM cannot extract medium-term changes in terms of the absolute magnitude of the prices. It instead follows the general movement of the market along an approximate average value for the time period. This indicates that the SVM may actually be a better predictor of long-term trends in the market, like CAPE. In addition, this indicates that the economic environment at any given time may create movement in the market or at least be a significant factor in it. However, the latest information about those companies and human sentiments determine the magnitude of those changes.

5.3 Deep Neural Net

The DNN averaged $0.12 below the true value on the training data but it did not generalize well and made predictions $6702.76 below the true value on average upon testing. It was found that the DNN guessed the same value for all training data to minimize the average error; that did not work on the testing set, which had new data.

5.4 Linear Regression

The linear regression was predicted prices $3545.68 lower than the true value and $10242.77 lower on the test data, on average. Cubic regression performed even worse.

6. Conclusion

An aggressive investment strategy that performed downturn management guided by CAPE and rebalanced every 3 years was shown to be a very successful combination of passive investing and dynamic asset allocation. This strategy, as well as both of the moderate and baseline fluid strategies, can be used to help financial advisors fulfill the long-term financial goals of an average family. Furthermore, the new portfolios allow advisors to stay conducive to the risk tolerance of a specific family by giving them 3 choices that should roughly match their situation. Even if there are only 3 portfolios, the process used in this paper could be used by advisors to adjust one of them to the needs of a specific family. In addition, all of the new portfolios mitigate the effects of a crash and both the aggressive and moderate 3-year strategies avoided the Great Recession and the crash following the 2000 Tech Bubble.

It was also determined that one-year forecasts created by NuSVR can be used as a guiding heuristic to generally validate long terms predictions made with CAPE. However, it is possible to use new methods of rescaling the predictions to make changes in magnitude more evident. Unfortunately, that is outside the scope of this project.

Thus, the first part of the hypothesis was confirmed by creating three new investment portfolios that were able to significantly beat current strategies used by financial advisors. However, the second part of the hypothesis regarding machine learning was proven false. Machine learning can give investors average values over a long period of time and give indications of market movement; however, it cannot approximate the magnitude of these changes nearly close enough to be called a useful predictor for the medium-term.

Works Cited

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X. and Google Research (2015) 'TensorFlow: Large-scale machine learning on heterogeneous distributed systems', .

Asness, C.S., Frazzini, A., Israel, R. and Moskowitz, T.J. (2015) 'Fact, fiction, and value investing by Clifford S. Asness, Andrea Frazzini, Ronen Israel, Tobias J. Moskowitz: SSRN', Journal of Portfolio Management, . doi: 10.2139/ssrn.2595747.

Basu, J.K., Bhattacharyya, D. and Kim, T.H. (2010) 'The Use of Artificial Neural Networks in Pattern Recognition', International Journal of Software Engineering and Its Applications, 4(2).

Ben-Hur, A. (2008) 'Support vector clustering', Scholarpedia, 3(6), p. 5187. doi: 10.4249/scholarpedia.5187.

Bennett, K.P. and Campbell, C. (2000) 'Support Vector Machines: Hype or Hallelujah?', SIGKDD Explorations, 2(2), pp. 2–13.

Canner, Niko, N. Gregory Mankiw, and David Weil, 1997, An asset allocation puzzle, American Economic Review 87, 181—191.

Case, K.E., Quigley, J.H. and Shiller, R.J. (2013) 'Wealth Effects Revisited 1975-2012', Critical Finance Review, 2(1), pp. 101–128. doi: 10.1561/104.00000009.

Chang, C.-C. and Lin, C.-J. (2009) Training ν-Support Vector Regression: Theory and Algorithms, 22 July. .

Cires ̧an, D., Meier, U. and Schmidhuber, J.̈rgen (2012) Multi-column Deep Neural Networks for Image Classification. Available at: https://arxiv.org/pdf/1202.2745.pdf (Accessed: 26 December 2016).

Cireşan, D., Meier, U., Gambardella, L. and Schmidhuber, J. (2010) 'Deep, big, simple neural nets for handwritten digit recognition', Neural computation., 22(12), pp. 3207–20.

Clevert, D.-A., Unterthiner, T. and Hochreiter, S. (2016) 'FAST AND ACCURATE DEEP NETWORK LEARNING BY EXPONENTIAL LINEAR UNITS (ELUS)', ICLR: . Available at: https://arxiv.org/pdf/1511.07289.pdf (Accessed: 27 December 2016).

Cortes, C. and Vapnik, V. (1995) 'Support-vector networks', Machine Learning, 20(3), pp. 273–297. doi: 10.1007/bf00994018.

Damodaran, A. (2016a) 'Active investing: Rest in peace or resurgent force?', Musings on Markets, 14 December. Available at: http://aswathdamodaran.blogspot.com/2016/12/active-investing-rest-in-peace-or.html#uds-search-results (Accessed: 30 December 2016).

Damodaran, A. (2016b) 'Mean Reversion: Gravitational super force or dangerous delusion?', Musings on Markets, Available at: http://aswathdamodaran.blogspot.fr/2016/08/mean-reversion-gravitational-super.html (Accessed: 30 December 2016).

The Economic Policy Institute (2012) The State of Working America. 12th Edition edn. Cornell University Press.

Elwell, C.K. (2013) CRS report for congress rebuilding household wealth: Implications for economic recovery. Available at: https://fas.org/sgp/crs/misc/R43228.pdf (Accessed: 24 December 2016).

Englund, P. and The Royal Swedish Academy of Sciences (2013) ‘Scientific background: Understanding asset prices’, Sweden: The Royal Swedish Academy of Sciences. Available at: https://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2013/advanced-economicsciences2013.pdf (Accessed: 4 February 2017).

Fama, E.F. (1965) ‘Random walks in stock market prices’, Financial Analysts Journal, 21(5), pp. 55–59. doi: 10.2469/faj.v21.n5.55.

Foody, G.M. and Mathur, A. (2004) ‘Toward intelligent training of supervised image classifications: Directing training data acquisition for SVM classification’, Remote Sensing of Environment, 93(s 1–2), pp. 107–117. doi: 10.1016/j.rse.2004.06.017.

Grennon, T. (2015) A Dynamic Asset Allocation Approach to Investing. Available at: https://papers.ssrn.com/sol3/Papers.cfm?abstract_id=2638805 (Accessed: 30 December 2016).

Guillemette, M., Finke, M.S. and Gilliam, J. (2012) ‘Which types of risk tolerance questions should be used when determining a client’s portfolio allocation preference?’, Journal of Financial Planning, . doi: 10.2139/ssrn.2009419.

Guiso, L., Sapienza, P. and Zingales, L. (2008) ‘Trusting the stock market’, Journal of

Finance, LXIII(6), pp. 2557–2600. doi: 10.2139/ssrn.811545.

Hansen, L.P. (2001) ‘Generalized method of moments estimation’, in International Encyclopedia of Social and Behavioral Sciences. Available at: http://larspeterhansen.org/wp-content/uploads/2016/10/Generalized-Method-of-Moments-Estimation-A-Time-Series-Perspective-published-title-22Method-of-Moments22.pdf (Accessed: 28 December 2016).

Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A.-R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T. and Kingsbury, B. (2012) ‘Deep neural networks for acoustic modeling in speech recognition’, .

Jain, A.K., Mao, J. and Mohiuddin, K.M. (1996) ‘Artificial neural networks: A tutorial’, Computer, 29(3), pp. 31–44. doi: 10.1109/2.485891.

Kantor, B. and Holdsworth, C. (2014) ‘2013 Nobel prize revisited: Do Shiller’s models really have predictive power?’, Journal of Applied Corporate Finance, 26(2), pp. 101–108. doi: 10.1111/jacf.12072.

Knutzen, E. (2012) ASSESSING ACTIVE VS. PASSIVE STRATEGIES IN THE CURRENT ENVIRONMENT CONSIDERING THE PAST TO LOOK FORWARD. NEPC.

Kotlikoff, L.J. (2008) ‘Is conventional financial planning good for your financial health?’, CFA Institute, .

Lahart, J. (2016) Shiller’s powerful market indicator is sending a false signal about stocks this time. Available at: https://www.wsj.com/articles/powerful-market-indicator-flashes-sell-heres-why-it-can-be-ignored-1475687806 (Accessed: 4 February 2017).

LeCun, Y., Kanter, I. and Solla, S.A. (1991) 'Eigenvalues of covariance matrices: Application to neural-network learning.', Physical Review Letters, 66(18), pp. 2396–2399.

Leigh, D., Igan, D., Simon, J. and Topalova, P. (2012) 'Chapter 3: Dealing with household debt', in World Economic Outlook (WEO): Growth Resuming, Dangers Remain. International Monetary Fund, pp. 89–124.

Litgens, G., Sánchez, C.I., Timofeeva, N., Hermsen, M., Nagtegaal, I., Kovacs, I., Hulsbergen- van de Kaa, C., Bult, P., van Ginneken, B. and van der Laak, J. (2016) 'Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis', Scientific Reports, 6. doi: 10.1038/srep26286.

Liu, J., Longstaff, F. and Pan, J. (2001) 'Dynamic asset allocation with event risk', The Journal of Finance, 58(1), pp. 231–259.

Macey, R.P. (2010) Dynamic Asset Allocation. Available at: http://redtortoisellc.com/wp-content/uploads/2016/03/iwm10janfeb_dynamicaa.pdf (Accessed: 30 December 2016).

MIT OpenCourseWare and Winston, P. (2014) 16. Learning: Support vector machines. Available at: https://www.youtube.com/watch?v=_PwhiWxHK8o (Accessed: 27 December 2016).

Montavon, G., Braun, M.L. and Müller, K.-R. (2011) 'Kernel analysis of deep networks', Journal of Machine Learning Research, 12, pp. 2563–2581.

Ng, A. (no date) Lecture Notes #1 CS 229. Available at: http://cs229.stanford.edu/notes/cs229-notes1.ps (Accessed: 24 October 2016).

Nielsen, M.A. (2015) 'Using neural nets to recognize handwritten digits', in Neural

Networks and Deep Learning. Determination Press, .

Nobel Media AB (2014) Nobelprize.Org. Available at: https://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2013/hansen-facts.html (Accessed: 28 December 2016).

Roberts, E. (2000) Neural networks - neuron. Available at: https://cs.stanford.edu/people/eroberts/courses/soco/projects/neural-networks/Neuron/index.html (Accessed: 30 December 2016).

van Rooij, M., Lusardi, A. and Alessie, R. (2011) 'Financial literacy and stock market participation', Journal of Financial Economics, 101(2), pp. 449–472. doi: 10.1016/j.jfineco.2011.03.006.

Sehgal, M.S.B., Gondal, I. and Dooley, L. (2004) 'Statistical neural networks and support vector machine for the classification of genetic mutations in ovarian cancer - IEEE Xplore document', Computational Intelligence in Bioinformatics and Computational Biology, .

Shiller, R.J. (2000) Irrational exuberance. Princeton, NJ, United States: Princeton University Press.

Shiller, R.J. (2016) 'Online data - Robert Shiller'. Available at: http://www.econ.yale.edu/~shiller/data.htm (Accessed: 8 December 2016).

Smola, A.J. and Schölkopf, B. (2004) Statistics and Computing. Available at: http://www.svms.org/regression/SmSc98.pdf (Accessed: 28 December 2016).

Steinwart, I. and Christmann, A. (2008) Support vector machines. New York: Springer-Verlag New York.

Tanner, E. and Abdih, Y. (2009) 'Rebuilding U.S. Wealth', Finance and development, 46(4).

Tian, M., Zhang, W. and Liu, F. (no date) 'On-line ensemble SVM for robust object tracking', Computer Vision – ACCV 2007, 4843, pp. 355–364. doi: 10.1007/978-3-540-76386-4_33.

Vornovitsky, M., Gottschalck, A. and Smith, A. (2012) Distribution of Household Wealth in the U.S.: 2000 to 2011. Available at: https://www.census.gov/people/wealth/files/Wealth%20distribution%202000%20to%202011.pdf (Accessed: 24 October 2016).

Wang, K.Q. (2011) 'Buy High and Sell Low', International Journal of Pension Management, . doi: 10.2139/ssrn.1099964.