# Foreign Exchange Forecasting via Machine Learning

Christian González Rojas
cgrojas@stanford.edu

Molly Herman
mrherman@stanford.edu

## Introduction

Finance has been revolutionized by the increased availability of data, the rise in computing power and the popularization of ML algorithms. Despite the boom in data-driven strategies, the literature analyzing ML methods in financial forecasting has focused on stock return prediction. Our intention is to implement machine learning methods in a relatively unexplored asset class: foreign exchange (FX).

## Objective

The objective of this project is to produce FX forecasts that are able to yield profitable investment strategies. We approach the problem from two different perspectives:

1. Classification of long/short signals
2. Point forecasts of FX levels that yield long/short signals.

## Market Variables vs. Fundamentals

We use two different datasets to explore the forecasting power of two types of variables that we define as:

- *Market variables*: Indicators with daily to weekly frequencies that have a close relationship with traded securities.
- *Fundamentals*: Indicators with monthly frequencies that are closely related to the macroeconomy.

We limit to forecasting the USD vs. MXN exchange rate. Our data is sequentially split into train (60%), validate (20%) and test (20%).

## Features

All data was gathered from either Bloomberg, the Global Financial Dataset or the Federal Reserve Bank of St. Louis (FRED). We center and scale all of our features. Furthermore, we can divide our 25-27 features into the following categories:


**Fig. 1:** Time series of the USD vs. MXN rate.

**Market variables**
- Fixed Income
- Stock Market
- Currency

**Fundamentals**
- Economic Activity
- Labor Market
- Debt
- Sentiment

## Models

We employ the same models for the market variables and the fundamentals. The following frameworks are considered for classification/regression:

1. *Logistic/Linear Regression*: This serves as our baseline model.

$$P(y=1|x;\theta) = \frac{1}{1+e^{-\theta^T x}};$$

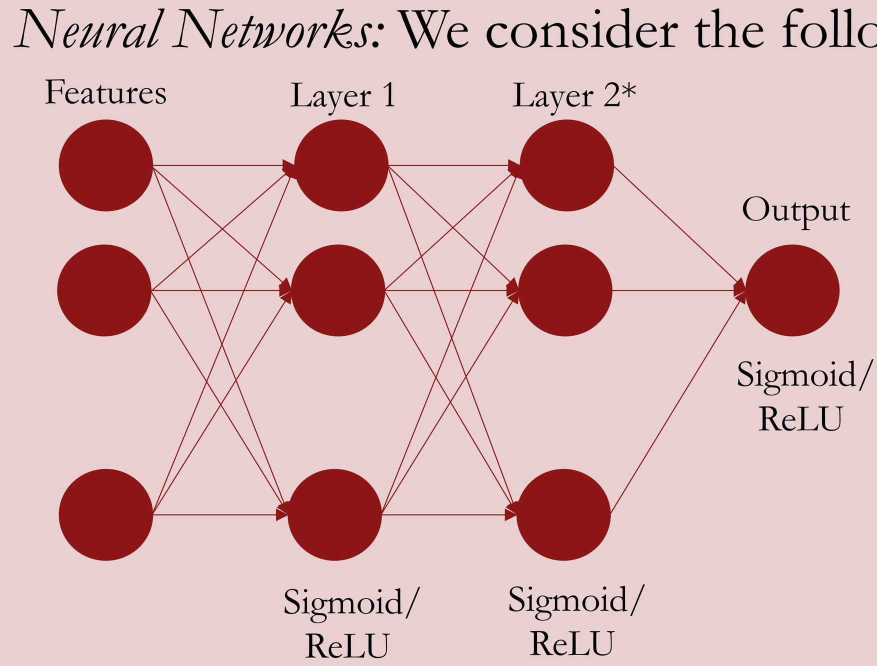
2. *Regularized Logistic/Linear Regression*: $L_1$ and $L_2$ regularization.

3. *Gradient Boosting Classifier/Regression:* We use GBC/GBR to capture non-linear relationships. Random Forests not appropriate since bootstrap does not preserve time-series nature of the data.

4. *Support Vector Machines/Regression:* We consider a Gaussian kernel. The non-linear boundary produced by the infinite-dimensional mapping could better capture FX dynamics.

$$K(x,z) = \exp\left(-\gamma \left\| x - z \right\|_2^2\right)$$

5. *Neural Networks:* We consider the following set-up:




* Only considered for fundamentals model.

**Market variable model**
- Logistic/MSE loss
- 255/500 neurons
- Dropout

**Fundamentals model**
- Logistic/MSE loss
- 100/50 neurons
- Regularized/Dropout

## Binary Experiments

Since we are interested in classifying long/short signals, we modify the target variable to a binary classification:

$$Signal_t = \mathbb{I}(USDMXN_{t+1} - USDMXN_t > 0)$$

Parameters are tuned in the validation set using accuracy.

## Continuous Experiments

We construct point-forecast model using the raw USDMXN data. We then modify the forecast output to produce long/short signals:

$$Signal_t = \mathbb{I}(\widehat{USDMXN}_{t+1} - \widehat{USDMXN}_t > 0)$$

Parameters are tuned in the validation set using MSE.

## Statistical Performance

**Table 1**: Accuracy (%)

| | Binary Experiments | | | | | | Continuous Experiments | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Market Variables | | | Fundamentals | | | Market Variables | | | Fundamentals | | |
| | Tr | V | T | Tr | V | T | Tr | V | T | Tr | V | T |
| **Logit/LR** | 62.5 | 55.2 | 53.0 | 67.8 | 39.1 | $\boxed{44.9}$ | 65.3 | 65.9 | 58.8 | 54.5 | 55.9 | 50.0 |
| **Lasso** | 59.1 | 58.8 | 53.6 | 58.5 | 53.6 | 34.8 | 63.2 | 67.1 | 57.0 | 50.5 | 63.2 | 52.9 |
| **Ridge** | 60.1 | 61.8 | 54.2 | 59.0 | 53.6 | 37.7 | 63.6 | 67.1 | $\boxed{60.0}$ | 52.0 | 52.9 | 50.0 |
| **SVM** | 59.1 | 60.0 | $\boxed{56.0}$ | 65.4 | 53.6 | 40.6 | 67.3 | 56.7 | 58.2 | 55.9 | 45.6 | $\boxed{54.5}$ |
| **NN** | 69.7 | 56.4 | 54.2 | 65.5 | 55.1 | 40.6 | 79.2 | 54.9 | 60.0 | 65.2 | 45.6 | 54.4 |
| **GB** | 81.9 | 52.1 | 48.2 | - | - | - | 73.9 | 50.6 | 56.4 | - | - | - |

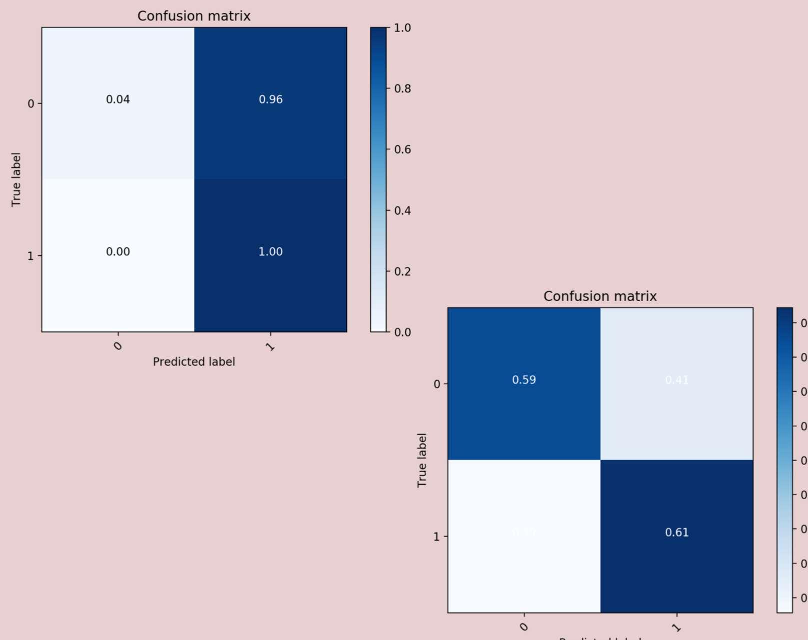



**Fig. 2:** Confusion matrix: Binary market SVM (upper) and continuous market Ridge (lower)

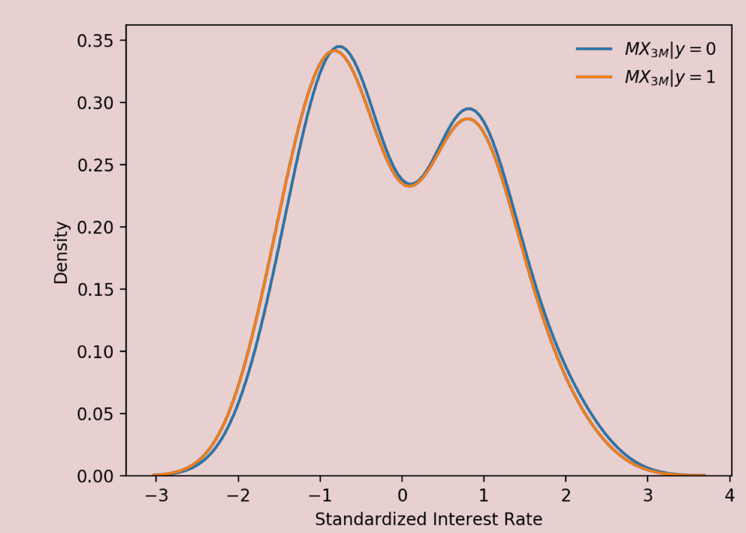



**Fig. 3:** Density plot of 3-month MX yield conditional on binary target

## Economic Performance

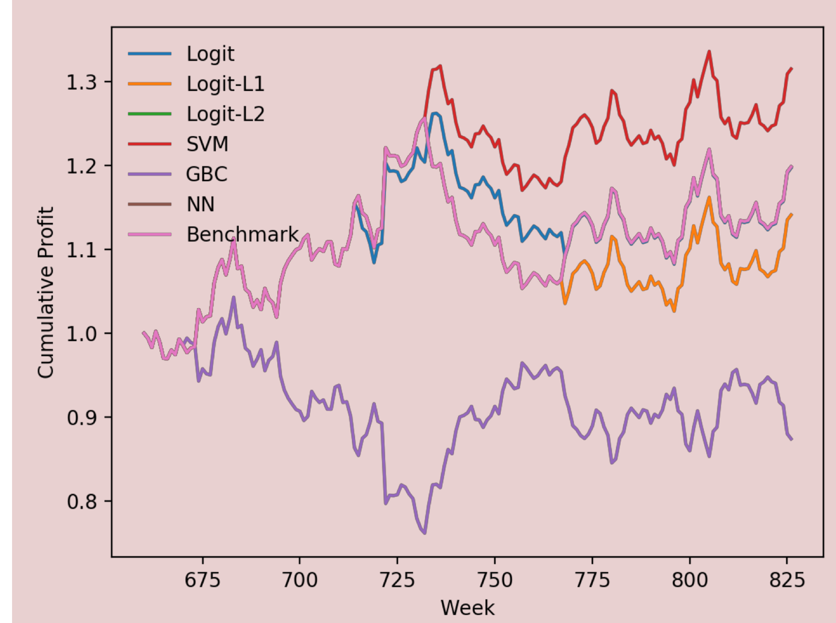



**Fig. 4:** Cumulative profits of binary market variable model

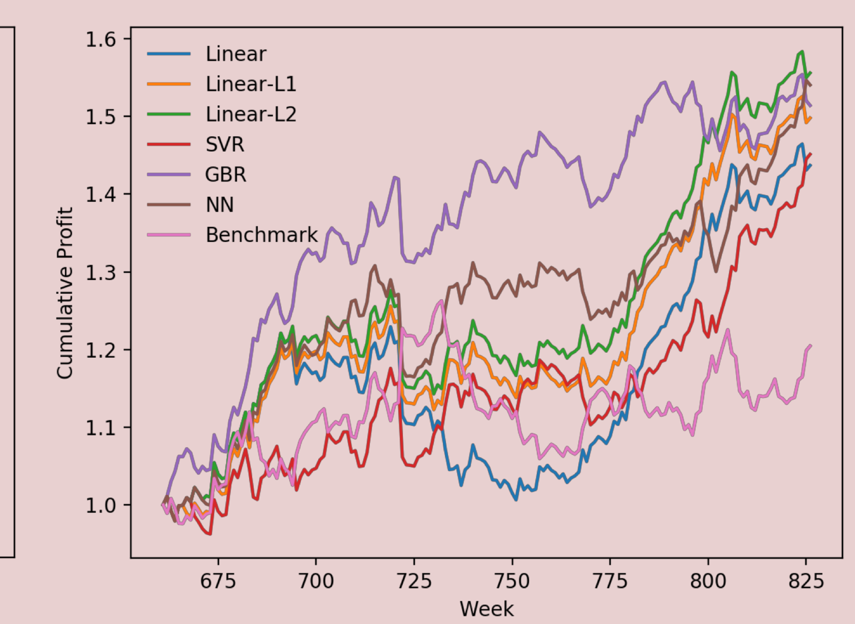



**Fig. 5:** Cumulative profits of continuous market variable model

## Conclusion and Future Work

ML methods are promising for FX forecasting, with continuous variable models outperforming binary classification. In addition, market variables better explain FX movements. Future work should explore improvement via recursive validation and ensemble methods.

**References**: Gu, S., Kelly, B. T., & Xiu, D. (2018). Empirical Asset Pricing via Machine Learning. *Chicago Booth Research Paper*, No. 18-04.