Predicting the Direction of Equity and Currency Index Movement using Artificial Neural Network

December 13, 2016

Chang Liu

Introduction

This paper examines the predictability of the stock and currency market index for a given country or region. According to the Efficient Market Hypothesis, past information can not be used to predict the future prices. The author's personal research as well as a body of literature in financial economics [5][6][7] suggests that international equity and currency markets can have cross-market momentum, where one market in a country correlates with another market in a different region in a lagged manner. This cross market momentum makes it possible to predict future prices using past prices in certain markets.

In addition to market prices, macroeconomic conditions are inherently linked to the performance of equity and currency market. Factors such as GDP, unemployment, and inflation reflect the fundamental conditions of an economy and thus determine financial asset prices of that economy. Bollerslev et al. [2] shows that macro fundamentals can predict the high-frequency dynamics of foreign exchange, using 5-minute exchange rate data. An interesting result is that the US macroeconomic announcement surprises have much better forecasting power on German Marks than the German macroeconomic indicators, presumably because of the sweeping influence of the US economy on other major economies. Their work further suggests investigation of possible cross-market dynamics in the international equity market.

In the computer science literature, researchers [3][4] using Artificial Neural Networks (ANN) have shown promising results in predicting currency and equity returns in developed markets. However, it seems that developing markets have less coverage by these methods. In sum, we aim to predict the direction of daily index movement in both developed and developing currency and equity markets.

Data

We use daily price series of currency and equity indices in 29 major developed and developing regions and countries, and 16 US macro fundamentals as input features (listed in the appendix). The regional indices are Morgan Stanley Capital International (MSCI) indexes and Bloomberg JP Morgan Latin America Currency Index. They are primary indexes that are broad benchmark for asset allocation of large institutional investors across the world. The price series starts from January 3, 2005 to December 12, 2015. The prices series are then transformed into daily percentage change (a.k.a. returns).

Since most macro indicators are only observed monthly, the macro variables are first transformed into percentage change using the last available observation. To avoid zeros as input features, they are then padded forward to match the daily frequency of our prices. The macro indicators start from January 3, 2005 to March 31, 2015. In the experiments below, if we include the macro variables, we will have 2616 observations lasting from January 3, 2005 to March 31, 2015; if we exclude them, for various purposes, we have 2808 observations from January 3, 2005 to December 31, 2015. All variables are normalized to zero mean and unit variance as input features.

Methodology

Forecasting method: We adopted the walk-forward method which produces overlapping training-validation-testing sets, suitable for financial time series with limited samples. To start, we will train pick the first fold of N + K observations: the first 80% of N are training set, the second 20% of N held out for validation, and the next K obervations is then the number of steps ahead forecast. For the second fold, we move our window K steps forward so that we can predict the next K observations following the last fold. All the K steps ahead make up the test set of our predictions. As we show later using validation and test set, we find N = 2500 and K = 50 to be a good choice. Then we have around 100 predictions as our test samples for the series. If we exclude macro variables and have a longer series lasting till the end of December, 2015, the test sample size will be around 300.

We consider using all or a subset of variables with 1 lag to predict the next-period return of a given index, because the currency and equity markets are generally quite liquid and efficient. This setting implicitly has a modeling assumption: we are parametrizing a Markov model that future prices are conditionally independent on the farther past given the previous observation.

Training the network: We will train a feedforward neural net using backpropagation algorithm implemented in Keras library[1] in Python. In consideration of our relatively small data sets, we will use the following architecture with regularization techniques:

- 1 to 3 layers of hidden units; each with 1,2,…,10, 15, 20, 30 hidden units[2]
- Activation and loss functions: Sigmoid for output layer; ReLu for the rest; binary cross entropy for loss function
- Alternative loss function: treat the problem as regression, using MSE as loss function and sigmoid for output layer[3]
- Initialization: normal distribution with zero mean and variance scaled with the incoming units
- Run two optimizers, Adam and Rmsprop, with learning rate fine-funed while other parameters are set to be:
    - Adam `(beta_1=0.9, beta_2=0.999, epsilon=1e-08, decay=0.0)`
    - RMSprop `(rho=0.9, epsilon=1e-08, decay=0.0)`
- Dropout: 20% at each layer
- L1 regularization with lambda = 0.01; constraint of the norm of a weight matrix to be under 3
- Early stopping when validation set does not improve over 2 epoches

We use the walk-forward validation set to pick the optimal parameters and report the test accuracy for the optimal model. We run 5 times the same model to report the average accuracy because different initializations might lead to unstable results. For comparison, we choose off-the-shelf classification algorithms Random Forest and Gradient Boosted Trees (parameters are described in Figure 4) as they are relatively easy to tune and have good results.

Experimental Results

We predict the daily directioned of a foreign exchange rate, equity or currency index given the 1-period-lagged observations of all the other variables. We find that only 3 indexes have statistically significant accuracy that is different from 50%, a random guess: MSCI Emerging Market Asia Equity Index, MSCI Pacific Equity Index, and Bloomberg JP Morgan Latin America

---

1 https://keras.io/

2 A rule of thumb to avoid overfitting is to have 50~75% of the number of input features as hidden units, assuming that we have at least half as many weights. With our small sample size, for example, if we pick N =2500 while there are 45 input features, we have only 2000 effective training set, and therefore can only choose below 2000/2/45 < 30 number of hidden units.

3 Empirically, we find that training the neural network with Y as real values has no predictability compared with training it with Y as binary values.

Currency index. We will call them EM ASIA, PACIFIC, EM LATAM for short. Figure 1 and 2 show the best-performing accuracies for each index compared with the baseline algorithms. The neural network slightly outperforms the baseline algorithms. The results also support our modeling assumption because variables with more than 1 lag produces no statistically significant predictions.

Validation accuracy	Test accuracy	Region	Number of layers	Number of nodes	Include macro factors?	Treated as regression?	Optimizer
70%	69%	EM ASIA	3	30	0	0	adam
71%	71%	EM ASIA	2	20	1	0	adam
70%	68%	EM ASIA	1	8	0	1	rmsprop
70%	69%	EM ASIA	1	20	1	1	adam
68%	67%	EM LATAM	1	20	0	0	rmsprop
70%	72%	EM LATAM	3	20	1	0	adam
74%	68%	EM LATAM	1	15	0	1	rmsprop
69%	71%	EM LATAM	1	10	1	1	adam
70%	71%	PACIFIC	2	30	0	0	rmsprop
71%	70%	PACIFIC	1	7	1	0	adam
70%	70%	PACIFIC	1	30	0	1	rmsprop
70%	71%	PACIFIC	1	30	1	1	rmsprop

Figure 3. Best-performing results of neural network and the architectures. All accuracy values are statistically significant at the 1% level, which rejects the null hypothesis that the neural network has no predictablity, i.e., no different than a random guess. Other parameters: training size = 2500, number of steps ahead = 50, learning rate for the optimizer = 0.01. Number of input features equals 45 and test size = 100 if we include macro factors (denoted by '1' on the corresponding column above) and number of input features equals 29 and test size = 300 otherwise. Regression means to use MSE as loss function.

	EM ASIA				EM LATAM				PACIFIC			
Algorithms	RF		GBT		RF		GBT		RF		GBT	
Test Accuracy	68%	64%	63%	60%	67%	63%	68%	62%	54%	60%	61%	66%
Include macro factors?	0	1	0	1	0	1	0	1	0	1	0	1

Figure 2. Best-performing results of Random Forest (RF) and Gradient Boosted Trees (GBT). All accuracy values are averaged over 10 trials and statistically significant at the 1% level except for the Random Forest when we exclude macro variables. Other fixed parameters: training size = 2500, number of steps ahead = 50. For GBT, number of estimators = 1000, learning rate = 0.01 and max features are sqrt(number of features), max_depth = 3. For RF, number of estimators = 100, learning rate = 0.01, max features are sqrt(number of features). Nodes are expanded until all leaves contain less than min_samples_split samples (set at 2); Gini impurity is used for split criterion and entropy for information gain. Number of input features equals 45 and test size = 100 if we include macro factors (denoted by '1' on the corresponding column above) and number of input features equals 29 and test size = 300 otherwise.

Robustness of Results: We now discuss the robustness of the results across the parameters. There are a number of architectures that offers comparable accuracy, as complexity of the models varies. Empirically, all training takes no more than 15~20 epochs to terminate, and more training almost always leads to overfitting. Therefore, we think the dynamics of data is relatively easy to capture by both small and large (regularized) models.

We also find that the number of steps ahead for predictions is stable at least over 100 trading days, as shown in Figure 3. This is a surprising result, as we would think that new information every day would require us to rerun our model frequently, and therefore a shorter predictive horizon would have higher accuracy. An economic explanation for the signal robustness is that under normal market conditions, the dynamics of the markets remains stable and can therefore be captured by simple models. Daily incoming information might not change much of the current information set unless events unexpected by market participants like market crashes happen.

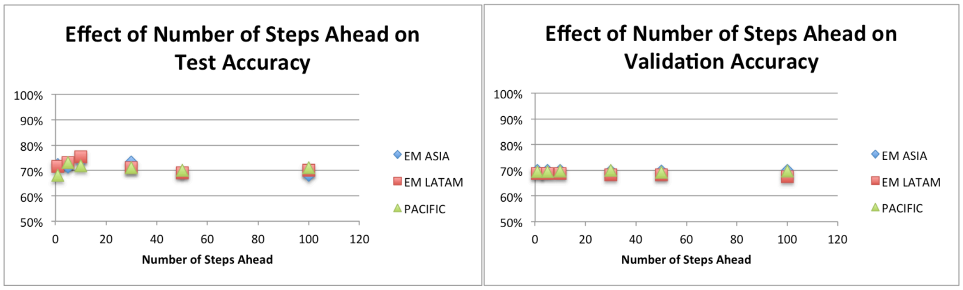

Figure 3. Stability of predictive power across large number of steps ahead forecast, in all three indexes.

We compare models with and without the macro variables to examine whether the much lagged economic indicators have predictive power[4]. It is evident that there is no added predictive power for the macro variables, primarily because of the stale nature of macro indicators that do not add new information other than what is reflected in the prices. As a result of dropping them, we also seek to find the most "important" among the other predictors as feature selection.

Feature Selecton: We group them into 3 buckets, developed market currencies, emerging market currencies, and regional equity indexes, each of which have salient characteristics. Combined with the knowledge about the financial variables, we perform greedy forward search within each group and select the best-performing set of features, as shown in Figure 4. Even though we only use one iteration of forward search, the result reveals high predictive power in these sets. The test accuray on each index using the corresponding set is: 68%, 70% and 67% in the same order as in the figure. Although the greedy approach is not guaranteed to be optimal in general, the economic grouping of the feature set is sound in this case: all three markets are highly influenced by North American markets as a financial center, while each index has economic connections particular to that region. For example, only PACIFIC equity market is influenced by Japan Yen (JPY). Emerging markets such as Brazil, Mexico, and Russia are tightly connected because of their trade relations such as commodity exports.

Predicting EM LATAM											
Predictor	USA	North America	EM LATAM Equity	EM EMEA Equity	South Africa	Brazil	Mexico	EM	Europe	Russia	Turkey
Validation	66%	65%	62%	62%	62%	62%	62%	61%	61%	61%	61%
Predicting PACIFIC											
Predictor	USA	North America	Mexico	JPY	Europe						
Validation	67%	67%	60%	60%	60%						
Predicting EM ASIA											
Predictor	USA	North America	South Africa	Mexico	EM	Russia	Turkey	Brazil	EM EMEA Equity	EM LATAM Equity	
Validation	65%	65%	64%	64%	63%	63%	62%	62%	61%	61%	

Figure 4. Most important features selected for each index ranked according to their individual validation accuracy using greedy forward search. The ANN is trained with 2 layer of 20 hidden units without macro variables using sigmoid as loss function. Other parameters remained the same as the optimal model in Figure 3.

[4] Dropping the macro variables (if they are shown to add no value) means we will have test size = 300 and increased statistical significance for the accuracy values.

We also look at whether the increased sample size improves our results as shown in Figure 5. There is little difference in validation accuracy with N = 2000 or N = 2500 training samples. While smaller training size seems to trade off performance on test set, the larger test size (see caption) as a result produces accuracy with little variance and show more stability across different number of steps ahead K. We then prefer larger training size, fixing N = 2500, as we can drop the macro variables to have 300 test samples, which is large enough.

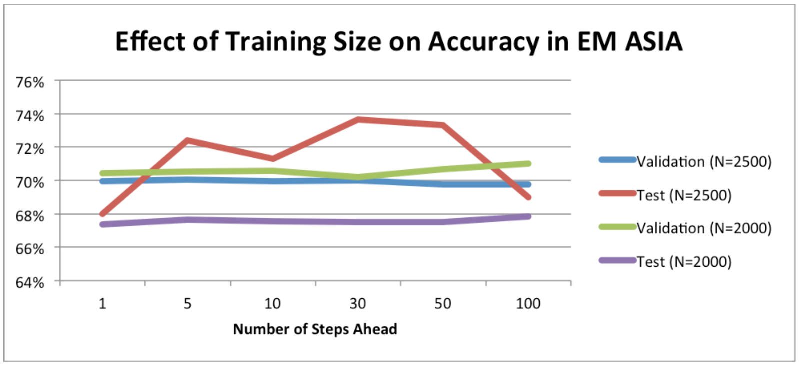

Figure 5. Effect of increased training size on accuracy in one index. When N = 2500, test size = 100; when N = 2000, test size = 500+100 = 600. The tradeoff between training size and test size is shown by the variance of and the magnitude of test accuracy on varying training size.

Trading Simulations

Predictability does not necessarily translate into profits. Market factors such as transaction cost, market impacts, and liquidity costs, as well as unexpected market events like the 2008 financial crisis, especially in emerging markets, could potentially undermine the effect of good forecasts. These market inefficiencies could partly explain why our moderately accurate forecasts exist in the first place. To simulate the realistic setting, we design a long-short strategies that trade the three indexes we considered based on the output probabilities of our neural network.

Strategy: If the forecasting probability is high for a given index, above 70%, we long that index today; if that probability is low, below 30%, we short that index today. Implementation of such trades will use future contracts or index ETF[5]; we consider trading costs from the perspective of a large institutional investor or a bank, which usually range from 0.05% to 0.1% for buy or sell. Hence, we consider 0.2% for entering and exiting a position. We at last combine the three strategies with equal weights for the diversification purpose. The backtest starts from October 16, 2014 to December 18, 2015[6].

Some salient features of the back-test shown in Figures 6 and 7 include 1) the small maximum drawdown compared to the index and 2) the profitability measured by various metrics due to the market timing ability of our prediction machine. This is most evident in the market crash in August and September 2015, when the MSCI EM index drops by 4.4% in a day. But the neural network was able to predict the crash, and shorted the indexes correctly, as we see the divergence in the gross return curve at that period. The profitability of the strategy reveals the highly predictable market momentum and that dynamics is captured by the neural network.

[5] Futures will be more suited to a daily trading strategies like the one considered in this paper due to the lower cost and tracking errors. For MSCI index futures, see details at https://www.theice.com/products/Futures-Options/Equity-Derivatives/MSCI-Indexes. For MSCI ETF, see details at https://www.ishares.com/us/products/239696/ishares-msci-world-etf .

[6] The slight mismatch with the end date of the year was due to the test size of a walk-forward window that was set to be 50 market days.

	Annualized Return (pre-cost)	Annualized Volatility (pre-cost)	Annualized Sharpe (pre-cost)	Maximum Drawdown	Positions Taken	Annualized Sharpe (post-cost)	Annualized Return (post-cost)
EM LATAM	64.6%	9.8%	6.6	-2.5%	122	4.8	44.1%
PACIFIC	74.9%	11.8%	6.4	-2.8%	132	4.7	52.7%
EM ASIA	83.4%	11.2%	7.5	-2.2%	135	5.8	60.7%
Diversified	74.3%	9.7%	7.7	-1.5%	389	5.8	52.5%

Figure 6. Performance metrics of the long-short strategy for each of the index that we forecast. Total market days = 300. For the diversifed strategy, we take positions across 3 indexes and thus it is possible that we have over 300 positions taken. Transaction cost considered is 0.2% for a round trip. Maximum drawdown for the benchmark indexes at the same period are -23%, -18%, and -26%, respectively.

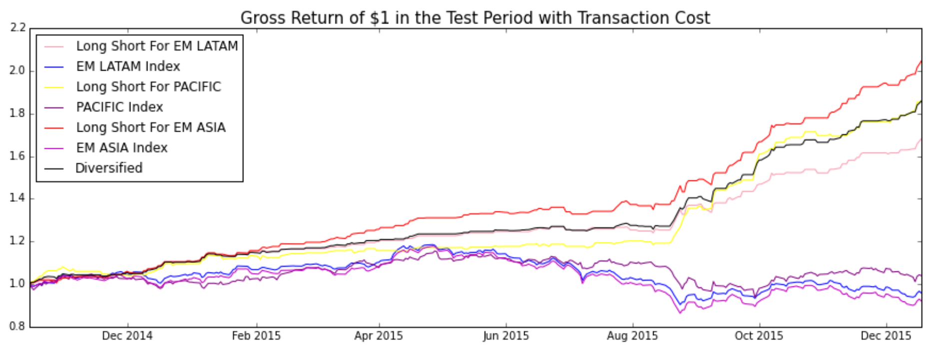

Figure 7. Gross return curve of long-short strategies versus their benchmarks. Transaction cost considered is 0.2% for a round trip

Conclusion

Using previous-period returns of international equity and currency markets, we showed that EM ASIA, EM LATAM, and PACIFIC are predictable up to around 70% out-of-sample from October 16, 2014 to December 18, 2015, outperforming baseline algorithms Random Forest and Gradient Boosted Trees. The signal is shown to be stable across dimensions such as the number of step ahead forecast, model size, and other model parameters. To gain a realistic sense of the predictability, we then formulate a trading strategy based on the probability output from ANN and confirm that the market timing ability of the prediction machine can in effect translate into profitability.

The Efficient Market Hypothesis states that prices reflect all available information including the past information set. One implication is that using past prices to predict future prices is impossible. However, the author argues that whether prices are predictable or not depends on various factors in market structure, such as liquidity in a specific market. Therefore, we should measure the efficiency of the markets rather than taking it for granted. However, the results shown here by no means is rejecting the Hypothesis. The fact that we are able to predict a small number of markets but not the rest points directly to the spectrum of efficiency in international equity and currency markets.

Appendix

US Macroeconomic Indicators[7]:

GDP
Core Inflation
Producer Price Index
Industrial Production
Money Supply
Consumer Confidence (MICH)
Retailers sales
Non-farm Payroll
Consumer Opinion (OECD)
Unemployment Rate
Housing Starts
10 year Treasury yield
1 year Treasury yield
3 month Treasury yield
20 year Treasury yield
30 year Treasury yield

Currency and Equity Index[8]

EM	MSCI currency index
EM LATIN AMERICA	Bloomberg JP Morgan Latin America Currency Index
EM ASIA	MSCI currency index
EM EMEA	MSCI currency index
EM Local Currency	MSCI currency index
EUROPE	MSCI equity index
EAFE + CANADA	MSCI equity index
PACIFIC	MSCI equity index
NORTH AMERICA	MSCI equity index
USA	MSCI equity index
EM ASIA Equity	MSCI equity index
EM LATAM Equity	MSCI equity index
EM EMEA Equity	MSCI equity index
CHINA	MSCI equity index
S&P 500	US equity index

Exchange rate[9]

Emerging Markets	**Developed Markets**
Indonesia	Australia
China	Euro
Czech	Japan
Russia	UK
Turkey	Switzerland
South Africa	
Brazil	
Israel	
Mexico	

[7] Downloaded from https://www.stlouisfed.org/

[8] Available from Bloomberg

[9] Available from Bloomberg

References

[1] Kaastra, I., & Boyd, M. (1996). Designing a neural network for forecasting financial and economic time series. *Neurocomputing, 10*(3), 215-236. doi:10.1016/0925-2312(95)00039-9

[2] Andersen, T., Bollerslev, T., Diebold, F., & Vega, C. (2002). Micro Effects of Macro Announcements: Real-Time Price Discovery in Foreign Exchange.

[3] Sermpinis, G., Theofilatos, K., Karathanasopoulos, A., Georgopoulos, E. F., & Dunis, C. (2013). Forecasting foreign exchange rates with adaptive neural networks using radial-basis functions and Particle Swarm Optimization. European Journal of Operational Research, 225(3), 528-540. doi:10.1016/j.ejor.2012.10.020

[4] Selmi, N., Chaabene, S., & Hachicha, N. (2015). Forecasting returns on a stock market using Artificial Neural Networks and GARCH family models: Evidence of stock market S&P 500. Decision Science Letters,4(2), 203-210. doi:10.5267/j.dsl.2014.12.002

[5] ASNESS, C. S., MOSKOWITZ, T. J. and PEDERSEN, L. H. (2013), Value and Momentum Everywhere. The Journal of Finance, 68: 929–985. doi:10.1111/jofi.12021

[6] Chan, K., Hameed, A. and Tong, W. (2000) ‘Profitability of Momentum Stragegies in the International Equity Markets’, *Journal of Financial and Quantitative Analysis*, 35(2), pp. 153–172. doi: 10.2307/2676188.

[7] Bhojraj, S., & Swaminathan, B. (2006). Macromomentum: Returns Predictability in International Equity Indices. *The Journal of Business, 79*(1), 429-451. doi:10.1086/497416