

RoboAdvisor Algo – Part 1

Predict 3 year Forward Asset Prices & Direction and Portfolio Return & Risk using Historic Macro data & Central Bank Forecasts

Legal Disclaimer – This research is not an investment advisory or a sales pitch.

Objective

- **Overall - build a RoboAdvisor algo** that can automatically construct and re-balance a multi-asset portfolio using forward predicted asset prices for different timeframes
- **Create a medium to long-term portfolio for small/medium investors** (low investment amount > \$500, covers majority of the population)
- **Part 1 - Predict 3 year forward monthly average asset prices and direction** (stock indices, commodities, bonds, currencies) using historic macroeconomic and central bank forecasts for US, UK and EU. Only the 5 main macro factors were used
- **Part 1 - Predict forward Return and Risk for a multi-asset portfolio** (stock indices, commodities, currencies, real estate, bonds) using **pre-defined asset weights** and predicted forward monthly asset prices for different time frames – 3 months to 3 years
- **Part 1 - Backtest** the results by comparing predicted historic asset prices against actual for historic actual macro data
- *Part 2 – Find optimal portfolio weights for maximum return and minimum risk*
- *Part 2 – Automate portfolio construction and re-balancing*
- *Part 2 – Use India, China. Not considered due to lack of historic govt bond price data*
- *Part 2 – Build a model to forecast macro data using all other macro data, to use for forward price prediction*

Data

- US, UK and EU monthly main macroeconomic data for last 21 years (Jan 99 – Dec 19) was used for independent variables (X).
- Average monthly close price over the same period was used for asset prices (main stock indices, commodities, bonds, currencies) as the dependent variable (Y).
- All 3 countries' macroeconomic data was combined to see their impact on each asset price. **Only the 5 main macro factors (forecasted by central banks) were used for 3 year predictions – interest rate, core & headline inflation, GDP, unemployment**
- Historic Asset price, Macroeconomic data has been sourced from public data sets and our website <https://datawisdomx.com>, which sources data from reliable well-known data providers.

Algorithm Logic - Forward Asset Price Prediction

- For 3 year forward price/direction forecasts, central bank forecasts were added to historic actual macro data in monthly forward loops
- Model is trained and tested using Historic + Forecast Macro data to predict forward asset price
- Predicted asset price is then added to historic data for re-training model and predicting for next forward timeframe

- For forward asset price **prediction for month 't'**
 - Model is trained/tested on asset price and macro data till 't-2'
 - Macro data for 't-1' is then used to predict asset price for 't'. This ensures it is not used for training/testing the model (prevent leakage)
- **Forward Price Asset model: $FPA_t = RFModel_{t-2} (Macro_{t-1})$**
 - $t = 1$ to 36 (3 years, 36 months). $i = 1$ to 251 (Jan '99 – Nov '19)
 - For $[t=1, i = i + (t-1)]$
 - **$RFModel_{t-2} = \text{Random Forest } (A_i, M_i)$**
 - **$Macro_{t-1} = Macro_{i+1}$** (Dec '19)
 - For $[t > 1, i = i + (t-1)]$
 - **$RFModel_{t-2} = \text{Random Forest } (A_i, M_i)$** . (Jan '99 – Dec '19)
 - $A_i = A_i + A_{t-1}$. A_i = Actual, A_{t-1} = Predicted using model
 - $M_i = M_i + M_{t-1}$. M_i = Actual, M_{t-1} = Central bank forecast
 - **$Macro_{t-1} = Macro_{i+1}$** (Jan '20). Central bank forecast
 - **This loop runs till entire 3 year forward prices are predicted**
- Eg: Asset price for Jan '20 was predicted using Macro data for Dec '19 and model was trained using asset price and macro data upto Nov '19
- **Limitation – Only the 5 main macro factors** (forecasted by central banks) were used for 3 year predictions – interest rate, core & headline inflation, GDP, unemployment
 - Forward 3 year forecasts for other macro data like industrial/manufacturing production, consumer/business confidence, retail sales is not available/reliable

Note - Tried to build a model to use all macro data to predict another macro data. **Results were not good / degraded** as time frame was increased beyond a few months

Algorithm setup details

- Standard python scikit-learn, pandas, numpy, visualization libraries were used for running the different algorithms in python
- **Regression algorithms used** - Random Forest. XGBoost was considered but not used
- 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.
- **Feature scaling** - Data wasn't standardized as it would require the new independent data vector for new price prediction to be standardized using the trained models mean/variance
 - Also, with Random Forest/XGBoost algorithms it's not necessary as they are based on decision tree ensemble model (Bagging/Boosting), which does not require standardization (not distance based)
- **Metrics** used for evaluating the algorithms were
 - MSE – Mean Squared Error, RMSE – Root Mean Squared Error
 - MAE – Mean Absolute Error, R2 – R-squared
- **Hyper parameter tuning** – was done using GridSearchCV
 - It ensures that multiple variations of the data sample are used by shuffling it randomly for k-fold splits, thereby preventing overfitting on the same test data and reducing bias
- **Below hyper parameters** gave the best results

- **RandomForest** - n_estimators=1000, criterion='mse', min_samples_leaf=2, max_depth=17, min_samples_split=2, max_features='sqrt', random_state=42, n_jobs=-1
- **Python code** for data wrangling, GridsearchCV hyperparameter tuning for each model, model explanation details for Random Forest and XGBoost and data visualization is given in github – **RoboAdvisorAlgoPart1_7Jan20.py**

Portfolio Construction, Return and Risk Logic

- **Portfolio is constructed** using a typical composition of different asset classes – Stock Index, Bonds, Gold, Real Estate, Currencies, Cash, Commodity Index
- **Create a medium to long-term portfolio for small/medium investors** (low investment amount > \$500, covers majority of the population)
- **Investment strategy** will determine asset class and weight allocation. Example:

Low risk	Medium risk	High risk
Conservative	Moderate	Aggressive

- **Highly liquid, Daily priced, Exchange traded Indices/Assets futures were considered**
 - Kind of assets most small investors understand/prefer
 - Easy and cheap to buy/sell, price, benchmark, backtest
 - Very low risk of price manipulation and default, unlike illiquid/alternative assets
- **Portfolio Return** = Weighted sum of asset returns, over a given period (months to years)
 - **Portfolio Return** = $\sum_{i=1 \text{ to } n} AR_i * W_i$
 - AR = Asset Return for period (Yearly/Monthly)
 - W = % Weight of Asset in Portfolio for period
 - Portfolio Weight grid example:

Period	StkIndex	REIndex	Gold	OilWTI	Bond10Yr	Cash	Total
3yr	0.4	0.05	0.2	0.1	0.2	0.05	1

- **Asset Annual Return rate** is calculated using Start/End of Year Asset Price
 - Asset Annual RR = $(APEoY - APSoy)/APSoY$
- **Asset Monthly Return rate** using Start/End of Month Asset Price
 - Asset Monthly RR = $(APEoM - APSoM)/APSoM$
- **Asset Annualized return** using Monthly Return rate of Asset
 - Asset Annualized Returns AAR = $[(1+amr_1)*(1+amr_2)...*(1+amr_n)]^{1/n} - 1$
 - amr = asset monthly return
- **Bond (10 year government - US, UK, EUBund) returns were calculated using futures price and average yield, instead of normal bond calculation (Coupon, YTM, Period, Par)**
 - As bond will most likely not be held to majority, futures price gain/loss is used for calculating return. Coupon is close to average yield of bond futures
 - Bond Annual RR = $[BondPrice(EoY-SoY) + MeanBondYield(EoY \text{ To } SoY)] / BondPrice(SoY)$
 - Bond Monthly RR = $[BondPrice(EoM-SoM) + MeanBondYield(EoM \text{ To } SoM)] / BondPrice(SoM)$
 - For UK, EUBund bond futures price data is available only from Mar '08

- **Portfolio Risk - Variance (Volatility)** is calculated using the standard Modern Portfolio Theory formula which considers both risk and correlation of the assets in the portfolio
 - **Variance(Portfolio) = Volatility**
 - $V = \sum_{i=1 \text{ to } n} (\text{Variance}(\text{AR}_i) * W_i^2) + \sum_{i=1 \text{ to } n} \sum_{j=i+1 \text{ to } n} [2 * W_i * W_j * \text{Covariance}(\text{AR}_i, \text{AR}_j)]$
 - **Covariance(AR_i, AR_j)** = CorrelationCoefficient(AR_i, AR_j) * SDAR_i * SDAR_j
 - **SDAR** = Standard Deviation of Asset Return
- **Portfolio Risk - Standard Deviation** measures deviation of the returns from the mean
 - **Standard Deviation (Portfolio)** = $[\text{Variance}(\text{Portfolio})]^{1/2}$
- **Explanation of risk terms**
 - Variance (x) = $\sum_{i=1 \text{ to } n} ((x_i - \text{Average } x)^2) / n$. Spread from mean. Excess weight to outliers
 - Covariance (x,y) = $\sum_{i=1 \text{ to } n} [(x_i - \text{Avg } x) * (y_i - \text{Avg } y)] / n - 1$. Directional relationship between returns of 2 assets (x,y)
 - Correlation Coefficient - Strength of relationship between 2 assets -1 to 1
 - Standard Deviation of Portfolio. Dispersion from mean. Same unit as data
- **Portfolio Weight grid different timeframes sample.**
 - *Portfolio weights have to be optimized for different time frames and asset mix – will be done in part 2*

Period	StkIndex	REIndex	Gold	OilWTI	Bond10Yr	Cash	Total
3m	0.3	0.1	0.2	0	0.3	0.1	1
6m	0.3	0.1	0.2	0	0.3	0.1	1
9m	0.25	0.1	0.2	0.05	3	0.1	1
1yr	0.2	0.1	0.2	0.05	0.35	0.1	1
2yr	0.2	0.2	0.1	0.05	0.35	0.1	1
3yr	0.4	0.05	0.2	0.1	0.2	0.05	1
5yr	0.25	0.25	0.1	0.05	0.3	0.05	1
7yr	0.3	0.25	0.05	0.05	0.3	0.05	1
10yr	0.25	0.3	0.05	0.05	0.3	0.05	1

Backtesting Logic

- Backtesting is used to measure the accuracy of the model (which uses central bank forecast data) by using the same logic for predicting asset prices using actual macro and asset price data
 - Predict asset's monthly average price using actual Macro data for US, UK and EU
 - Loop in reverse, start at last year and month of actual data, drop 1 month each time, till last 1 year (1999) of Macro data is left
 - Use macro, asset data till t-2 to train model. Then t-1 macro data to predict asset price for t
 - We need at least 1 year's macro, asset data to predict future asset price
- **Accuracy** - This allows us to measure how accurately the model is predicting asset prices
 - (Predicted monthly asset price - Actual monthly price) for historic data
- **Direction** - Calculate Predicted vs Actual direction using actual prices and macro data

- Since **predicted prices are consistently below actual price** due to random forest algorithm's limitation, we can use asset price monthly direction as another measure to backtest
- $t = 1$ to [(last year and month)-1] of actual data
- $\text{Actual direction}_t = \text{Actual Price}_t - \text{Actual Price}_{t-1}$
- $\text{Predicted direction}_t = \text{Predicted Price}_t - \text{Predicted Price}_{t-1}$
- If $\text{Actual direction}_t = \text{Predicted direction}_t$, then **Direction = 1. Same direction**
- If $\text{Actual direction}_t \neq \text{Predicted direction}_t$, then **Direction = 0. Different direction**
- We can now compare the Direction historically to see if we can get the trend right, so as to **determine Long/Short position for the month**
- We can count the total correct directions for each year and for each month across years to see **accuracy of direction**

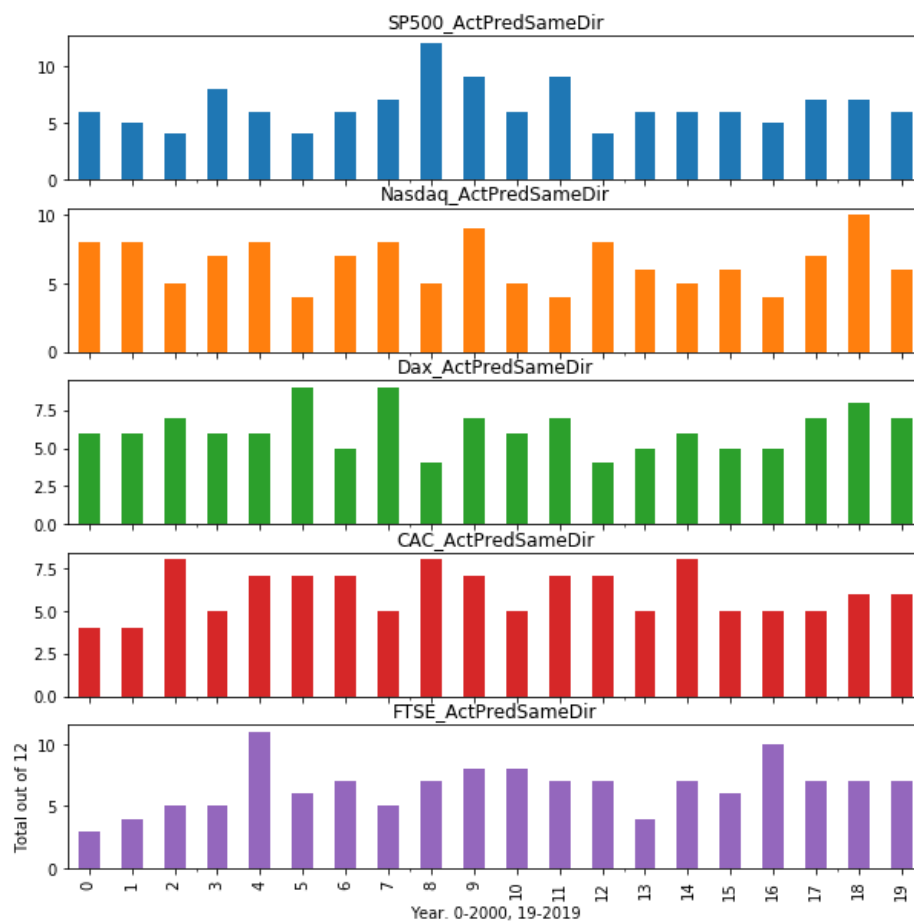
Results, Backtesting Analysis

- Results are given for all assets in the spreadsheet - *RoboAdvisorAlgoPart1_7Jan20.xlsx*
- **Actual Price/Direction Prediction - Results are good but not accurate enough**
 - **Predicted price** consistently below actual price, irrespective of asset type and its date in the series
 - **Direction prediction is quite good, with atleast 50% accuracy for most assets**
- **Forward Price/Direction Prediction - Results are good but not accurate enough**
 - **Predicted price** consistently below actual price, irrespective of asset type and its date in the series
 - **Direction prediction is quite good, with atleast 75% accuracy for most assets for now (Jan '20)**
- **Portfolio Return and Risk – Logic has been built but not backtested**
 - Tested for a few periods and weight combinations. Results are varied/not good
 - **Note - Will be done as part of part 2 optimization of weights**
- **AllAssetAnnualNRR, AllAssetMthlyNRR** sheets – contain the actual annual and monthly returns for 1999-2019 for all assets
- **AllAssetsActPredPrices** sheet – Backtested actual data. 75%-85% of Predicted monthly values are within +/- 20% difference from actual
- **AllAssetsActPredDirection** sheet – Backtested for Actual and Predicted prices, +/- Diff values indicate monthly asset price move higher/lower in current month compared to previous
 - Dir values indicate if the direction (same/different, 1/0) is same between actual and predicted prices
 - Price and Direction for each year are shown graphically below, split into asset class group – equities, commodities, currencies, bonds
- **AllAssetSameDirYrlyTot** sheet - Backtested for Actual and Predicted prices, directional match total for each year
 - **Total > 5 indicates atleast 50% of the time (6 months out of 12) in the year the direction was predicted accurately**
 - Results are varied, but for **Equities and Commodities/Currencies Total > 5 for most years**
 - Bond results are slightly worse, possibly due to lack of entire 21 year price data

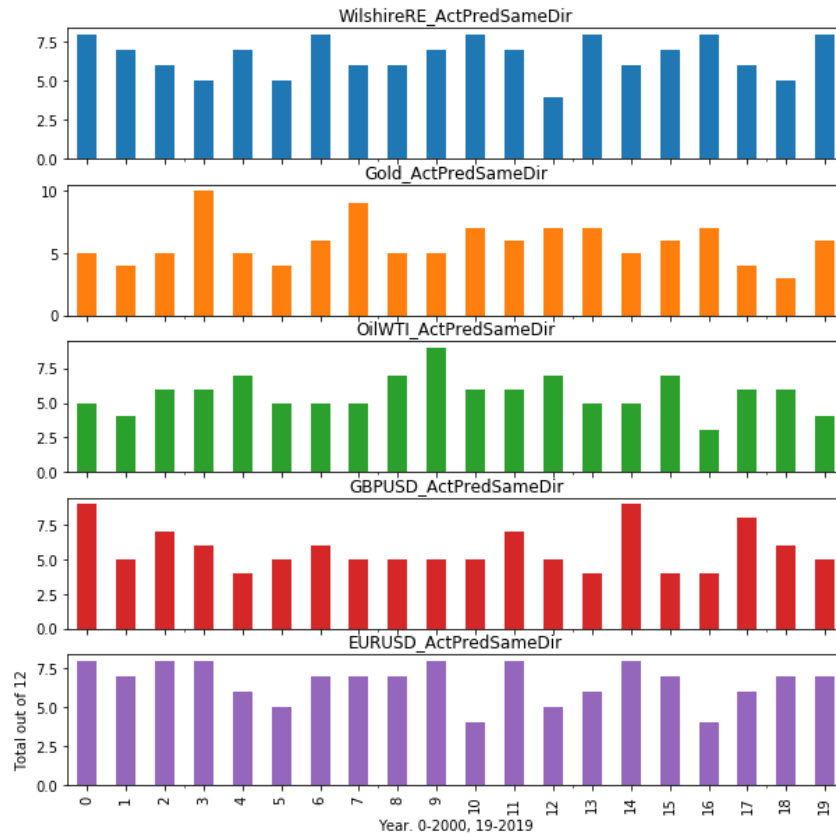
- Total for each year are shown in a grid and graphically below, split into asset class group – equities, commodities, currencies, bonds

Year	SP500	Nasdaq	Dax	CAC	FTSE	WilshireRE	Gold	OilWTI	GBPUSD	EURUSD	UST10YrYield	UST10YrPrice	UKGilt10YrYield	UKGilt10YrPrice	EUBund10YrYield	EUBund10YrPrice
2000	6.0	8	6	4	3	8	5	5	9	8	6	4	3	11	2	11
2001	5	8	6	4	4	7	4	4	5	7	3	6	5	12	6	12
2002	4	5	7	8	5	6	5	6	7	8	6	7	5	12	4	12
2003	8	7	6	5	5	5	10	6	6	8	4	5	6	12	4	12
2004	6	8	6	7	11	7	5	7	4	6	5	5	7	12	6	12
2005	4	4	9	7	6	5	4	5	5	5	7	5	6	12	6	12
2006	6	7	5	7	7	8	6	5	6	7	6	6	7	12	6	12
2007	7	8	9	5	5	6	9	5	5	7	6	6	5	12	7	12
2008	12	5	4	8	7	6	5	7	5	7	4	4	4	6	4	12
2009	9	9	7	7	8	7	5	9	5	8	5	7	5	7	1	6
2010	6	5	6	5	8	8	7	6	5	4	8	8	6	4	6	7
2011	9	4	7	7	7	7	6	6	7	8	4	7	7	7	4	5
2012	4	8	4	7	7	4	7	7	5	5	4	8	5	4	6	5
2013	6	6	5	5	4	8	7	5	4	6	8	6	7	6	6	7
2014	6	5	6	8	7	6	5	5	9	8	4	3	4	8	4	7
2015	6	6	5	5	6	7	6	7	4	7	6	5	4	7	6	6
2016	5	4	5	5	10	8	7	3	4	4	6	3	4	5	7	4
2017	7	7	7	5	7	6	4	6	8	6	5	6	7	8	6	7
2018	7	10	8	6	7	5	3	6	6	7	6	7	2	6	6	4
2019	6	6	7	6	7	8	6	4	5	7	7	8	8	6	7	3

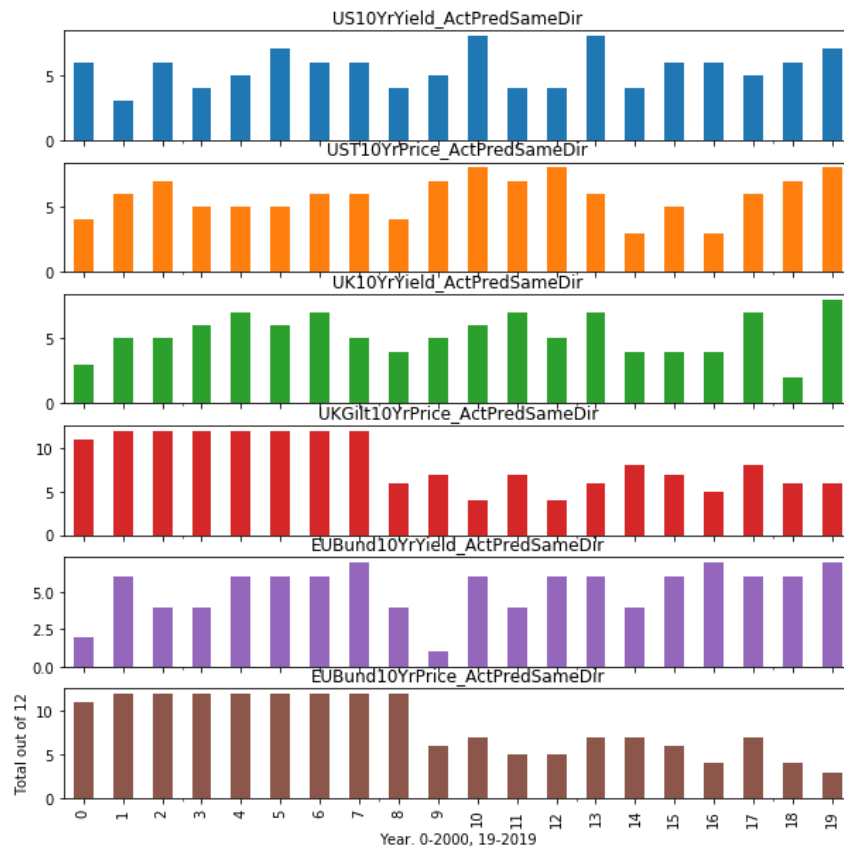
Act vs Pred Same Direction Year Total



Act vs Pred Same Direction Year Total



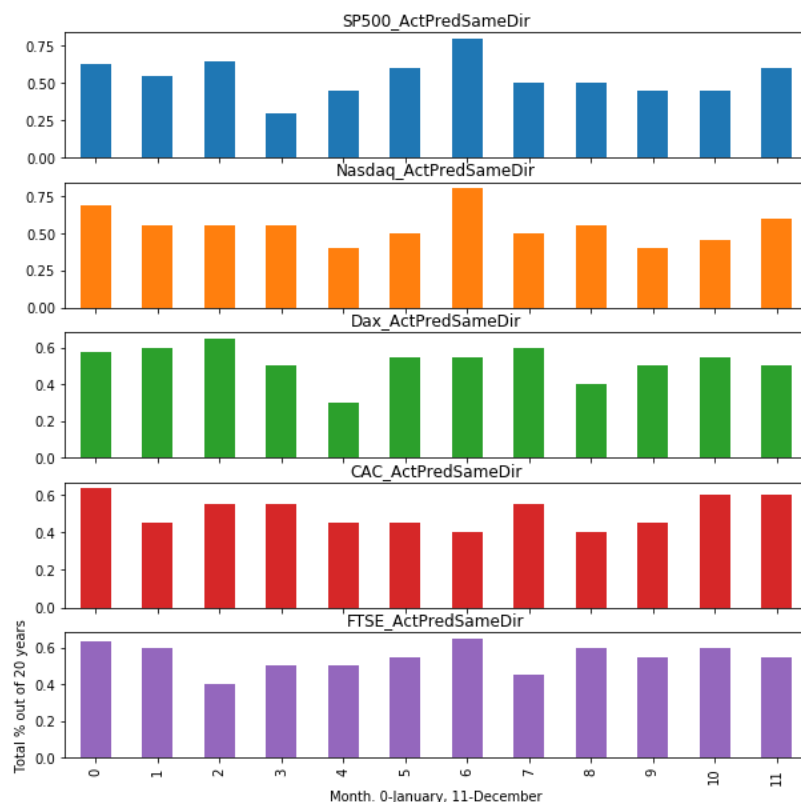
Act vs Pred Same Direction Year Total



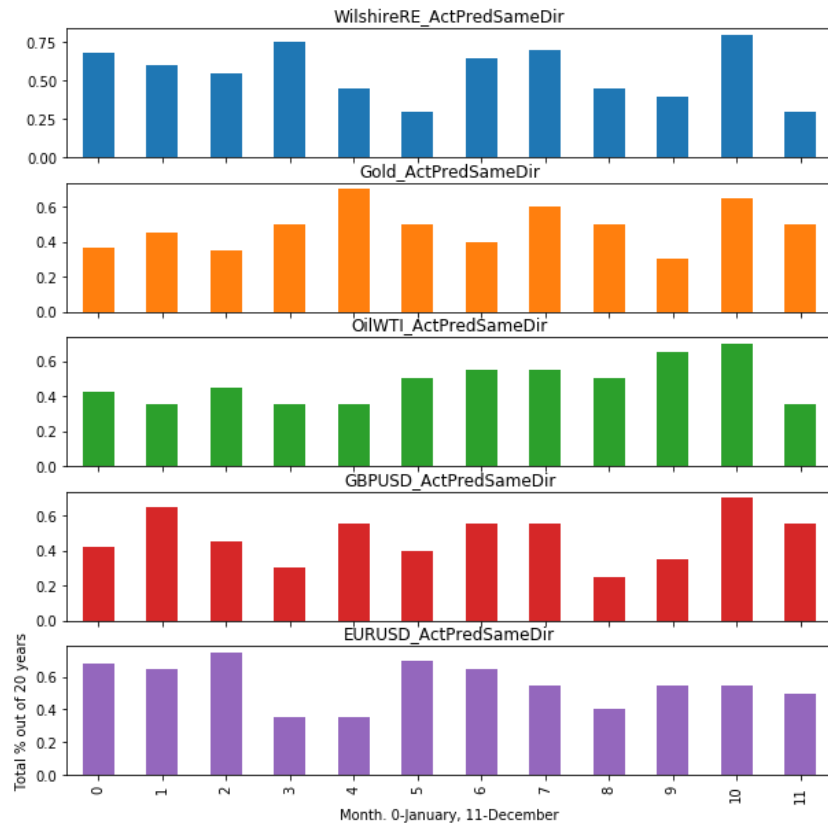
- **AllAssetSameDirMthlyTotPct** – Backtested for Actual and Predicted prices, % total of directional match for each month across all years
 - This is used to measure if there are specific months where predicted prices are more accurate
 - **Eg:** Total % same direction matches across January from 2000 – 2019 (20 years)
 - **Total % > 0.5 indicates atleast 50% of the time (10 years out of 20) that month had same predicted direction as actual**
 - Results are varied, but only for **Equities Total % > 0.5 for most years**
 - Rest asset types results are slightly worse, possibly due to monthly/seasonal factors
 - Total % for each year are shown in a grid and graphically below, split into asset class group – equities, commodities, currencies, bonds

Month	SP500	Nasdaq	Dax	CAC	FTSE	WilshireRE	Gold	OilWTI	GBPUSD	EURUSD	UST10YrYield	UST10YrPrice	UKGilt10YrYield	UKGilt10YrPrice	EUBund10YrYield	EUBund10YrPrice
Jan	0.63	0.68	0.58	0.63	0.63	0.68	0.37	0.42	0.42	0.68	0.32	0.32	0.37	0.53	0.42	0.63
Feb	0.55	0.55	0.60	0.45	0.60	0.60	0.45	0.35	0.65	0.65	0.50	0.60	0.45	0.70	0.50	0.80
Mar	0.65	0.55	0.65	0.55	0.40	0.55	0.35	0.45	0.45	0.75	0.35	0.55	0.45	0.70	0.70	0.75
Apr	0.30	0.55	0.50	0.55	0.50	0.75	0.50	0.35	0.30	0.35	0.65	0.60	0.60	0.70	0.40	0.60
May	0.45	0.40	0.30	0.45	0.50	0.45	0.70	0.35	0.55	0.35	0.40	0.55	0.40	0.80	0.30	0.75
Jun	0.60	0.50	0.55	0.45	0.55	0.30	0.50	0.50	0.40	0.70	0.35	0.30	0.50	0.70	0.60	0.70
Jul	0.80	0.80	0.55	0.40	0.65	0.65	0.40	0.55	0.55	0.65	0.30	0.55	0.40	0.85	0.50	0.90
Aug	0.50	0.50	0.60	0.55	0.45	0.70	0.60	0.55	0.55	0.55	0.70	0.60	0.30	0.60	0.20	0.65
Sep	0.50	0.55	0.40	0.40	0.60	0.45	0.50	0.50	0.25	0.40	0.45	0.35	0.60	0.80	0.35	0.65
Oct	0.45	0.40	0.50	0.45	0.55	0.40	0.30	0.65	0.35	0.55	0.25	0.45	0.35	0.60	0.35	0.70
Nov	0.45	0.45	0.55	0.60	0.60	0.80	0.65	0.70	0.70	0.55	0.45	0.35	0.55	0.70	0.60	0.75
Dec	0.60	0.60	0.50	0.60	0.55	0.30	0.50	0.35	0.55	0.50	0.80	0.60	0.40	0.80	0.30	0.55

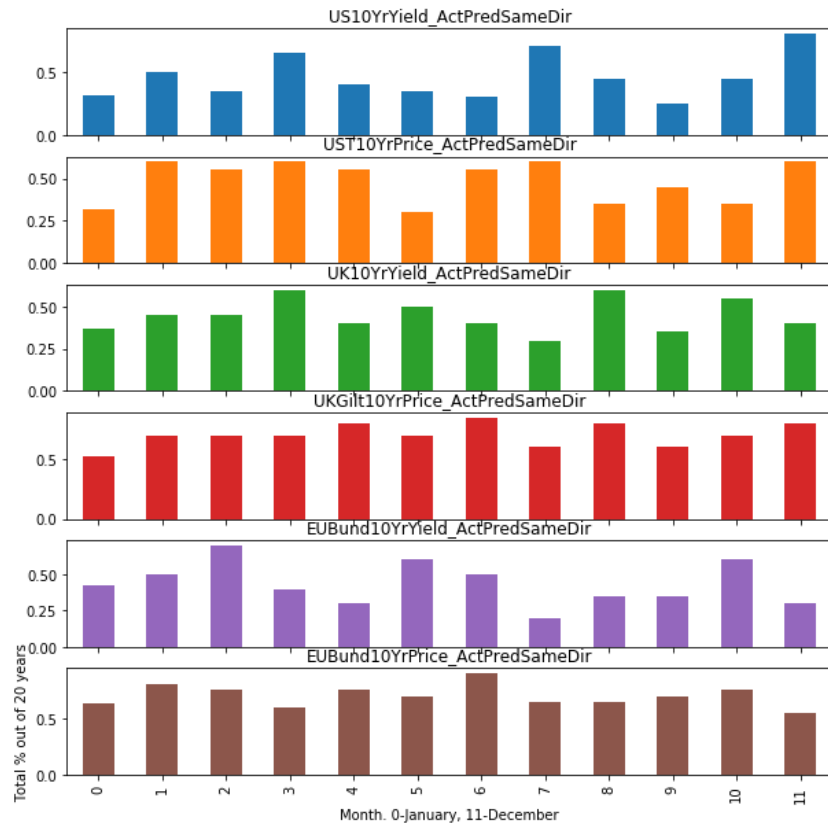
Act vs Pred Same Month All Years Total %



Act vs Pred Same Month All Years Total %



Act vs Pred Same Month All Years Total %



- **AllAssetsForecast3YrPrices** sheet – contains for all assets the predicted monthly forward prices for 3 years using all historic macro + central bank macro forecasts
 - *Predicted forward prices cannot be compared to actual until more real market data becomes available in 2020 and beyond*
- **AllAssetsForecast3YrDirection** sheet - contains for all assets the predicted monthly direction of forward prices for 3 years
 - **Overall results are quite good, with atleast 75% accuracy in predicted direction for most assets**
 - **Eg: SP500, Gold** - Nov'19, Dec'19 prices were predicted rise / go higher and that was the actual market move. Jan '20 is predicted lower/fall, and is valid for now
 - **Eg: OilWTI, GBPUSD** – Correct for Nov'19 wrong for Dec'19. Jan '20 correct for now
 - But results are varied/not correct for some equities, commodities and bonds
 - These values will improve as actual macro data becomes available
 - Results for a few future months are given in a grid below Results will be published daily in live on a website to compare with actual market regularly
- **Note** - The current algorithm doesn't incorporate geopolitical/trade conflicts, elections, comments by politicians, etc. These have an impact and will be incorporated in the algo later in part 2

Date	SP500	PredDiffDir	Nasdaq	PredDiffDir	Dax_Frc	PredDiffDir	CAC_Frc	PredDiffDir	FTSE_Frc	PredDiffDir	WilshireRE	PredDiffDir	Gold	PredDiffDir	OilWTI	PredDiffDir
01/11/2019	1.06	1.00	-67.43	-1.00	64.81	1.00	-73.15	-1.00	-3.34	-1.00	-2.76	-1.00	-8.71	-1.00	0.80	1.00
01/12/2019	143.83	1.00	575.74	1.00	836.92	1.00	-87.10	-1.00	272.01	1.00	14.19	1.00	36.96	1.00	-3.46	-1.00
01/01/2020	-1.13	-1.00	18.78	1.00	143.93	1.00	48.75	1.00	-109.52	-1.00	0.95	1.00	-4.88	-1.00	0.09	1.00
01/02/2020	2.99	1.00	106.76	1.00	-281.38	-1.00	71.49	1.00	47.74	1.00	-2.45	-1.00	15.50	1.00	-0.33	-1.00
01/03/2020	-3.37	-1.00	-101.87	-1.00	85.68	1.00	-49.74	-1.00	-16.53	-1.00	2.33	1.00	-6.58	-1.00	0.22	1.00

Date	GBPUSD_Frc	PredDiffDir	EURUSD_Frc	PredDiffDir	UST10YrYield	PredDiffDir	UST10YrPrice	PredDiffDir	UKGilt10YrY	PredDiffDir	UKGilt10YrP	PredDiffDir	EUBund10Yr	PredDiffDir	EUBund10Yr	PredDiffDir
01/11/2019	0.00	-1.00	0.00	1.00	0.01	1.00	1.78	1.00	0.01	1.00	-1.45	-1.00	-0.04	-1.00	-0.28	-1.00
01/12/2019	-0.06	-1.00	-0.05	-1.00	-0.15	-1.00	2.47	1.00	-0.19	-1.00	2.74	1.00	-0.25	-1.00	7.23	1.00
01/01/2020	0.00	1.00	0.01	1.00	0.06	1.00	0.09	1.00	0.09	1.00	0.19	1.00	0.05	1.00	-0.44	-1.00
01/02/2020	0.01	1.00	0.00	1.00	-0.05	-1.00	-0.10	-1.00	-0.06	-1.00	-0.16	-1.00	-0.04	-1.00	1.03	1.00
01/03/2020	0.00	-1.00	0.00	-1.00	0.02	1.00	-0.13	-1.00	0.00	1.00	-0.35	-1.00	-0.01	-1.00	-0.29	-1.00

Conclusion

- While the predicted direction results are quite good but price prediction not as accurate as desired, they still show validity of the algorithm and logic for all assets
- As all possible data has not been considered, this was expected
- As other macro/market data and more samples are included, expectation is that it will improve the results
- It provides a good base to work with to improve the results
- Results further validate prior research on price prediction and correlation between asset prices and macroeconomic data

Further research, ways to improve algorithms

- This is a good starting point as the **model can now be improved by**
- Further analysis to bring predicted price closer to actual price
- Adding other data types (central bank statements, political statements, etc) and other countries' data (China, India, Japan, etc) can improve the results, as they are relevant
- Use dimensionality reduction like PCA to see if it improves the results, though it reduces explainability

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/RoboAdvisorAlgoPart1>

Published research – Asset Price / Macro data prediction and relationship

- Results of this new research further validate earlier research articles already published
- <https://github.com/datawisdomx/Predict-Monthly-Asset-Price-and-Direction-using-Macroeconomic-Data> - Predicts monthly asset price for future next 1 month using historic actual macro and asset data
- <https://github.com/datawisdomx/Monthly-Asset-price-relationship-to-Macroeconomic-Data> - Analyses correlation and coefficient of determination relationship between them

Note - This research is based on a very simple premise and small data set. This by itself is not sufficient for all possible variations to the relationship between macroeconomic data, countries, asset prices, timeframes, algorithms used and other factors like political & central bank data, etc. Users can test that on their own and use as they see fit.

Disclaimer

Please use this research 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