Final Project: Modeling Foreign-Exchange Rates

David Modjeska and Dominic Murphy¶

CSCI E-82, Harvard Extension School, Autumn 2017

Abstract

FX rates are challenging for prediction, but potentially lucrative and intrinsically interesting. In this report, we present a project to predict FX rates for the world's six most-traded currencies against the USD: AUD, CAD, CHF, EUR, GBP, and JPY. Data on major economic indicators was obtained from the Federal Reserve Bank of St. Louis, the Energy Information Administration, Quandl, and *The Economist*. We fitted prediction models for both regression and classification, using classic machine learning (including ARIMA, GAM, random forest, boosting, and SVM) and neural networks (including LSTM and FFNN). Results showed that regression on rates couldn't overcome the 'random walk' in seven-day FX rate movements. Classification of the direction of FX rate movement six months in advance, however, resulted in an average ROC AUC score near 80%. In general, the most resource-intensive economies offered the best predictions, while the least resource-intensive economies offered the worst predictions. Methodologically, a robust and streamlined software framework (in Python) was needed for efficient exploration of the modeling space.

Introduction

Foreign exchange (FX) allows goods and services to move around the world. The FX market is global and decentralized. It is the world's largest market, with the main participants being large banks. The FX market's three main participants are traders, corporate CFOs, and governments. For traders, rates support hedging risk, pursuing profit, pricing FX derivatives, and algorithmic trading. For CFOs, rates support assessing risk and cross-border ventures. For governments, rates support budgeting and setting monetary policy.

Given the criticality of FX rates to the global economy, and the sizeable profit-making opportunities, both explanation and prediction are interesting and valuable. Explanations offer the promise of helping with key tasks above. Prediction offers more -- the hope of profit-making.

This report will focus on the exchange rate between US and the world's other six most heavily-traded currencies by volume. The US is a major trading partner for much of the world, and the USD is often seen as the benchmark on global markets. So it will be the base currency in our work. Key questions in this domain revolve around currency-pair rates (e.g., USD/CAD FX), both direction and magnitude of rate movements.

Research Questions

As we consider the issues above, they can be focused into five research questions driving the bulk of this work:

1. Which feature categories add the most value for each machine-learning model type? (Also, for each currency, which features are the most important for predicting FX rates?)
2. Which major machine-learning technique is the most effective for predicting FX rates?
3. What is the relative effectiveness of prediction at different time lags (e.g., day, week, month, year)
4. How effective is the best machine-learning technique for market use?

To answer the questions above and to further our educational goals for this project, we adopted a replication-oriented approach. Generally consistent data acquisition, modeling framework, and results presentation are adopted to allow comparability of results. We utilized both regression and classification methods in addressing the research questions above.

Related Research

Previous research around predicting FX rates falls into two broad areas: economic theory and machine learning. Two useful selections from this research are discussed below.

First, the economic theory of purchasing power parity (Taylor 2004) suggests that if arbitrage exists with a common good between two countries, the long-term exchange rates move to equalize the cost of this good. This idea motivated the *The Economist* to create their Big Mac index. In our feature set, we included this concept as a ratio showing whether a currency is over- or undervalued relative to USD.

Second, SVM models are common in the FX field. (Kamruzzaman, 2003) found that a polynomial kernel was particularly effective for predicting with SVMs. In this work, we replicate the application of a polynomial SVM kernel to capturing FX rate direction.

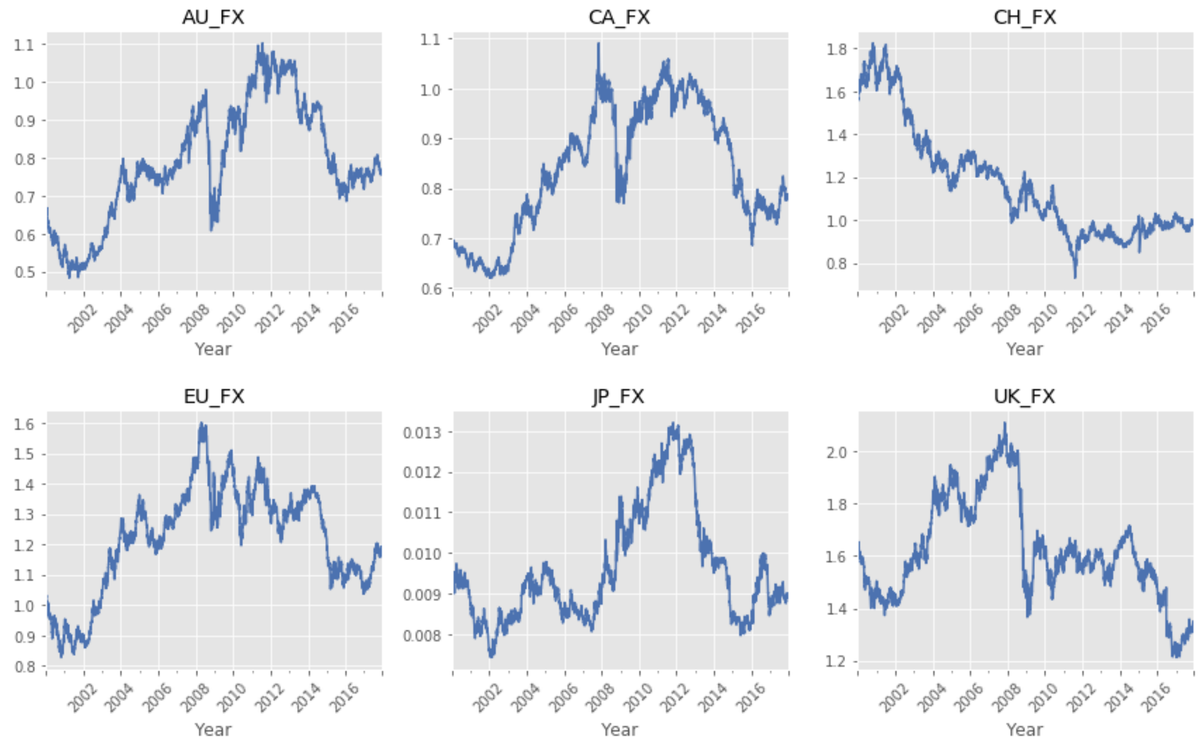
A general approach and some implementation details with using LSTM in time series analysis were guided by the author Jason Brownlee (2017).

FX Rates

To ground this work, a time-series visualization of the relevant FX rates is shown below. Several of the rates move in tandem (e.g., AUD and CAD), while others have quite different patterns (e.g., CHF). These differences may represent the fundamentally different natures of the economies concerned, whether based primarily on resources, manufacturing, or services.

Methodology

The two pillars of our work were data acquisition and data modeling. For data acquisition, we acquired a range of predictors from canonical sources, in order to consider as many perspectives as possible. For modeling, we explored both regression and classification algorithms, in order to re-trace some of the exploration involved in FX prediction during recent years.



FX Rates Relative to USD (2000-2017)

Data Acquisition

In selecting the feature set, we collected data needed to support common trading strategies and economic theories.

The top 7 currencies by trading volume account for approximately 82% of global flows. For analysis, we therefore selected the top 6 flows relative to the US dollar (USD):

- Australian Dollar (AUD)
- Canadian Dollar (CAD)
- Swiss Franc (CHF)
- Euro (EUR)
- English Pound (GBP)
- Japanese Yen (JPY)

The starting date for this analysis is January 1st, 2000, as this date coincides with the introduction of the Euro -- a fundamental shift in FX trading patterns.

Data was sourced from reliable, standard, market sources:

- The Federal Reserve Bank of St. Louis (FRED)
- U.S Energy Information Administration (EIA)
- Quandl
- *The Economist*

Most of these sources (except for *The Economist*) are accessible via APIs that support automated data refresh.

Major Indicators for Foreign Exchange Rates

Indicator	Effect	Rationale
Interest Rates	Interest rate up → currency up	Higher interest rates relative to other countries have a positive effect on currency value.
Gold	USD up → gold down	Gold is seen as safe store of value and often moves inversely with USD.
Unemployment Rate	Unemployment up → currency down	Increase in unemployment is a negative indicator for currency value.
Gross Domestic Product (GDP)	GDP up → currency up	A difference in GDP growth between countries indicates a difference in economic strength
Consumer Price Index (CPI)	CPI up → currency down	High CPI indicates that the cost of production is rising.
Government Debt	Debt up → currency down	Large government debt increases risk in the national currency
US Trade Weighted Index (TWI)	US TWI up → USD up	Purchasing power of USD relative to a basket of goods
Loan Defaults	Defaults up → currency down	Indicator of relative economic health
Energy Prices	Prices up → petrocurrencies up	High energy prices benefit resource economies, while harming other economies that require this energy
FX Futures	N/A	Indicates market expectations relative to current FX
Moving FX Averages	N/A	Rolling averages of FX rates
Big Mac Index (PPP)	N/A	Indicates relative over/under-valuation of currency
Public holidays	N/A	Little market movement is expected on holidays
Season and month	N/A	Seasonal or monthly market variations may be present in the data

Feature names for these indicators were harmonized to support the software framework discussed below.

Data Modeling

Broadly speaking, modeling of FX rates can proceed as either regression or classification. Regression has the advantage of offering both rate sign and magnitude, potentially, for market investment. Information signal may be hard to find in the data, however. Classification lacks rate magnitude, but even rate direction can help to steer major market decisions. Signal seems easier to find than with regression.

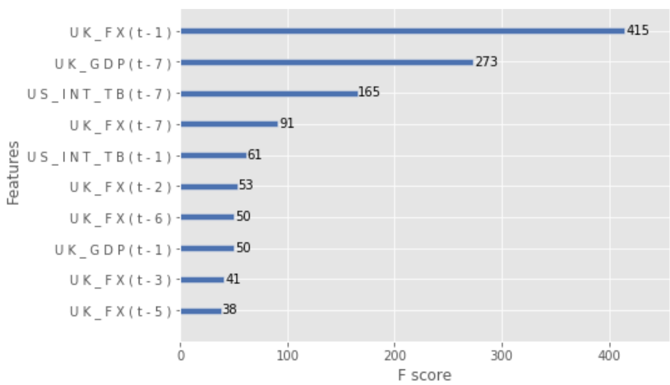
Given the range of relevant modeling options available, it was essential to develop a general software framework and pipeline for efficiency. For this purpose we implemented a master notebook in Python, which we used as a template for rapid expansion of our model set. This notebook included the six currencies evaluated in this work. The framework contained code to reframe features as predictors; to scale, reduce, split, and extract data; to fit models; to score models; and to visualize feature importances (where available), actual/predicted change, and test/naive score. The configurable elements in this framework included:

- Features to use in each model
- Past time lags (1-365 days)
- Future time lags (1-365 days)
- Regression vs. classification
- Scaling - range vs. standard
- PCA - yes/no
- Shuffling data - yes/no
- Ratio of training to testing data
- Algorithm - GAM, random forest, boosting, LSTM, FFNN, SVC
- Individual vs. stacked models
- Scoring metric
- Visualization of process
- Visualization of result

As an aside, our initial modeling didn't use the framework above, but instead used automatic ARIMA in R. Each ARIMA model incorporated one of the FX rate time series. The models were configured to require stationarity, and to seek seasonality where relevant.

Leveraging input from the course staff and from the research literature, we divided our modeling into two broad categories: regression and classification. For regression with deep learning, the most appropriate technique for time series is LSTM. For regression with classic machine learning, effective techniques include generalized additive models (GAMs), random forest regressors, and boosting regressors. These regression techniques can benefit from enhanced features such as moving averages and purchasing-power parity. A stacked ensemble can leverage the classic ML algorithms. For classification of time series, SVC is often favored, and LSTM and FFNN are worth trying as well.

GBP/USD - XGBoost Model - Top 10 Feature Importances



Different models and currencies benefitted from different features, as expected. For all models, the basic FX rate was the most important feature for prediction. A sample feature importances graph from the XGBoost model for GBP/USD demonstrates this. For some other models, the difference in importance was even stronger.

Past lags were constructed by time shifting each of the *k* predictors forward once for each of the *p* days in the past lag, for a total of $k * p$ predictors. The future lag was constructed by time shifting the FX rate column *f* days backward to create a single response variable. Resulting rows containing null values were removed. (For details and examples, please see the implementation notebooks.)

For all regression in the framework, we used past and future lags of one week. This period allowed us to avoid the challenge of the one-day 'random walk' in FX rates, while restricting data contamination with the classic ML algorithms. For classification, we chose a longer period - six months - to investigate the impact of purchasing power parity (PPP), following the literature. Research on PPP suggests that despite arbitrage and other deviations from historical norms, rates tend to gravitate towards a long-term average.

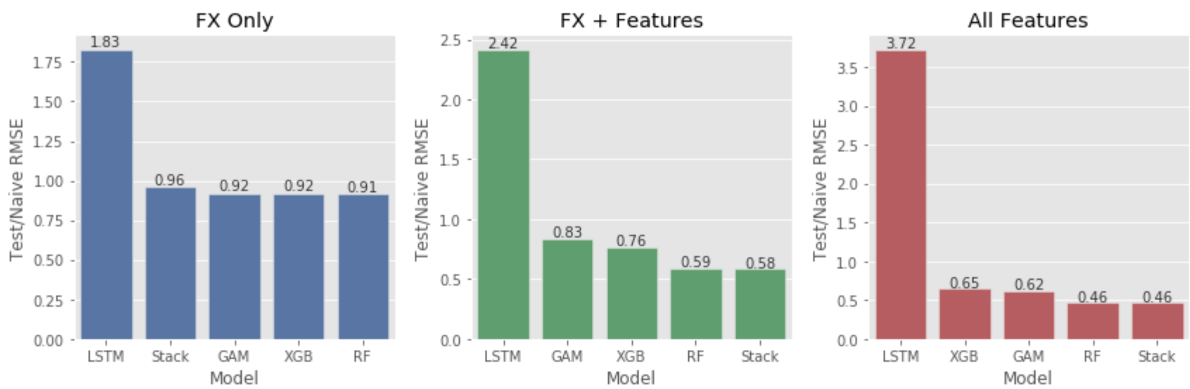
Results

Results from the data modeling are presented below in two subsections - regression and classification. On the basis of variations mentioned with the software framework above, these results cover 6 currencies * 8 algorithms * 3 feature sets = 144 individual models. (Individual results are provided in the appendices.) All results refer to a testing data set.

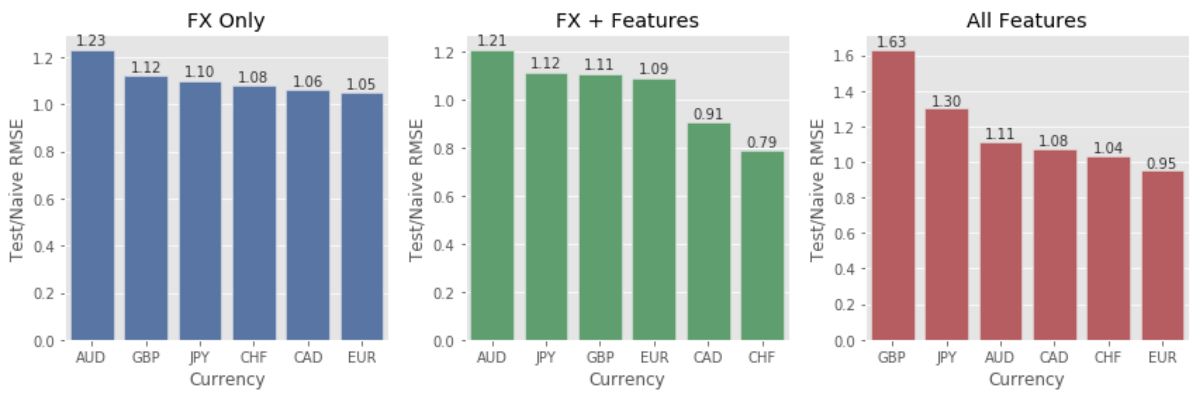
Regression

Results showed that the classic regression methods XGboost, GAM and random forest substantially outperformed LSTM. These classic methods all scored relatively similarly to each other for a given feature set: 92%-96% of naive error using FX only, 58%-83% of naive error using FX and other features, and 46%-65% of naive error using all features. Because of these scores, our stacked ensemble omitted the LSTM model.

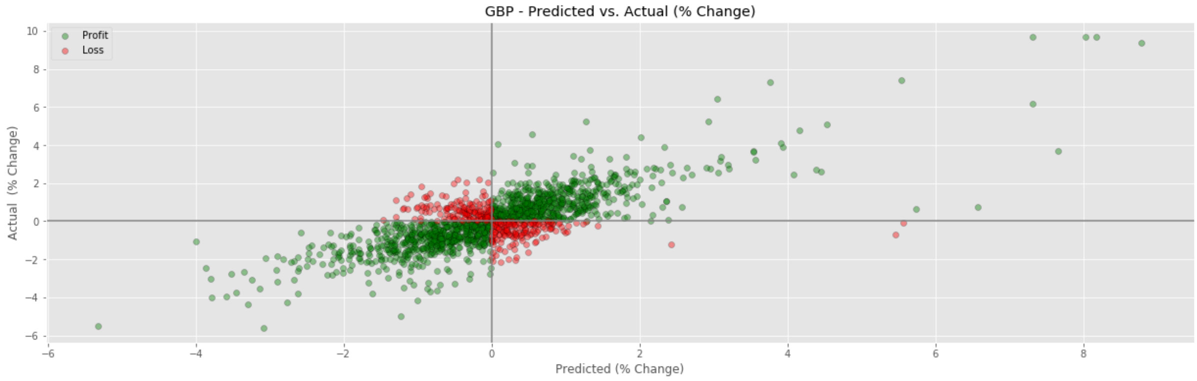
Average test/naive RMSE by model (7 Days)



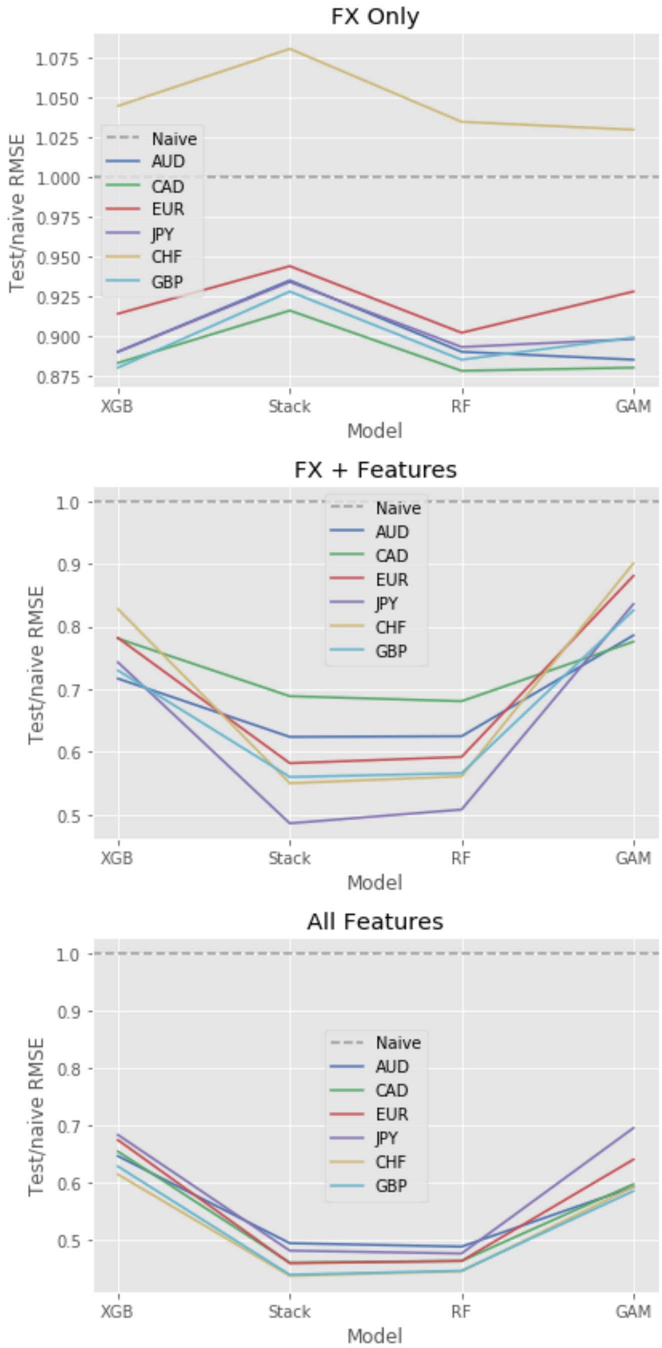
Average test/naive RMSE by currency (7 days)



GBP - XGBoost Model - Predicted vs. Actual (7 Days % Change)



Regression Results by Currency (7 Days)



The above graph shows XGBoost prediction with the GBP/USD FX rate (using FX and other features). This graph compares the predicted and observed percentage change during a seven-day period. Quadrants with green points show where rate direction was correctly predicted. The red points indicate where the predicted direction was incorrect. The visible preponderance of green points indicates that in most cases, XGBoost correctly predicted direction. We also note that when predicted change is large (> |2%|), the predicted direction was almost entirely correct. In practical terms, if the model predicted a change of > 2% either up or down, an investor could be reasonably confident of breaking even or making a small profit.

In reviewing the three graphs on the right, it is clear that adding features reduced RMSE for all models in general. (This observation is most pronounced with CHF.) Using all features, XGB and GAM scored between 60% and 70% of naive error, while random forest did better with 40%-50% of naive error. Stacking didn't worsen scoring, but stacking didn't exceed its best component model.

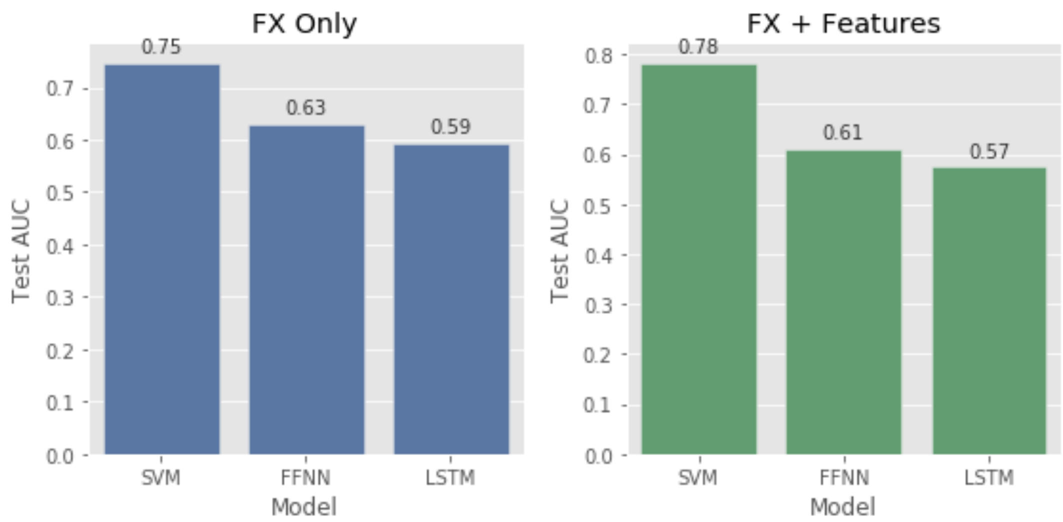
ARIMA

Before leaving regression, let's note that ARIMA failed to achieve the same level of performance as our predictor/response regression models. ARIMA results showed AIC and BIC scores in the range of 15,000-35,000. While models can't be compared directly across such variation, the AIC and BIC scores achieved by GAM, random forest, and XGBoost regressors were approximately 15. This represents a difference of three orders of magnitude in favor of the predictor/response style models. ARIMA may have potential for long-term modeling in this domain, and further investigation is suggested.

Classification

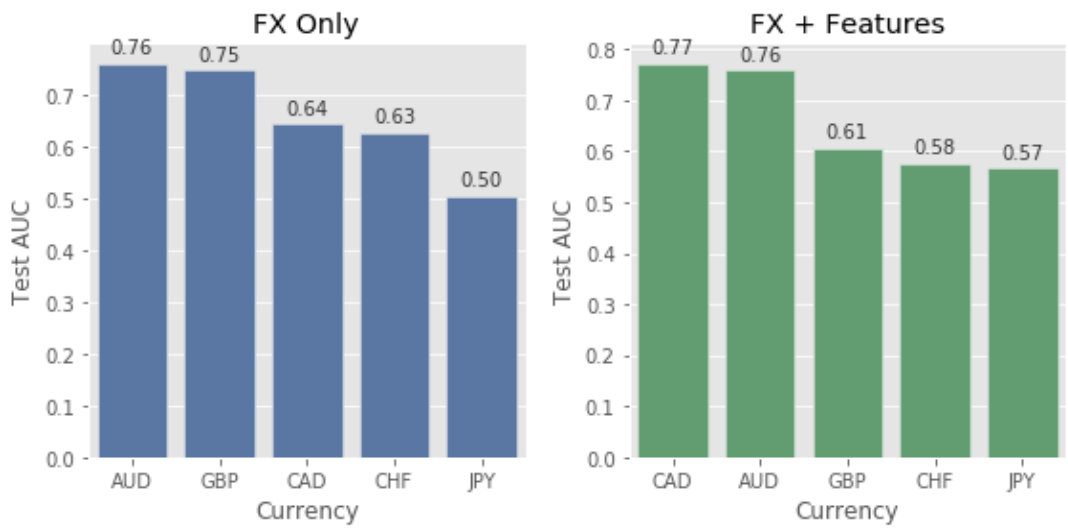
Of the three classifiers, SVM (with a polynomial kernel) scored best, with an average AUC of 0.78 using all features. FFNN scored next best with an average AUC 0f 0.62, and LSTM performed somewhat worse with an average AUC of 0.57. Results are relatively similar using FX and other features.

Classification Results by Model Type (7 Days)



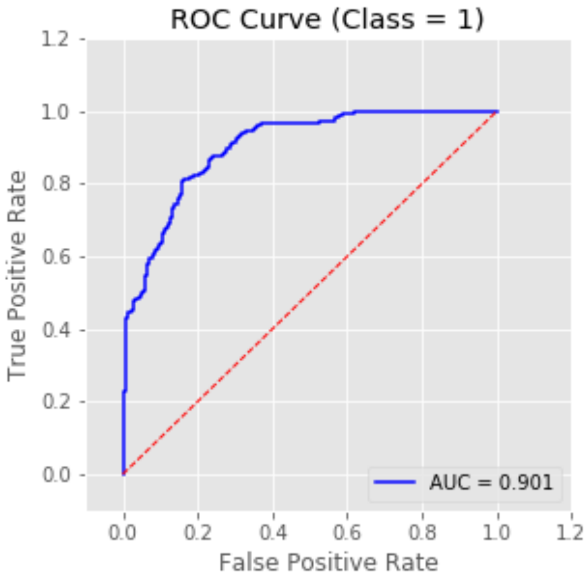
While results were relatively consistent across currencies, AUD showed the most success in predicting the direction of rate movements.

Classification Results by Currency (7 Days)



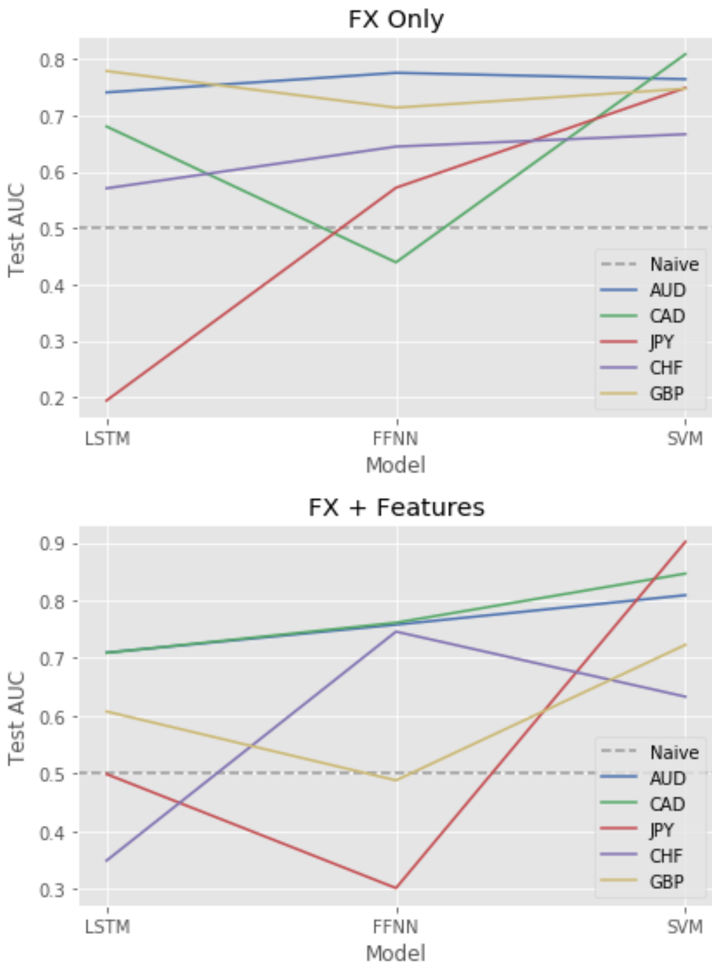
ROC Curve - JPY with SVM - FX + PPP

(180 Days)



As an example of successfully predicting direction with SVM, the ROC curve to the left shows results for the JPY/USD FX rate. Features included in this model were the FX rate itself and the PPP ratio.

Classification Models for all Currencies (180 Days)



As noted above, SVM consistently outperformed LSTM and FFNN. Individual currencies were problematic for specific neural networks - JPY for FFNN generally, and CHF for LSTM with FX + Features. As it turns out, adding PPP as a feature caused some scores to worsen. The sheer inconsistency of NN performance here is worth noting, especially for a potential investor. (Further research into some of these inconsistencies might be interesting.)

Unlike with regression models, these classification models seem to be sensitive to the addition of features.

Regression vs. Classification

Results showed that regression on rates couldn't overcome the 'random walk' in seven-day rate movements. By contrast, classification of the direction of rate movement six months in advance resulted in a reasonable average ROC AUC score around 80%.

Discussion

Predicting FX rates is widely acknowledged to be a difficult problem. (For example, two Australian banks have recent forecasts in opposite directions for a twelve-month time frame.) The primary difficulty is that daily FX rate movements follow a random walk. Addressing this challenge requires making trade-offs between length of feature history, distance of future prediction, and timeline splitting between training and testing data. For regression, we found that no combination of these trade-offs allowed a fitted model to beat a naive model, without some potential contamination of the data from different time periods. For classification, we found that the farther into the future that a prediction was made, the more that a fitted model could beat a naive model. With classification, the important thing was to preserve a large enough testing set to robustly validate the predictions.

A secondary difficulty in addressing FX rate prediction is the wide range of options available - how should one structure the inquiry? Our approach was to include both regression and classification methods, for completeness, being sure to try techniques suggested by the research literature. We also developed a robust and streamlined modeling framework using Python, which enabled rapid exploration of the space of modeling options. Finally, we obtained a wide range of economic indicators, carefully categorized, to support flexible consideration for possible feature engineering.

During this work, each model was fit three times for each currency - once with only the FX rate, once with the FX rate and other hand-chosen features, and finally with all features (using PCA for data compression). Compromises were necessary in tuning hyperparameters that applied across a given set of models, e.g., all currencies and features variations for XGBoost. For optimal scoring and/or market application, it would be necessary to tune each model individually for each currency and each feature combination. Such fine grain tuning was beyond the scope of this inquiry.

On the subject of different currencies, it was striking how much the currencies differed during modeling. The primary difference was between the most resource-intensive economies (i.e., Canada and Australia) and the least resource-intensive economies (i.e., Japan and Switzerland). Techniques that worked well for one type of economy often failed for the other type. As it turns out, SVM classification was the most robust technique for bridging these differences.

While this work included both regression and classification approaches, classification clearly scored more successfully. Regression always remained borderline in its ability to detect information signal, while three classification algorithms clearly detected signal with varying degrees of success. To make money in the market doesn’t require perfect prediction, fortunately - it just requires gaining an edge over other market players. The detection of FX rate movement described in this report is the sort of information that could gain one such an edge, depending on the knowledge and tools of other players. So the regression results in this report are interesting from an analytical or academic perspective, and the classification results offer that and the possibility of market application.

Conclusion

The work presented in this report addressed FX rate prediction for seven major currencies over the period 2000-2017. Both regression and classification models were used from classic machine learning and neural networks. While precise prediction of rates proved elusive, substantially correct prediction of rate movements over a six-month period was successful. This result is academically interesting, potentially valuable in the market, and promising for future research.

Future research might pursue one of several directions. First, tuning models individually has definite potential to improve scoring. Such work is labor intensive, but necessary for any concrete application. Second, research literature has mentioned the usefulness of CNN's for classification, so this avenue could be explored. Third, new ways of structuring the time-series problem might allow greater success by previously-used or other models. Finally, additional features could enrich any model - such features might potentially include media feeds for sentiment analysis, impacts of major political or economic events (with consideration for impact decay rate), and additional indicators for manufacturing-intensive countries (such as Japan).

Research with FX rate prediction is distributed among financial services organizations, universities, and technology companies. Incentives for publishing results vary among these research venues, which can make information dissemination difficult. One suspects that effective techniques exist in-house, privately, that might solve or shed light on the problems discussed in this report. For the moment, we can conclude with some satisfaction with having made at least a small step in the right direction.

References

Machine Learning

Maknickienė, N. and Maknickas, A., 2012, May. Application of neural network for forecasting of exchange rates and forex trading. In The 7th international scientific conference" Business and Management (pp. 10-11).

Taylor, A.M. and Taylor, M.P., 2004. The purchasing power parity debate. The Journal of Economic Perspectives, 18(4), pp.135-158.

Kamruzzaman, J., Sarker, R.A. and Ahmad, I., 2003, November. SVM based models for predicting foreign currency exchange rates. In Data Mining, 2003. ICDM 2003. Third IEEE International Conference on (pp. 557-560). IEEE.

Eberhard Schoeneburg, PWC, HK 2017, *Financial Time-Series Predictions and AI Models (Part 1): Deep Learning - or not ?*, Linkedin accessed 2 December 2017 <www.linkedin.com/pulse/financial-time-series-predictions-ai-models-part-1-deep-schoeneburg/>

The Economist 2017, *The Big Mac index*, accessed 7 December 2017 <www.economist.com/content/big-mac-index>

Jason Brownlee 2017, *Multivariate Time Series Forecasting with LSTMs in Keras*, Machine Learning Mastery, accessed accessed 26 Nov 2017 <https://machinelearningmastery.com/multivariate-time-series-forecasting-lstms-keras/>

Quant Corner 2014, *Downloading EIA's data with Python*, accessed 27 Nov 2017 source: <https://quantcorner.wordpress.com/2014/11/18/downloading-eias-data-with-python/>

Finance

Currency Shop, The Forecast for the Australian Dollar in 2018, <http://www.thecurrencyshop.com.au/guides/australian-dollar-aud-forecast>, accessed Dec 14 2017

Data Sources

Board of Governors of the Federal Reserve System (US), U.S. / Economic Indicators, retrieved from FRED, Federal Reserve Bank of St. Louis; https://research.stlouisfed.org/docs/api/fred/, December 10, 2017.

U.S. Department of Energy, Energy Information Administration, Oil Futures Data. retrieved from Energy Information Administration <http://api.eia.gov/>, December 10, 2017.

Quandl, Futures Trading Data, retrieved from Quandl; <https://www.quandl.com/tools/api>, December 10, 2017.

The Economist, Big Mac Index, retrieved from The Economist, <http://infographics.economist.com/2017/databank/BMFile2000toJul2017.xls>, December 10, 2017.

U.S National Centers for Environmental Information (NCEI), weather data, <https://www.ncdc.noaa.gov/> accessed on Dec 1, 2017

Appendix 1: Test RMSE for all FX rates (relative to USD)

	Lag	Model	Currency	Naive	FX Only	FX + Features	All Features
0	7	Stack	AUD	0.0155	0.0145	0.0097	0.0077
1	7	Stack	CAD	0.0121	0.011	0.0083	0.0056
2	7	Stack	EUR	0.0181	0.0171	0.0105	0.0083
3	7	Stack	JPY	0.0001	0.0001	0.0001	0.0001
4	7	Stack	CHF	0.0165	0.0179	0.0091	0.0072
5	7	Stack	GBP	0.0229	0.0213	0.0128	0.0101
0	7	GAM	AUD	0.0155	0.0137	0.0122	0.0092
1	7	GAM	CAD	0.0121	0.0106	0.0094	0.0072
2	7	GAM	EUR	0.0181	0.0168	0.0159	0.0116
3	7	GAM	JPY	0.0001	0.0001	0.0001	0.0001
4	7	GAM	CHF	0.0165	0.017	0.0149	0.0098
5	7	GAM	GBP	0.0229	0.0206	0.0189	0.0134
0	7	XGB	AUD	0.0155	0.0138	0.0111	0.01
1	7	XGB	CAD	0.0121	0.0106	0.0094	0.0079
2	7	XGB	EUR	0.0181	0.0166	0.0142	0.0122
3	7	XGB	JPY	0.0001	0.0001	0.0001	0.0001
4	7	XGB	CHF	0.0165	0.0173	0.0137	0.0102
5	7	XGB	GBP	0.0229	0.0201	0.0167	0.0144
0	7	RF	AUD	0.0155	0.0138	0.0097	0.0076
1	7	RF	CAD	0.0121	0.0106	0.0082	0.0056
2	7	RF	EUR	0.0181	0.0163	0.0107	0.0084
3	7	RF	JPY	0.0001	0.0001	0.0001	0.0001
4	7	RF	CHF	0.0165	0.0171	0.0093	0.0074
5	7	RF	GBP	0.0229	0.0203	0.013	0.0102
0	7	LSTM	AUD	0.0104	0.0266	0.0343	0.0347
1	7	LSTM	CAD	0.0087	0.0152	0.014	0.0279
2	7	LSTM	EUR	0.0135	0.021	0.0353	0.0338
3	7	LSTM	JPY	0.0001	0.0002	0.0004	0.0005

4	7	LSTM	CHF	0.0156	0.0187	0.0171	0.0481
5	7	LSTM	GBP	0.0183	0.037	0.0526	0.111

Appendix 2: Test/naive RMSE for all FX rates (relative to USD)

	Lag	Model	Currency	FX Only	FX + Features	All Features
0	7	XGB	AUD	0.89	0.717	0.647
1	7	XGB	CAD	0.883	0.781	0.655
2	7	XGB	EUR	0.914	0.782	0.675
3	7	XGB	JPY	0.89	0.743	0.684
4	7	XGB	CHF	1.045	0.828	0.615
5	7	XGB	GBP	0.88	0.73	0.629
0	7	LSTM	AUD	2.563	3.301	3.339
1	7	LSTM	CAD	1.744	1.602	3.197
2	7	LSTM	EUR	1.558	2.618	2.508
3	7	LSTM	JPY	1.885	3.009	4.163
4	7	LSTM	CHF	1.199	1.101	3.087
5	7	LSTM	GBP	2.019	2.866	6.054
0	7	Stack	AUD	0.935	0.624	0.495
1	7	Stack	CAD	0.916	0.689	0.461
2	7	Stack	EUR	0.944	0.582	0.46
3	7	Stack	JPY	0.934	0.486	0.482
4	7	Stack	CHF	1.081	0.55	0.438
5	7	Stack	GBP	0.928	0.56	0.44
0	7	RF	AUD	0.89	0.625	0.489
1	7	RF	CAD	0.878	0.681	0.464
2	7	RF	EUR	0.902	0.592	0.464
3	7	RF	JPY	0.893	0.508	0.477
4	7	RF	CHF	1.035	0.561	0.446
5	7	RF	GBP	0.885	0.566	0.447
0	7	GAM	AUD	0.885	0.786	0.593
1	7	GAM	CAD	0.88	0.776	0.598
2	7	GAM	EUR	0.928	0.881	0.641

3	7	GAM	JPY	0.898	0.836	0.696
4	7	GAM	CHF	1.03	0.901	0.593
5	7	GAM	GBP	0.899	0.826	0.586

Appendix 3: Test AUC for FX rates (relative to USD)

	Lag	Model	Currency	FX Only	FX + Features
0	7	LSTM	AUD	0.7407	0.7094
1	7	LSTM	CAD	0.6802	0.709
2	7	LSTM	JPY	0.194	0.499
3	7	LSTM	CHF	0.5705	0.3492
4	7	LSTM	GBP	0.7785	0.6075
0	7	FFNN	AUD	0.7754	0.7579
1	7	FFNN	CAD	0.4392	0.7613
2	7	FFNN	JPY	0.5715	0.3015
3	7	FFNN	CHF	0.6445	0.7459
4	7	FFNN	GBP	0.7136	0.4882
0	7	SVM	AUD	0.7641	0.8089
1	7	SVM	CAD	0.8084	0.8464
2	7	SVM	JPY	0.7486	0.9015
3	7	SVM	CHF	0.6664	0.6331
4	7	SVM	GBP	0.7472	0.723