

# Foreign Exchange Forecasting via Machine Learning

Christian González Rojas  
cgrojas@stanford.edu

Molly Herman  
mrherman@stanford.edu

## I. INTRODUCTION

The finance industry has been revolutionized by the increased availability of data, the rise in computing power and the popularization of machine learning algorithms. According to *The Wall Street Journal* (2017b), quantitative hedge funds represented 27% of total trading activity in 2017, rivaling the 29% that represents all individual investors. Most of these institutions are applying a machine learning approach to investing.

Despite this boom in data-driven strategies, the literature that analyzes machine learning methods in financial forecasting is very limited, with most papers focusing on stock return prediction. Gu, Kelly, and Xiu (2018) provide the first comprehensive approach to quantifying the effect of using machine learning (ML) to the prediction of monthly stock returns. Our intention is to implement machine learning methods in a relatively unexplored asset class: foreign exchange (FX).

The objective of this paper is to produce directional FX forecasts that are able to yield profitable investment strategies. Hence, we approach the problem from two perspectives:

- 1) Classification of long/short signals.
- 2) Point forecasts of FX levels that translate into long/short signals.

These frameworks allow us to exploit different machine learning methodologies to solve a single problem: designing a profitable FX strategy based on ML-generated forecasts.

## II. RELATED WORK

Machine learning methods have long been used in stock return prediction. For instance, variations of Principal Component Analysis, an unsupervised learning technique, have been applied by Connor and Korajczyk (1988), Fan, Liao, and Wang (2016), Kelly, Pruitt, and Su (2018) and Lettau and Pelger (2018) to identify latent risk factors that can explain the dynamics of stock returns. Moreover, Gu et al. (2018) have found that regularization, dimension reduction and the introduction of nonlinearities significantly improve stock return predictions.

Nevertheless, despite the large adoption of machine learning in stock return forecasting, ML applications in FX prediction have been widely ignored by the literature. Few exceptions are available. Ramakrishnan, Butt, Chohan, and Ahmad (2017) find that, when trained with commodities prices, Random Forests outperform Support Vector Machines

and Neural Networks in forecasting the Malaysian FX. Furthermore, Amat, Michalski, and Stoltz (2018) conclude that economic fundamentals gain power to forecast exchange rate even at short horizons if ML methods are applied. Finally, Hryshko and Downs (2004) apply Reinforcement Learning to create FX trading strategies based on technical analysis.

The main contribution of this paper is the assessment of the statistical and economic performance of ML-generated directional forecasts.

## III. DATASETS

We make use of two different datasets to explore the forecasting power of two types of variables: market and fundamentals. We define a market variable as an indicator with daily to weekly frequency that has a close relationship with traded securities. On the other hand, we define a fundamental variable as an indicator with monthly frequency that is closely related to the macroeconomy.

Finally, we limit the scope of our project to forecasting the USDMXN, which is the exchange rate between the US Dollar (USD) and the Mexican Peso (MXN), expressed in MXN per USD. However, the exercise can be generalized to other currencies. All data was retrieved either from Bloomberg, the Global Financial Dataset or the Federal Reserve Bank.

### A. Market Variables Dataset

We obtained the weekly closing price of the USDMXN currency pair, which we use as our target variable. In addition, we consider 25 features across both Mexico and the United States. A summary is shown in Table I.

TABLE I  
MARKET FEATURES: WEEKLY DATASET

Type	Country	Variables
Fixed Income	Mexico	Bond yields (3m, 6m, 1Y and 10Y) Debt holdings
	US	Bond yields (3m, 6m, 1Y and 10Y)* Bond Index Federal Funds Rate*
	Global	Global High-Yield Indices Emerging Market Bond Index
Stock Market	Mexico	Mexican Stock Exchange Index*
	US	S&P 500 Index*
	Global	Volatility Index*
Currency		Dollar Index* Trader positions on USDMXN
Other	Global	Economic Surprise Indices Commodities Index*

\* Also considered in the monthly dataset

The dataset spans between the first week of January 2003 and the second week of November 2018.

### B. Fundamental Variables Dataset

The fundamental variables data uses the monthly closing price of the USDMXN currency pair as our target variable. We use 27 features that describe the macroeconomic conditions of both the US and Mexico between March 1990 and October 2018. The additional features that are considered in this dataset are detailed in Table II.

TABLE II  
FUNDAMENTAL FEATURES: MONTHLY DATASET

Type	Country	Variables
Economic Activity	Mexico	IP, Industrial Production Trade Balance (Exports - Imports)
	US	IP, Industrial Production Trade Balance (Exports - Imports)
Labor Market	US	Unemployment Non-farm Payroll
	Prices	Mexico
US		CPI, Consumer Price Index PPI, Producer Price Index
Debt	Mexico	National Debt
	US	National Debt
Sentiment	US	PMI, Purchasing Managers Index Investor Sentiment
Other	Mexico	M2 Money Supply
	US	M2 Money Supply

### C. Data Processing

Almost all data processing is identical in both datasets. We first split the data into 60% train set, 20% validation set, and 20% test set. These subsets are taken sequentially in order to keep the time-series nature of the data and to guarantee our algorithms train exclusively on past data.

To translate our problem into a classification problem, we introduce the  $Signal_t$  variable which we set to 1 if the USDMXN was higher tomorrow than today. This is:

$$Signal_t = \begin{cases} 1 & \text{if } USDMXN_{t+1} - USDMXN_t \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

We also perform data processing on the features. In particular, we standardize using the mean and standard deviation of the training set for every covariate.

For the fundamentals dataset, covariates are lagged by an additional period. This is done to approximate the fact that it is extremely rare to obtain real-time macroeconomic data. By lagging the features by one month we ensure we are not peeking into the future by including unpublished data.

## IV. FRAMEWORKS AND MODELS

### A. Frameworks

First, we perform binary classification on the  $Signal_t$  variable we constructed in the data processing step. This essentially transforms what initially is a continuous variable problem into a classification task.

On a second exercise, we use ML algorithms to construct point forecasts for our raw continuous target variable,

$USDMXN_t$ . We then construct an estimated long/short signal by computing:

$$\widehat{Signal}_t = \begin{cases} 1 & \text{if } \widehat{USDMXN}_{t+1} - \widehat{USDMXN}_t \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Both strategies yield a binary signal output that we can execute as a trading strategy.

### B. Models

The performance of different machine learning algorithms is tested for each framework. In particular, we considered:

1) *Logistic/Linear Regression*: We use logistic and linear regression as our benchmark models.

2) *Regularized Logistic/Linear Regression*: We consider  $L_1$  and  $L_2$  regularization applied to logistic and linear regression. This allows to reduce overfitting in the validation set. The hyperparameter  $\lambda$ , which penalizes large coefficients, is tuned using the validation set accuracy.

3) *Support Vector Machines/Regression (SVM/SVR)*: It is highly likely that fitting FX dynamics requires a non-linear boundary. SVM/SVR with a Gaussian kernel provide the flexibility to generate a non-linear boundary as a result of the infinite-dimensional feature vector generated by the kernel.

4) *Gradient Boosting Classifier/Regression (GBC/GBR)*: Tree-based models allow us to capture complex interactions between the variables. Unlike Random Forests, which require bootstrapping, GBC allows us to keep the time-series structure of the data while considering non-linearities. It is important to notice that GBC and GBR is just considered for the market variables dataset, due to the division of work between the authors (See section IX).

5) *Neural Networks (NN)*: Neural networks can model complex relationships between input features, which could improve the forecasting performance. We consider fully-connected networks. The architecture is shown in Fig. 1.

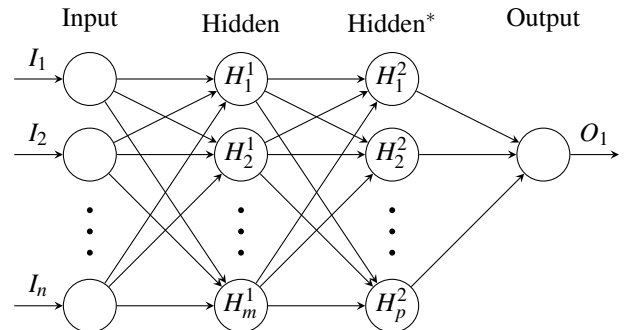


Fig. 1. NN architecture.  
\*Second hidden layer only for the market variables model.

Gu et al. (2018) show that shallow learning outperforms deeper learning in asset pricing applications. We follow this result and only consider shallow architectures. In particular, we use a network with two hidden layers for the market

variables dataset and a neural net with one hidden layer for the fundamentals dataset.

Our choice for loss depends on the framework. We select logistic loss for classification and mean squared error for the continuous target variable problem. We choose the proper activations in the same fashion: sigmoid is used for classification, while ReLU is used for the continuous target variable. Finally, we use dropout or activation regularization to avoid overfitting.

## V. HYPERPARAMETER TUNING

All model parameters are tuned using the validation set. We use accuracy as our performance evaluation in the binary classification model and mean squared error in the continuous target variable model. The resulting parameters are detailed in Table III.

TABLE III  
SELECTED PARAMETERS

Model	Framework	Market	Fundamentals
Regularized Regression	Binary	$\lambda_{LASSO} = 0.39$ $\lambda_{Ridge} = 0.14$	$\lambda_{LASSO} = 0.0785$ $\lambda_{Ridge} = 1.13$
	Continuous	$\lambda_{LASSO} = 0.0002$ $\lambda_{Ridge} = 0.0071$	$\lambda_{LASSO} = 0.75$ $\lambda_{Ridge} = 0.29$
SVM/SVR	Binary	$C = 1000$ $\gamma = 0.0001$	$C = 11.5$ $\gamma = 0.001$
	Continuous	$C = 100$ $\gamma = 0.00001$	$C = 14.5$ $\gamma = 0.0014$
NN	Binary	Neuron = 250 Epoch = 1000 Batch = 64 Dropout = 0.2	Neuron = 100 Epoch = 5000 Batch = Full $\lambda = 5, \alpha = 0.03$
	Continuous	Neuron = 500 Epoch = 2000 Batch = 32 Dropout = 0.2	Neuron = 50 Epoch = 7000 Batch = 32 Dropout = 0.2
GBC/GBR	Binary	Trees = 100 Depth = 7 $\alpha = 0.0005$	
	Continuous	Trees = 500 Depth = 3 $\alpha = 0.01$	

## VI. STATISTICAL PERFORMANCE

### A. Binary Experiments

Table IV shows the statistical performance of every model for the binary classification framework applied to the market variables dataset and the fundamentals dataset.

TABLE IV  
BINARY CLASSIFICATION: ACCURACY (%)

Model	Market			Fundamentals		
	Train	Validate	Test	Train	Validate	Test
Logistic	62.5	55.2	53.0	67.8	39.1	44.9
Lasso	59.1	58.8	53.6	58.5	53.6	34.8
Ridge	60.1	61.8	54.2	59.0	53.6	37.7
SVM	59.1	60.0	56.0	65.4	53.6	40.6
NN	69.7	56.4	54.2	65.5	55.1	40.6
GBC	81.9	52.1	48.2			

Note: Best performance on test set marked in red.

The results provide evidence that market variables have a stronger forecasting power than fundamentals when it comes to classifying long/short signals. The largest test accuracy (56.0%) for the market variables was obtained by the SVM, while the maximum test accuracy (44.9%) is achieved by logistic regression for the fundamentals data.

There is, however, an important caveat when interpreting the results. Being a measurement of the fraction of predictions that we can correctly forecast, accuracy does not differentiate between true positives and true negatives. A successful trading strategy should exploit true positives and true negatives, while minimizing false positives and false negatives.

To discern between these cases, Fig. 2 shows the confusion matrix for the SVM model in the market variables dataset. The plot suggests a bad performance on the classification of short signals, as well as a prevalence of long predictions.

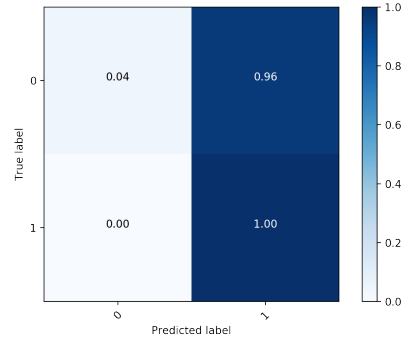


Fig. 2. Confusion matrix of the SVM model on the market variables dataset

We further explored why this would be the case, even after significant efforts were made to reduce overfitting via regularization. Fig. 3 shows the density of the standardized 3-month yield of Mexican Treasury Bills computed using kernel density estimation, conditional on the binary target variable. The plot provides evidence that both conditional densities are very similar, a pattern that we observed was recurrent across all features. This complicates the classification task and likely induces underperformance in short signals.

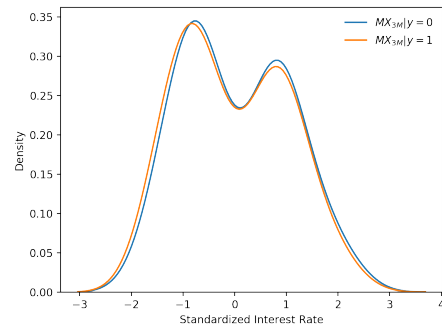


Fig. 3. Conditional density of 3-month Mexican T-Bills

## B. Continuous Experiments

Table V presents the statistical performance of every model for the continuous target framework applied to the market variables and the fundamentals datasets.

TABLE V  
CONTINUOUS TARGET: ACCURACY (%)

Model	Market			Fundamentals		
	Train	Validate	Test	Train	Validate	Test
Linear	65.3	65.9	58.8	54.5	55.9	50.0
Lasso	63.2	67.1	57.0	50.5	63.2	52.9
Ridge	63.6	67.1	60.0	52.0	52.9	50.0
SVR	67.3	56.7	58.2	55.9	45.6	54.5
NN	79.2	54.9	60.0	65.2	45.6	54.4
GBR	73.9	50.6	56.4			

Note: Best performance on test set marked in red.

The outperformance of the continuous variable target with respect to the binary classification models is significant. The improvement between the accuracy of the best performing models in the market variables test set is of around 7%, while of 21% for the fundamentals test set. All continuous target models outperform the binary classification in terms of accuracy and all market-variables models outperform fundamentals models.

Given the bad results of the confusion matrix for the binary classification problem, we explore the results of the continuous experiments. Fig. 4 shows the confusion matrix of the best performing model in terms of accuracy on the market variables data for the continuous variable framework, Ridge regression.

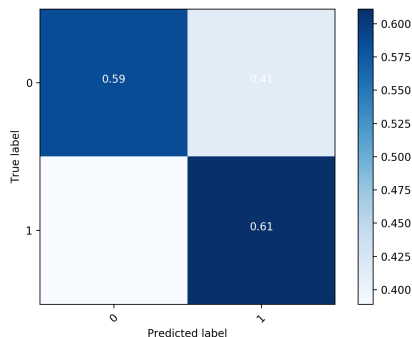


Fig. 4. Confusion matrix of the Ridge model on the market variables data

It is easy to observe that the change with respect to the continuous model is dramatic. From a 4% true negative rate obtained in the best model for binary classification, this new continuous target framework yields a 59% rate. This is obtained at the expense of a lower true positive rate. However, the true positive rate still yields a reasonable performance of 61%.

## VII. ECONOMIC PERFORMANCE

A model with very successful statistical performance of long/short signals does not imply positive economic implications. This is an inherent problem in directional forecasts.

A profitable investment strategy requires algorithms that correctly predict the direction of very large movements in the price of the asset. In our case, if an algorithm correctly predicts most small changes but misses large jumps in the exchange rate, it is very likely that it will produce negative economic performance upon execution. This issue has been previously assessed in the literature by Kim, Liao, and Tornell (2014).

Therefore, to assess the economic performance of our models, we compute the cumulative profits generated by the execution of the ML-generated strategy in the test set. The implemented strategy is simple: we start with enough cash in MXN to buy a unit of USD. We then execute the following for every time  $t$ :

$$\text{Strategy}_t = \begin{cases} \text{Long USD 1} & \text{if } \widehat{\text{Signal}}_t = 1 \\ \text{Short USD 1} & \text{if } \widehat{\text{Signal}}_t = 0 \end{cases}$$

At the end of every period, the position is closed, profits are cashed-in and the strategy is repeated. Finally, we use a long-only strategy as our benchmark for economic performance.

### A. Binary Classification

Fig. 5 plots the cumulative profits of executing the binary classification algorithms on the market variables dataset as a trading strategy.

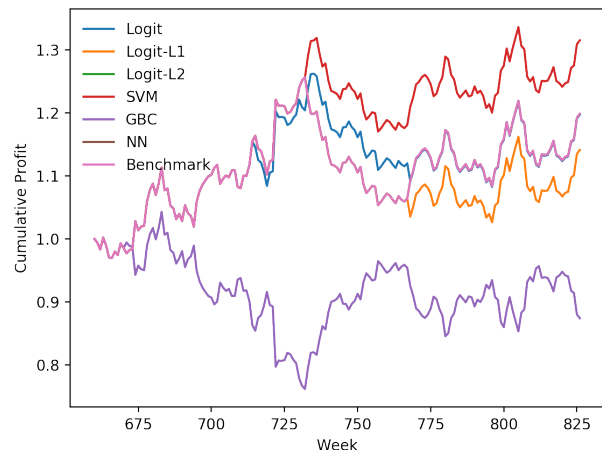


Fig. 5. USD cumulative profits of the market variables dataset

The statistically best performing model corresponds to the economically most profitable specification. However, it is important to notice that this positive result is mostly driven by a single correct bet made between weeks 725 and 750. All other strategies produce profits that are equal to or worse than the long-only benchmark.

These results can be explained by the bad performance of the models in terms of the confusion matrix. Due to the very low true negative rate of most models, all specifications are close to the long-only benchmark and the departures are a consequence of few correct or incorrect short bets.

## B. Continuous Variable Target

Fig. 6 plots the cumulative profits of executing the continuous variable target algorithms on the market variables dataset as a trading strategy.

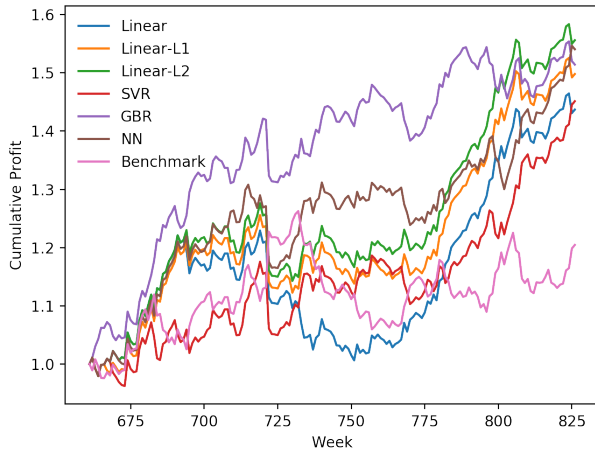


Fig. 6. USD cumulative profits of the market variables dataset

The differences with respect to the binary classification results are, once again, significant. The final cumulative return in the continuous target variable framework is around 15% higher than under the binary classification framework. Furthermore, all strategies outperform the long-only benchmark with the best strategy being Ridge regression.

In addition, the economic effect of an improved true negative rate is considerable. Unlike the binary classification case, the outperformance of all strategies with respect to the benchmark is not driven by few correct short positions. Moreover, the reduction in the true positive rate observed for the continuous target variable framework does not significantly penalize cumulative profits. The gains of a high specificity outweigh any losses derived from the reduction in sensitivity.

A natural question to address is which variables explain exchange rate forecasts the most. Fig. 7 shows the relative importance of the features in explaining FX dynamics.

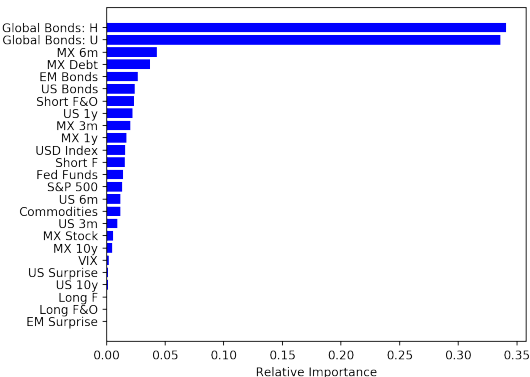


Fig. 7. Variable importance for ridge regression on the market variables dataset under the continuous target framework

It is no surprise that fixed income variables are the most relevant features. The result is consistent with the idea that the exchange rate is closely related to interest rates, as explained by the Uncovered Interest Rate Parity condition widely studied in economics.

Finally, another interesting insight is that the USDMXN reacts strongly to global and emerging-market (EM) fixed income indicators. In theory, the bilateral exchange rate should react strongly to the interest rate differential between the two countries. We believe the observed result provides evidence of investor behavior. As documented in recent years by Bloomberg (2015), The Wall Street Journal (2017a) and The Financial Times (2018), the high liquidity of the Mexican Peso has allowed its role as a hedge for long EM positions. Our results are consistent with these findings.

## VIII. CONCLUSION AND FUTURE WORK

This paper makes use of machine learning methods to forecast the US Dollar against Mexican Peso exchange rate. We use an innovative framework to find the best possible performance. First, we consider a market variables dataset and a fundamentals dataset on which we train ML algorithms. Second, we conduct binary classification experiments and continuous target experiments to produce the same output: a binary long/short signal on which we are able to execute a simple trading strategy.

Our results suggest that continuous target prediction outperforms binary classification not only in terms of accuracy, but also in terms of specificity and sensitivity. The economic results are in line with this finding, with all algorithms outperforming a long-only benchmark. The best results are produced by SVM in the binary classification case and Ridge regression in the continuous target case, both in terms of accuracy and cumulative profits. Last, we find that the fundamentals dataset yields poor results.

Future work could focus in several areas. First, the recursive validation procedure proposed in Gu et al. (2018) for time-series data could be implemented. This would allow to obtain classifiers and models that perform better out-of-sample. Second, a major improvement on model performance could be achieved through model ensembling. Finally, using more complex neural network models, such as LSTMs could increase the forecasting power of our features.

## IX. CONTRIBUTIONS

The team worked on the same problem but used different datasets. The contribution to this work was as follows:

**Christian González Rojas** was in charge of data collecting, data processing, algorithm selection and algorithm implementation on the market variables dataset for both the continuous and the binary framework. He decided to consider GBC/GBR as an additional model to further test the value of nonlinear relationships. He was also responsible for writing the CS229 poster and the CS229 final report. His data and code can be found at [this link](#).

**Molly Herman** worked on data collection, data processing and algorithms for the fundamentals dataset. She was responsible for modifying the CS229 poster to create an alternative version for the CS229A presentation and was in charge of writing her own final report for CS229A.

The division of work for the poster and the final report was done to provide deeper insight on the results to which each author contributed the most.

#### REFERENCES

- Amat, C., Michalski, T., & Stoltz, G. (2018). Fundamentals and exchange rate forecastability with simple machine learning methods. *Journal of International Money and Finance*, 88, 1 - 24.
- Bloomberg. (2015). *Why Traders Love to Short the Mexican Peso*.
- Connor, G., & Korajczyk, R. A. (1988). Risk and return in an equilibrium APT: Application of a new test methodology. *Journal of Financial Economics*, 21(2), 255 - 289.
- Fan, J., Liao, Y., & Wang, W. (2016, 02). Projected principal component analysis in factor models. *Ann. Statist.*, 44(1), 219–254.
- Gu, S., Kelly, B. T., & Xiu, D. (2018). Empirical Asset Pricing via Machine Learning. *Chicago Booth Research Paper, No. 18-04*.
- Hryshko, A., & Downs, T. (2004). System for foreign exchange trading using genetic algorithms and reinforcement learning. *International Journal of Systems Science*, 35(13-14), 763-774.
- Kelly, B., Pruitt, S., & Su, Y. (2018). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics, Forthcoming*.
- Kim, Y. J., Liao, Z., & Tornell, A. (2014). Speculators' Positions and Exchange Rate Forecasts: Beating Random Walk Models. *Working Paper*.
- Lettau, M., & Pelger, M. (2018). Factors that fit the time series and cross-section of stock returns. *Working Paper*.
- Ramkrishnan, S., Butt, S., Chohan, M. A., & Ahmad, H. (2017). Forecasting Malaysian exchange rate using machine learning techniques based on commodities prices. In *2017 International Conference on Research and Innovation in Information Systems (ICRIIS)* (p. 1-5).
- The Financial Times. (2018). *Mexico's Peso remains the bellwether for Emerging Markets*.
- The Wall Street Journal. (2017a). *The Mexican Peso: A Currency in Turmoil*.
- The Wall Street Journal. (2017b). *The Quants Run Wall Street Now*.