24 Apr 19

Daily Currency Price prediction using Daily Macroeconomic Data by applying Regression (Random Forest, SVR, ANN) Algorithms - Results Summary

Objective

Observe and improve the results of running regression algorithms (Random Forest Regression, Support Vector Regression, Artificial Neural Networks) on daily macroeconomic data to predict daily currency prices for GBPUSD, EURUSD and USDINR.

Note** – **This is a follow up research from the earlier research article** MacroCcyRegressionRandomForestResultsSummary_3Apr18* where random forest regression was applied to predict *average monthly currency prices using monthly macro data along with a time lag* between them. ***This research considers daily macro and currency prices without any time lag and applies different regression algorithms to them.

Data

- UK, EU, India and US ***daily macroeconomic data*** for last 19 years (Jan 00 – Feb 19) was used as the independent variable (X) and daily currency price over the same period for GBPUSD, EURUSD and USDINR was the dependent variable (Y).
- Total no of X variables and Y values varies depending on country and available data
    - X = 77, 78, 60 for EU, UK, India
    - Y = 6805, 6804, 5769 for EURUSD, GBPUSD, USDINR
- Core macroeconomic data (interest rate, inflation, GDP, unemployment, etc) that is published monthly was used ignoring the quarterly and annual data.
    - ***Daily macroeconomic data*** was calculated by ***resampling*** from monthly values using ***forward fill method***
    - This resampling method is sensible/relevant as the macroeconomic data considered is released monthly and is relevant for that period till the next value is released in the next month
- ***Note*** that each countries' macroeconomic data has to be considered separately and combined with US data as it will impact the currency pairs.
- For each currency pair daily End of Day price for each business day in the month (that had published prices) was used
    - ***Simple Moving Average (SMA)*** for currency prices was also considered as that can give better results for some currency pairs
    - However, the results were not better than EOD prices so was not used for analysis
    - Sample results are given for random forest algos, but for reference only
- Historic macroeconomic data and currency prices were taken from our website https://datawisdomx.com, which sources data from reliable well-known data providers.

Algorithms

- Standard scikit-learn, keras and TensorFlow libraries were used for running the different algorithms in python.
- ***Regression algorithms used*** - Random Forest Regression, Support Vector Regression and Artificial Neural Networks

- The data was split into test and training sets, with a test_size = 0.25, as that gave better results compared to 0.2 or 0.33 or other variations.
- ***Metrics*** used for evaluating the algorithms were
    - MSE – Mean Squared Error
    - MAE – Mean Absolute Error
    - MAPE – Mean Absolute Percentage Error
    - R2 – R-squared
- ***Hyper parameter tuning*** – the below hyper parameters gave the best results for the different algorithms
    - ***Random Forest Regressor*** - (n_estimators=100, criterion='mse', min_samples_leaf=5, max_depth=10, min_samples_split=10, max_features=8, n_jobs=-1)
    - ***Support Vector Regressor*** - (kernel = 'rbf', gamma='auto')
    - ***Artificial Neural Network Regressor*** –
        - Input layer and the first hidden layer - Dense(units = 32, activation = 'relu', kernel_initializer = 'normal', input_dim = 76)
        - Second hidden layer - Dense(units = 16, kernel_initializer = 'normal', activation = 'relu'))
        - Output layer - Dense(units = 1, kernel_initializer = 'normal'))
        - Compiler - (optimizer = 'sgd', loss = 'mean_squared_error', metrics = ['mse', 'mae', 'mape'])
        - Model parameters - batch_size = 10, epochs = 100
- Results are close between the 3 algorithms. However**, Random forest gave the best results compared to SVR and ANN**
    - ***Lower MSE, MAE and higher R-squared values indicate higher accuracy and closer prediction***
    - ***Comparing the training-test model predicted values with actual test values within a +/- 5% difference band were between 60% - 70% depending on the currency pair***
    - For example, sample Random Forest metrics are given below
    -

RF Algo	MSE	MAE	R2	5% Err
Daily EURUSD	0.00929138865	0.0656349430	0.6751851262	**63.34%**
Daily GBPUSD	0.0128631108	0.0759042113	0.6747875696	**69.25%**
Daily USDINR	20.9744805334	2.9440463647	0.7326326496	**67.01%**

    - This is a good starting point as the model can now be improved upon by adding other data types (central bank monetary policy statements, political statements, etc) and using different time lags
    - ***Note*** – USDINR error metrics are quite bad as the macro data set is smaller, has more missing values in the original data set and currency prices used are cash prices from global exchanges. USDINR reacts better to futures prices from Indian exchanges like NSE. This will be published as a separate set of results
- The results are given in the spreadsheet - ***AlgoResults_24Apr19.xlsx***
    - It contains the data set for each Macro/Ccy pair, difference between predicted and actual test values, error and accuracy metrics for each algorithm

- For Random Forest it contains the feature importance list to assess which are the most relevant independent variables to consider
  - Most of the main features (with higher *feature_ importance_* values) being picked look correct/relevant
  - For example, EURUSD important features are given below
  - _x = US, _y = EU data

EURUSD	feat	val	
unemployed_persons_x	37	0.0715485	
households_debt_to_gdp_y	57	0.0657752	
gdp_y	51	0.0650408	
government_bond_10Y_x	15	0.0605869	
unemployment_rate_x	38	0.0543141	
youth_unemployment_rate_x	40	0.0513423	
core_inflation_rate_y	47	0.0446172	
government_budget_x	16	0.0443791	
consumer_confidence_x	4	0.0291154	
disposasble_personal_income_x	10	0.0279436	
households_debt_to_gdp_x	19	0.0240215	
interest_rate_y	61	0.0227004	
wages_x	39	0.0222760	
private_debt_to_gdp	32	0.0182911	
unemployed_persons_y	71	0.0168997	
interest_rate_x	23	0.0163281	
business_confidence_y	42	0.0161799	**0.651360**

Sample code, Data and Results

Data used for this analysis along with sample code and results are given in the below Github location - https://github.com/datawisdomx/CurrencyPricesMacroDataRegressionAlgos

- **Macro data** - usmacrodata.csv, eurmacrodata.csv, gbpmacrodata.csv, indmacrodata.csv
- **Ccy data** - eurusd_Jan00Feb19.csv, gbpusd_Jan00Feb19.csv, usdinr_Jan00Feb19.csv
- **Sample code** –
  - MacroCcyPrediction_RF.py
  - MacroCcyPrediction_SVR.py
  - MacroCcyPrediction_ANN.py
- **Results comparison** – AlgoResults_24Apr19.xlsx

Make sure you point the file loader to the correct location of the data file on your local drive.

Some possible Data and Algorithm Logic Variations

- Try changing the number of independent variables (X) considered for random forest using the values from *feature_ importance_*. It gives much better results in some cases, despite not being necessary for random forest.

- Try varying the combination of macroeconomic data for each currency pair. For example, UK, US data impacts GBPUSD, so you can try them together or separately to see their individual impact.
- Try varying the time series considered between the macroeconomic data (independent variable X) and the average currency price (dependent variable Y). This can be done by creating a time lag between the two variables.
    - So effectively, we try and use current macroeconomic data to study their impact on future currency prices, with time lags of 6 months, 12 months, etc.
    - This hypothesis is based on the premise that markets are forward looking and start adjusting their view on future expectations using current data trend.
    - ***This time lag hypothesis has not been considered for this analysis*** was already tested in the earlier paper that was published - *MacroCcyRegressionRandomForestResultsSummary_3Apr18.docx.*
    - Check the Github location for the earlier analysis data and results - https://github.com/mobicloudtrees/Macroeconomic-Data-and-Currency-Regression

Note - This daily currency price prediction using resampled daily macroeconomic data is a very simple premise and by itself not sufficient for all possible variations to the relationship between the macroeconomic data, currency prices and timeframe. There are many other variations that can be tried with the variables, data, time lag, different algorithms used and their parameters. Users can test that on their own and use as they see fit.

Note – The predicted and test results are not an exact match, which is quite difficult for such scenarios and data sets. But the variance reduces further with a larger data size, different time lags and other variables not considered here (central bank monetary policy statements, political statements, etc).

Please use them keeping in mind the disclaimer below.
https://datawisdomx.com/index.php/disclaimer/
Please get in touch if you see any errors or want to discuss this further at
nitin@datawisdomx.com