TRADING STRATEGIES BASED ON PREDICTING PRICES OF FUTURES CONTRACTS

Jan MROZOWSKI

Promotor: Prof. zw. dr hab. **Marian NOGA**

INTRODUCTION

Derivative markets for commodities and currencies play vital part in modern day international financial world by allowing market participants to manage their risk or engage in trading activities to generate or maximize profits. Futures contracts for high value metals such as Gold are used extensively by organizations that Gold in their manufacturing processes as well as financial institutions to manage their risk due to well established inverse correlation between market volatility and price of the metal.

Recent developments in machine learning algorithms gained wide adoption in some industries but their applicability to trading of derivatives have not been extensively covered in academic literature.

Building statistical models to predict price of a commodity and use it in trading is typically limited to trading houses that keep most of their research in-house as means of competitive advantage. Peer reviewed literature in this area is scarce as well.

A book on derivatives trading, *Money Management Strategies for Futures Traders* by Nauzer Balsara [1], states "A buy signal is generated when the shorter of two moving averages exceeds the longer one; a sell signal is generated when the shorter moving average falls below the longer moving average." It follows with "Armed with this information, the trader can estimate a cutoff value, beyond which it is highly unlikely that the unrealized loss will be recouped, and the trade will end profitably."

Some researchers [2] experimented with predicting prices of oil futures using Convolutional Neural Network (CNN). Their research concluded CNN produces better results than traditional economic models, but not accurate enough to be useful. That is due to complex nonlinear characteristics of the relationships coming from "many complex natural, economic, and political factors". No similar research related to Gold Futures was found.

All non-academic sources for traders suggest using maximum allowable loss from the trade when betting on a new trend. There is however no concept of "minimum target profit" or minimum duration of time that the contracts are targeted to be held. The author suggests that establishing these new metrics and calculating statistical probability of profit happening can enhance trading strategy by foregoing some of the trades when signals happen. or allocating smaller amount of capital to them or exploring low risk strategies such as calendar spreads.

List of Keywords:

Futures contracts, Machine Learning, Neural Networks, Autoencoders, Bayesian Structured Time Series, Short-term Trend Forecasting, Futures Prices Forecasting.

---

[1] Balsara, Nauzer. (1992). Money Management Strategies for Futures Traders. New York: John Wiley & Sons, Inc.
[2] Chew, D. H. *Corporate Risk Management*. Columbia University Press, 2008. p. 23

OBJECTIVE

Accurately predicting the price of Gold Futures, one of the world's most actively trading commodities, has always been important for academics and traders. The author reviews development of derivatives markets and trading strategies as well as statistical techniques traditionally used in forecasting. A number of well-known models such as Black-Scholes failed traders and markets during financial crises of 2009 so development and application of new approaches is critical in order for users of derivatives to have confidence in prices and settlement procedures.

This research aims to answer the question "Can traders utilize the data created from historic prices to improve profitability of bets on upcoming upward trend?" The analysis builds machine-learning models that can be applied to future trades. It tests the predictive power of these models and created variables. Models can be tested on other Futures contracts and adjusted accordingly in order to diversify trading and increase volume of trades. The hypothesis is that created data can help to identify 90% of successful trades with accuracy over 50%.

The author evaluates performance of machine learning algorithms, when trying to identify formation of a new upward trend in price of Gold Futures at the very beginning. Application of machine learning methodologies shows scarcity of linear relationship between historic prices and new trend development. Black box models such as Neural Networks and specifically Autoencoders allow traders and analysts to classify observations in a way that can be used by entities engaged in trading of gold futures.

RESEARCH AREA

In the sector of finances, a futures contract, also called simply futures, is a type of a forward contract that's been standardized as a legal agreement to purchase or sell an item at a previously determined price (the forwards price) of purchase and at a defined time in the future (delivery date). Futures contracts carry out transactions of assets that are usually commodities or instruments used in finances[3].

Everything connected with futures contracts is negotiated at special exchanges called futures exchanges that act as marketplaces for sellers and purchasers. The latter is considered a long position holder and the former a short position holder. There is, however, a risk that both parties of the agreement may decide to terminate it or simply walk away if the negotiated prices are not favourable to them. Therefore, it is possible for the parties to lodge a margin of the contract value with a neutral third party. For instance in the gold futures trading, the margins is between two per cent and twenty per cent[4].

As for its origins, the beginning of futures contracts can be traced back to 1972, when they were mostly used to negotiate agricultural commodities. Later on, they were mostly applied to transactions that concerned natural resources like oil. Over time, this type of contracts has developed and now we can come across such terms as

---

[3] Chew, D. H. *Corporate Risk Management*. Columbia University Press, 2008. p. 23

[4] Valdez, S., *An Introduction To Global Financial Markets* (3rd ed.). Basingstoke: Macmillan Press, 2000. p. 34-36

currency futures, interest rate futures and stock market index futures, which play a major role in the overall futures market[5].

Initially, the main purpose of futures contracts was to mitigate the risk associated with price and exchange rate movements through letting the parties fix prices or rate transactions, which were to be finalized at a later time, in advance. It came in handy when parties expected payments in advance, which came in foreign currencies.

Strategies for Futures Trading

From a wide variety of strategies covered in literature and described in the paper 2 were shortlisted for further exploration to be used in conjunction with proposed models:

- Swing Trading - with this approach, the investor deliberately leaves transactions open on the account for a period of more than a day, sometimes much longer (sometimes even several weeks). Swing Trading assumes using "swings", i.e. clearly marked sections on the chart. In principle, it is rather used with higher intervals such as daily or weekly. Others are capital management principles, Stop Loss and Take Profit methods.
- Calendar Spread - it is also important to define the Calendar Spread method in detail typically used on the Futures markets. It is also known as intracontract, intracommodity, intermonth or time spread as it involves entering into same number of opposite positions expiring in different months.

PROPOSED SOLUTIONS

Proposed Strategy

The common strategy for entering to long positions on gold has been developed by studying relationships between moving averages of different lengths prior the development of new trends. When the 4-day moving average crosses 9-day average on the way up, a new positive anomaly may develop in the next couple of days.

Trades are entered into at the market price following the crossover with a view of holding the contract for 6 or more days if upward trend develops. The minimum profit target is measured by averaging differences between high and low daily prices over the last 9 days. The same measure is used for maximum loss from the trade and the position is exited in the 6 day window if needed. This approach is moderately profitable as in ~30% of instances new upward trend materializes with average profit significantly higher than the average loss.

Research Objective

The author would like to improve upon this approach and develop a probabilistic model that can be used as a guide to risk of proposed trades at the time of entering positions. The model should include available historic price, volume and open interest trend data which can be derived from available variables. They are momentums of

[5] Chew, Donald H. *op. cit.*, 2008, p 26

the variables for various lengths of time, ratios of current price to maximums and minimums of time periods and other technical indicators used by traders. The author would like to explore the predictive power of these derivative variables and develop methodologies that will incorporate the data in the decision management process.

The purpose is to determine whether derivative historic variables can better predict a new upward trend in Gold Futures and improve the profitability of these investments. The findings from this research will help traders understand the predictive power of derivative variables and help with decisions about entering trades and allocating capital. The model features will be engineered from available Gold Futures data. This research will explore the relationship between historic prices and future prices on dates with preset criteria.

Variables and Scope

For this study, the author/we only considered dates when crossover of 4 and 9-day moving averages crossed. Prices from 6 days after crossover are used to classify which trades achieved minimum profitability target. If price is above target then dependent variable equals 1, else it equals 0.

Available daily price data since 1975 was extracted from Barchart.com using API:

- Open Price
- High Price
- Low Price
- Close Price
- Volume
- Open Interest

Data features were engineered for independent variables by calculating various types of momentums available variables and ratios of last known prices to minimums and maximums of different lengths of time. Below are groups of the feature-engineered, independent variables:

- Rate of change (momentum) for all available variables using 1,4,9,15,30 and 60 days and its moving average using 4,9,15,30 and 60 days
- Rate of change (momentum) for close price weighted by Open Interest, Volume and both. Its moving average using 4,9,15,30 and 60 days
- Ratio of Close price to Minimums and Maximums of Close price in the last 4,9,15,30, 60,90,180 and 360 days
- Ratio of Close price to Bollinger Bands and its moving averages for 4,9,15,30, 60 and 90 days
- Daily difference between price of Futures contract and physical gold and its moving averages for 4,9,15,30, 60 and 90 days

Modeling Methods

Firstly, a binary distribution is explored to investigate if any of machine learning algorithms can produce a result suitable for usage with proposed strategy. Methodologies included:

- Logistic Regression
- Decision Trees
- Random Forest
- Gradient Booster Classifier
- Neural Networks

Secondly, time-series approach, using structural time series model is defined by two equations. The observation equation relates the observed data y_t to a vector of latent variables α_t known as the "state."

$$y_t = Z_t^T \alpha_t + E_t.$$

The transition equation describes how the latent state evolves through time.

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t.$$

The error terms E_t and η_t are Gaussian and independent of everything else. The arrays Z_t , T_t and R_t are structural parameters. This approach has been considered due to availability of well-developed set of MCMC algorithms for doing Bayesian inference with time-varying paraments (BSTS R package).

MODELING

Analysis Framework

This research follows the Knowledge Discovery and Data Mining (KDD) structure, illustrated below.

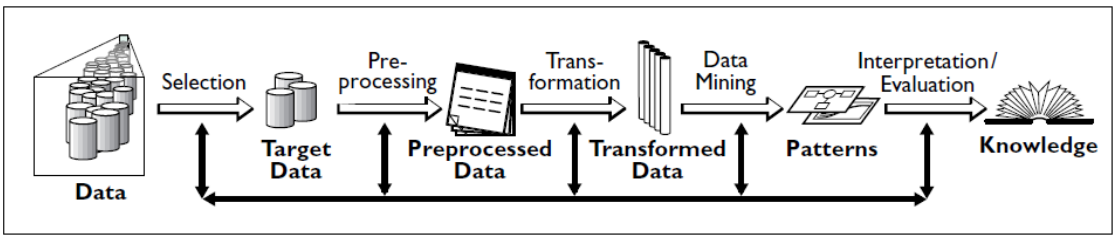

Figure 1. KDD Framework

The framework comprises finding the data source, pre-processing and transforming the data, followed by descriptive analysis and modeling, and finally interpreting the outcomes. The best outcomes lead to knowledge which can be acted upon. All steps are described in the following sections.

Data Description

Barchart Inc. has been identified as potential data source. Their free API service was used for downloading of the initial data set. It contained over 11,000 daily observations of Gold Futures prices as well as Open Interest and Volume of trading by day. Each day had 4 price points: open, low, high and close. The data was downloaded in structured spreadsheets format using R interface. Spot prices for Gold were downloaded separately from the same source and merged to initial dataset.

Development of Dependent Variable

The application of supervised machine learning algorithms requires for each observation to have a dependent variable as outcome measure and independent variables as predictors. Daily price data was used to create both dependent and independent variables. Moving averages of Close price have been calculated using 9 and 4-day periods, followed by calculation of difference between them using shorter average as basis. All observations when the difference turned positive from being negative the day before were identified and Signal date. A measure of volatility was selected to be the average difference between high and low daily prices for last 9 days, including Signal date. This measure became a minimum profit and maximum loss targets for the trades and was added to Close price on each signal date. Target price was calculated by adding Minimum profit target to Close price on Signal date. This calculated Target price was then compared with actual Close price on day 6 from Signal date. If actual price was above target price, then the trade was regarded as profitable and dependent variable was coded as 1. If the condition was not satisfied, then dependent variable was coded as 0.

Development of Independent Variables

It is popular between traders and analysts to use Rate if Change (ROC), also known as Momentum, to define trends in prices and make decisions. With that in mind ROC was calculated for all available variables utilizing commonly used periods of 1, 4, 9, 15, 30, 90, 180 and 300 days. These new variables were used to create to create moving averages of them. For price variables weighted moving averages were also calculated using Volume and Open Interest as weights.

Additional features related to potential levels of price resistance and market sentiment were created:

- Ratios of Close price to Minimums and Maximum prices of periods of different lengths: 9, 30, 60, 90 and 180 days
- Daily difference between Future Close price and Spot price of gold and its moving averages for 4, 9, 15, 30, 60 and 90 days

A total of 163 independent variables were created and attached to dependent variable. Due to some missing data and the fact that some of the independent variables could not be calculated for the first year of available data, 634 out of available 650 observations were selected for analysis. They were split into Training and Testing datasets using 70/30 ratio.

Modeling

Descriptive Analysis

The Weight of Evidence (WOE) and Information Value (IV) framework was used for exploratory analysis and variable screening for binary classifier, being dependent variable. According to Kim Larsen [6] WOE and IV enable to:

- Consider each variable's independent contribution to the outcome.
- Detect linear and non-linear relationships.
- Rank variables in terms of "univariate" predictive strength.
- Visualize the correlations between the predictive variables and the binary outcome.

WOE and IV play two distinct roles when analyzing data:

- WOE describes the relationship between a predictor and a binary target variable.
- IV measures the strength of that relationship.

Using Testing dataset WOE and IV were calculated. Below table depicts top 12 variables by IV. Their distribution suggests a non-linear relationship between dependent and independent variables.

[6] Larsen, Kim. (2016, April 07). Uplift Models: Optimizing the Impact of Your Marketing. Retrieved July 15, 2019, from: https://www.predictiveanalyticsworld.com/sanfrancisco/2016/uplift_modeling.php

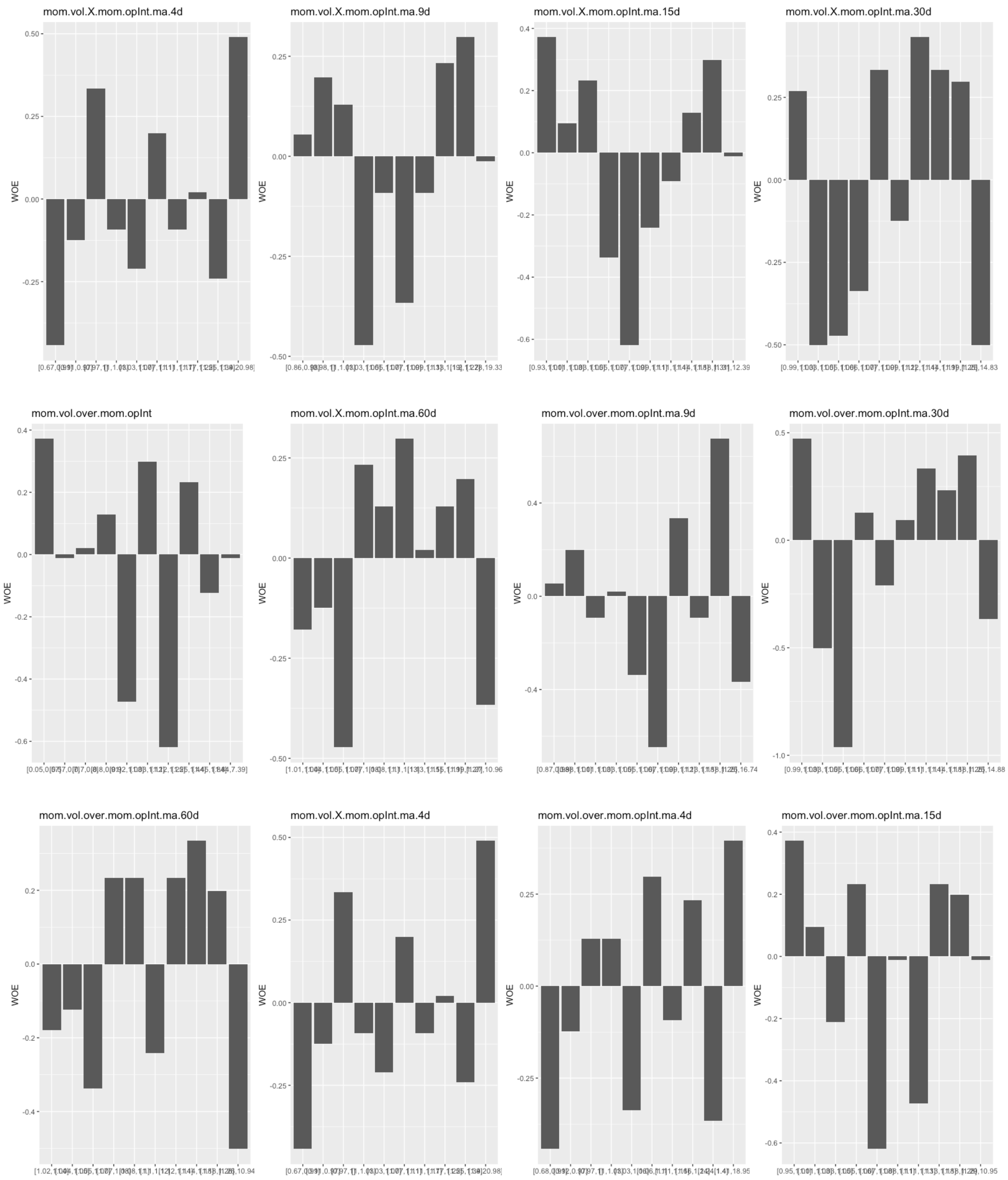

Figure 2. Information Value and Weight of Evidence

In order to understand if the same patterns show in Testing dataset, the Net Weight of Evidence (NWOE) and Net Information Value (NIV) were calculated by deafferenting WOE and IV of Training and Testing datasets. The table below shows top 10 variables by NIV. It appears that the penalty introduced by the differences between datasets is very high comparing to original values calculated from Testing dataset and NIV is less than half of what was initially expected. That further complicated research as it implies that findings from Testing are not likely to be validated and suggest that better results may be achieved from unsupervised models such as Neural Networks.

	Cluster	Variable	IV	PENALTY	AdjIV
62	45	mom.49d.ma30cl	0.21430317	0.12541399	0.0888891803
52	35	mom.90.180.ma30cl	0.24042815	0.15998139	0.0804467672
76	59	cl.max15	0.27075677	0.22341167	0.0473450940
72	55	cl.max60	0.09868773	0.05796630	0.0407214349
29	19	mom15c9ma	0.17362132	0.13898644	0.0346348816
74	57	cl.max30	0.13911815	0.10484595	0.0342722027
69	52	cl.min180	0.12282260	0.08897672	0.0338458741
131	91	ma.fut.spot9d	0.17035145	0.13852668	0.0318247693
130	90	ma.fut.spot15d	0.24107789	0.21012616	0.0309517249
139	94	mom1c	0.06906315	0.03937413	0.0296890180

Table 3. Top 10 Variables by Adjusted Information Value

Variables Selection

Independent variables were created with high degrees of correlation as the same original variables were used in different manners with expectation that some of them may have marginally better predictive power then others. A high degree of correlation causes issues with most supervised MALs so dimensionality reduction needed to be performed.

In order to achieve that and select most usable predictors, Variable Clustering was performed, using previously calculated IV. It divided variables into mutually exclusive clusters such that:

- the correlations between variables assigned to the same cluster are maximized.
- the correlations between variables in different clusters are minimized.

Using different cutoff levels for variables correlation 2 datasets were created:

- 24 variable dataset with no correlation above 0.8, most suitable for MLAs sensitive to multicollinearity such as regressions and trees
- 123 variable dataset with no correlation above 0.92, most suitable for MLAs not sensitive to multicollinearity such as Neural Networks (NN).

Logistic Regression

Binary Logistic Regression is a type of regression in which the binary response variable is related to a set of discrete or continuous explanatory variables. It was tested with 24 explanatory variables for completeness purposes despite exploratory analysis findings of lack of linear relationship. The best model had 3 significant variables and it failed to predict more than 2% as “1”s in both Training and Testing.

Logistic Regression was also explored with Principal Components (PC) of 123 variables and Regularized regression that penalizes coefficients that do not add value. Only one out of 18 PCs that explained 80% of the variance was selected as significant. All coefficients were suppressed to “0” in Regularized regression. Neither predicted over 3% as “1” in Testing or Training.

Decision Tree

CART Decision Tree is a type classification methodology, in which the tree is obtained by recursively partitioning the data and fitting a prediction model within each partition. In order to avoid overfitting, the minimum split was set to 20 in this study; the tree will only split when there are at least 20 observations in each branch. The result was promising in Training stage with 29% observations predicted as “1” with 74% accuracy. It did not hold for Test dataset with accuracy of prediction of “1” dropping to 30% in line with random choice. This kind of result was expected based on differences between WOE and IV between Training and Testing datasets.

Random Forest

Random Forest is an ensemble method that constructs multiple decision trees and outputs the single tree, which is the mode of the individual trees. The number of trees built was set to 200 given the relatively small data set. Different max depth values from 2 to 15 were tried and based on the best 5-fold cross-validation, the tree with a depth setting of 5 resulted in an 29% predicted as “1” with 87% accuracy. Similarly, to CART this result reverted 28% accuracy with 30% observations predicted as “1”.

Gradient Boosting Classifier

Gradient Boosting (GB) is an ensemble method that produces a prediction model in the form of weak prediction models, typically decision trees. The model descends various stages until it finds an optimal value for the coefficients. Like Random Forest. the number of trees was set to 200 with various depth settings. Unlike Random Forest, a learning rate was required for the boosting process. Various settings for the depth and learning rate were tried, however none of them managed to predict any “1” in Training. The Gradient Boosting Classifier was therefore discarded as option.

Neural Networks

Neural Networks (NN) are computing systems based on a collection of connected units or nodes called artificial neurons. Figure 4. A multilayer perceptron (MLP) is a class of feedforward NN consisting of at least three layers: input, output and at least one hidden layer. MLP utilizes a supervised learning technique called backpropagation for training [7]. Its multiple layers and non-linear activation functions allow to distinguish data that is not linearly separable [8].

[7] Goodfellow, Ian; Bengio, Yoshua; Courville, Aaaron (2016) *Deep Learning*. MIT Press. p. 196.
[8] Cybenko, G. 1989. Approximation by superpositions of a sigmoidal function. 2(4), 303–314.

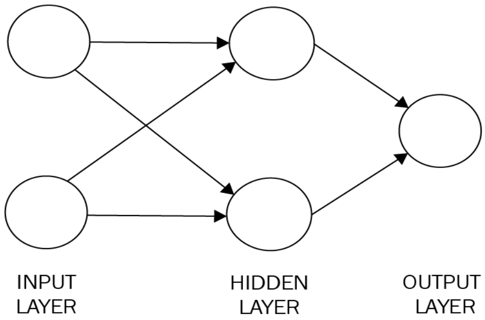

Figure 4. Generic Neural Network Diagram

MLPs with a number of layers between one and six were tested with various activation functions for layers. Best results for dataset with 123 variables were produced by a model with 3 hidden layers and Rectifier Linear Unit (ReLU) activation function. Table 5. It produced accuracy of over 50% in identification of “1” in both Training and Testing. Overall percentage of predicted “1”s was low in both instances ranging between 11% and 14%. Lowering probability cutoff did not help as increase of predicted “1” was accompanied with drop of accuracy to Random Choice level.

Event (Random Choice)	Holdout		Training	
	Predicted 0 (Accuracy %)	Predicted 1 (Accuracy %)	Predicted 0 (Accuracy %)	Predicted 1 (Accuracy %)
0 (70%)	132 (76%)	11	280 (73%)	24
1 (30%)	41	17 (61%)	104	25 (51%)
Predicted % of Total	84%	14%	89%	11%

Table 5. Confusion Matrix of NN model with 123 Variables

A model with the same parameters was trained using dataset with 24 variables. It gained results slightly worse in terms of accuracy of predicting “1;” however it predicted significantly more of them in both Training and Testing. Accuracy was 55% for Training and 43% for Testing with proportion of predicted “1” being 19% and 21%.

Event (Random Choice)	Holdout		Training	
	Predicted 0 (Accuracy %)	Predicted 1 (Accuracy %)	Predicted 0 (Accuracy %)	Predicted 1 (Accuracy %)
0 (70%)	115 (74%)	24	267 (76%)	37
1 (30%)	41	18 (43%)	83	45 (55%)
Predicted % of Total	79%	21%	81%	19%

Table 6. Confusion Matrix of NN model with 24 Variables

Seeking further improvements, other types of NNs were tested with one Deep Learning (DL) model producing interesting results. The process, called autoencoder, depicted in figure 7, often used for anomaly detection. It is a symmetric feedforward Neural Network with the purpose of reconstructing its inputs instead of predicting the target values. For this research it involved creating a model with Training dataset consisting only with observations where dependent variable was “0”. It was then trained with inputs equaling outputs in order to learn what the “normal” process looks like. After that “1”s from Training and Testing datasets were introduced to the model and their error terms were measured.

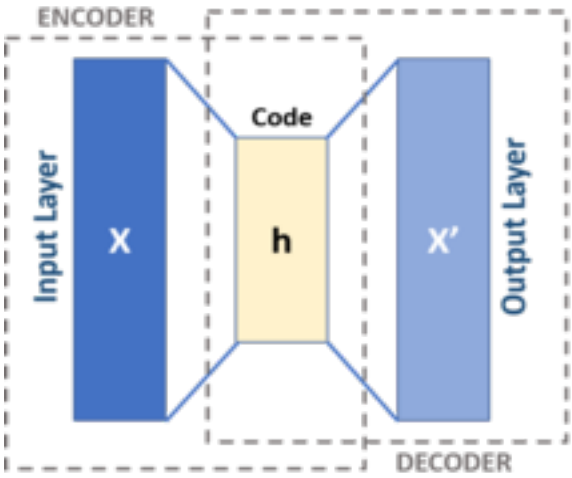

Figure 7. Generic Autoencoder Diagram

Below is reconstruction error by class for Autoencoder with 4 dense layers and 24 variables.

Figure 8. Reconstruction Error of Autoencoder with 24 Variables

The best results from this model is achieved using Reconstruction Error cutoff of 0.02. Almost 20% of observations lay below the cutoff line with proportion of “1 of only 8%. It still has 35% accuracy in identifying 1 in the remaining population, but the accuracy of identifying 0 below the line is the best seen in this analysis.

Assessing model performances

Among all classification methods, only Neural Network models produced results with accuracy of identifying 1 of greater than 40%, when predicting 20% observations as 1 and accuracy in identifying 0 of over 90% for 20% of the observations. Model with 123 variables achieved 50%+ accuracy in both Training and Testing but the proportion of predicted 1 was 11% and 14%. Due to Neural Networks being black-box methods, understanding of the importance of variables cannot be achieved.

Validation of assumptions

No assumptions were made about the form of either the true functional dependence or the form function estimate. These models did not need a linear relationship between the dependent and independent variables. In fact, these models supported many types of relationships. Also, the variances did not need to be heteroscedastic for each level of the independent variables. Instead, the analysis focused only on the predictive ability of the models.

Model validation

Traditional Verification and Validation (V&V) techniques such as cross-validation cannot be applied to Neural Networks due to their black box structure [9]. The only validation of the model developed using Training dataset was to apply holdout Testing dataset.

Time-series approach

The logic to define prediction in confusion matrix is as follows:

- If prediction is above target price and actual comes above target price than it is a True Positive
- If prediction is above target price and actual comes below target price than it is a False Positive
- If prediction is below target and actual comes below target price than it is a True Negative
- If prediction is below target price and actual comes above target price than it is a False Positive

The model was first trained with all data starting in 1975 through first Signal in 2005. Then it was re-fit for every signal until the end of 2014 and Forward Validation was performed using data from 2015-2019. Each Trend and Regression component was used separately, then they were ranked by predictive power. For final model best Trend component was selected first, then Trend components were added until model stopped improving. Then Regression components were added the same way.

[9] Tim Menzies, Charles Pecheur. Verification and Validation and Artificial Intelligence. *Advances in Computers,* **2005,**65.

Final model includes two trend components (Local Trend of Close Price and Local Trend of High Price) and two regression components (3-day lag of Close Price and 3-day Lag of High Price). Accuracy of predictions for Training and Validation periods are significantly higher than for any model previously considered. Table 9. Direct comparison is not possible, however, due to differences in approaches and how dependent variable is used for modeling.

Event (Random Choice)	Holdout		Training	
	Predicted 0 (Accuracy%)	Predicted 1 (Accuracy %)	Predicted 0 (Accuracy%)	Predicted 1 (Accuracy%)
0 (70%)	33 (86%)	4	99 (92%)	10
1 (30%)	5	16 (75%)	9	30 (75%)
Predicted % of Total	66%	34%	73%	27%

Table 9. Confusion Matrix of BSTS model with 4 components

Descriptive Analysis Results

As was explained in descriptive analysis section, the relationship between dependent and independent variables is weak and non-linear. That prevented us from understanding of importance of specific predictive variables.

Predictive power of historic data

Through the construction of multiple predictive models using various machine learning techniques, historic price data was shown to have a moderate predictive power of development of a new upward trend.

The best Neural Network model successfully predicted 35% of all instances of 1 in Training and 31% in Testing. In this case, historic price information cannot be used as a sole predictor for development of new upward trend. However, it can be used as a guide to potential profitability and help with decisions of how much capital to allocate to each trade.

Time series approach predicts over 70% of successful trades and can be used as a basis for trading strategy, but its performance needs to be monitored to assure continuous accuracy.

Variable importance

24 variables selected using IV and WOE methodologies appeared to have similar, if not better, predictive power as 123 variables. IV and WOE methodology therefore can be used regardless of the nature of the relationship between predictors and predicted variables. Individual results of the 24 variables selected are not possible to obtain due to black box nature of Neural Networks.

FINDINGS

This analysis faced numerous challenges. There was no existing research in and the methodologies used were experimental. Though accuracy of prediction of over 40% is definitely better than Random Choice at 30%, the overall number of events is not high enough for the model to be used as single decision-making tool.

This research helps to understand the relationship between historic data and development of new trends. Despite the small dataset, the predictive power that historic prices have on new trends has been proven. The models can be used as a guide to capital allocation by traders.

The best binary model from this research resulted in successful prediction of 33% of events with accuracy of 43% and the best Time Series model achieved 75% success in both accuracy and percentage predictions. Based on these metrics, we can reject the hypothesis that historic prices can be used to predict 90% of events with accuracy of above 50%.

Practical uses

The machine learning models in this research can be used as logic to capital allocation. Autoencoder model can be used to identify dates when probability of success is extremely low 8% and either exclude them from consideration or use a low-risk strategy such as calendar spread. If observation is not rejected by Autoencoder model, then NN can be used to allocate capital based on probability of success. As it stands observations with calculated probability below 0.5 have 26% likelihood of success, when those with probability above 0.5 have 42% likelihood. It means that the likelihood for observations from 0.5+ probability group to be profitable is 1.6 times higher. To achieve balanced approach to risk trader can buy 2 contracts every time probability is under 0.5 and 3 contracts when it is above 0.5.

Time series model can be used to make investment decisions on ongoing basis, by making trades every time the model suggests.

Future extensions

The research can be expanded to explore the same approach on other Futures contracts. Future studies can also further explore hidden layers structure of Neural Network and tune their hyperparameters. Due to the time series nature of the dataset, time series approaches using Neural Networks with Long Term Memory and further extensions of Bayesian Structured Time Series should be explored.

BIBLIOGRAPHY

A Review of Warren Buffett's Investing Strategy. (2019). Retrieved from https://www.investopedia.com/investing/warren-buffetts-investing-style-reviewed/

Balsara, Nauzer. (1992). Money Management Strategies for Futures Traders. New York: John Wiley & Sons, Inc.

Bernstein, W. (2010). *The four pillars of investing*. New York: McGraw Hill.

Best Forex Scalping Strategies. (2019) Retrieved from https://www.theforexarmy.com/forex-scalping-strategy/

Bollerslev T. (1986), Generalized autoregressive conditional heteroscedasticity, Journal of Econometrics, 31, 307-327.

Box G.E.P., Jenkins G.M. (1983), Analiza szeregów czasowych. Prognozowanie i sterowanie, PWN, Warszawa.

Charles Duchock, (1990), "Evidence of Efficiency in United States Futures Oil Prices," dissertation, United States International University.

Chew, D. H. *Corporate Risk Management*. Columbia University Press, 2008.

Competitivefutures.com. (2019). [online] Available at: https://www.competitivefutures.com/wp-content/uploads/2015/07/Case-Shiller-Q1-2013.jpg [Accessed 5 Jun. 2019].

Cox, D. R. (1955). "Some Statistical Methods Connected with Series of Events". *Journal of the Royal Statistical Society*. 17 (2): 129–164. doi:10.2307/2983950

Cybenko, G. 1989. Approximation by superpositions of a sigmoidal function Mathematics of Control, Signals, and Systems, 2(4), 303–314.

Dezfouli, A. and Bonilla, E. V. (2015). Scalable inference for Gaussian process models with black-box likelihoods. In Advances in Neural Information Processing Systems.

Di Lorenzo, R. (2012). *How to Make Money by Fast Trading*. New York: Springer.

Doerr C., Blenn N., and Van Mieghem P. (2013). Lognormal infection times of online information spread. PLOS ONE, 8.

Engle R. F. (1982), Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, Econometrica, 50, 987-1006.

Futuresforecasts.com. (2019). *eMini S&P500 Futures Prices & Chart Forecasts*. [online] Available at: http://www.futuresforecasts.com/emini-sp-500-futures.html [Accessed 5 Jun. 2019].

Gannon, N. (2010). *Investing strategies for the high net-worth investor*. New York, NY: McGraw-Hill.

Gelman A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003). Bayesian Data Analysis. Chapman and Hall/CRC.

Goodfellow, Ian; Bengio, Yoshua; Courville, Aaaron (2016) *Deep Learning*. MIT Press. p. 196.

Granger C.W.J., Swanson N.R. (1997), An Introduction to stochastic unit–root process, Journal of Econometrics, 80, 35-62.

Jones C.R., Marriott J.M. (1999), A Bayesian analysis of stochastic unit root models, Bayesian Statistics, 6, s. 785-794.

Jorgenson, Dale. *Capital Theory and Investment Behavior.* American Economic Review 53, no. 2 (1963): 247-259.

Larsen, Kim. (2016, April 07). Uplift Models: Optimizing the Impact of Your Marketing. Retrieved July 15, 2019, from: https://www.predictiveanalyticsworld.com/sanfrancisco/2016/uplift_modeling.php

LeBaron, D., & Vaitilingam, R. (2002). *Dean LeBaron's treasury of investment wisdom*. New York: Wiley.

Lee, S. H., Crawford, M. M., and Wilson, J. R. (1991). Modeling and simulation of a nonhomogeneous Poisson process having cyclic behavior. Communications in Statistics Simulation, 20(2):777–809.

Leybourne S.J., McCabe B.P.M., Mills T.C. (1996), Randomized unit root processes for modelling and forecasting financial time series: theory and applications, Journal of Forecasting, 15, 253-270.

Leybourne S.J., McCabe B.P.M., Tremayne A.R. (1996), Can economic time series be differenced to stationarity? Journal of Business and Economic Statistics, 14, 435-446.

Lloyd, C., Gunter, T., Osborne, M. A., and Roberts, S. J. (2015). Variational inference for Gaussian process modulated Poisson processes. International Conference on Machine Learning.

Luo, Z.; Cai, X.; Tanaka, K.; Takiguchi, T.; Kinkyo, T.; Hamori, S. Can We Forecast Daily Oil Futures Prices? Experimental Evidence from Convolutional Neural Networks. *J. Risk Financial Manag.* **2019**, *12*, 9.

Lynn, G. (2006). *A dictionary of derivatives and financial engineering*. Oakland, CA: Mason Pub.

Markham, Jerry W., (1987), The History of Commodity Futures Trading and Its Regulation, Praeger

Redhead, Keith (1997). *Financial Derivatives: An Introduction to Futures, Forwards, Options and Swaps*. London: Prentice-Hall

Research.cs.aalto.fi. (2019). *GPstuff - log Gaussian Cox process demo*. [online] Available at: https://research.cs.aalto.fi/pml/software/gpstuff/demo_lgcp.shtml [Accessed 5 Jun. 2019].

Ross, J. (2004). *Day trading*. Cedar Park, TX: Ross Trading, Inc.

Scalping: Small Quick Profits Can Add Up. (2019). Retrieved from https://www.investopedia.com/articles/trading/05/scalping.asp

Schap, K. (2005). *The complete guide to spread trading*. New York: McGraw-Hill.

Simma, A. and Jordan, M. I. (2010). Modeling events with cascades of poisson processes. In UAI, pages 546–555.

Sollis R., Leybourne S.J., Newbold P. (2000), Stochastic unit roots modelling of stock price indices, Applied Financial Economics, 10, 311-315.

Tim Menzies, Charles Pecheur. Verification and Validation and Artificial Intelligence. *Advances in Computers ,* **2005,**65,153-201.

Titsias, M. K. (2009). Variational learning of inducing variables in sparse Gaussian processes,. In Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics.

Tong H. (1990), Non-linear time series: a dynamical system approach, Oxford University Press, Oxford.

Unofficialgoogledatascience.com. (2019). *Fitting Bayesian structural time series with the bsts R package*. [online] Available at: ttp://www.unofficialgoogledatascience.com/2017/07/fitting-bayesian-structural-time-series.html [Accessed 5 Jun. 2019].

Van Blokland, P. (1995). *Using the futures market to lower the farm management risks of producing for unknown market prices in the USA*. Gainesville, FL: University of Florida, Institute of Food and Agricultural Sciences, Food and Resource Economics Dept.

Vanhatalo J., Riihimaki J., Hartikainen J., Jylanki P., Tolvanen V., Vehtari A. (2013). Gpstuff: Bayesian modeling with gaussian processes. J. Mach. Learn. Res., 14(1):1175–1179.