VECTOR VICTORS PROJECT REPORT

Members:

1. Raghav Govind [220852]

2. Parnapalli Sarath Kumar [220754]

3. Priyanshu Gupta [220826]

4. Jiya Verma [220480]

5. Suresh Rana [231080096]

Modeling and Forecasting Exchange Rate

Project Overview: Modelling and Forecasting Exchange Rate

Objective:
This project aimed to develop robust models for predicting exchange rates (INR/USD) using three different methods and subsequently evaluating their performance metrics.

Approach:
1. Data Preprocessing:
   - Z-score normalization was applied to the data to standardize it.
   - Correlation matrices were utilized to visualize relationships and identify trends among the variables.

2. Multicollinearity Analysis:
   - Variance Inflation Factor (VIF) analysis was conducted to assess multicollinearity.
   - The analysis identified and addressed potential redundancy among predictors, enhancing the model's reliability.

3. Variable Selection:
   - Independent variables like GDP growth, Trade Deficit (% of GDP), FDI (% GDP), Inflation, Real Interest Rates, and Crude Oil Prices (dollar per barrel) were considered.
   - Variables with high correlation with the target variable and VIF scores less than 5 were retained, while Population Growth and Total Reserves were excluded.

4. Model Implementation:
   - Time series forecasting models, including ARIMA and ARMA, were implemented.
   - Multiple Linear Regression was employed for predicting exchange rates.
   - The performance of each model was evaluated using metrics such as MAE, RMSE, and MAPE.

Outcome:
- Linear Regression Performance:
  - Achieved an impressive R-squared value of 0.96, indicating a strong linear relationship between predictors and the exchange rate.
  - The adjusted R-squared value, considering model complexity, was around 0.7.

- Time Series Models (ARIMA, ARMA):
  - ARIMA outperformed other models, exhibiting the lowest RMSE, MAE, and MAPE.ARIMA performed better than the ARMA because of the extra differencing as the data was not stationary. That differencing step made the data stationary, which is required for the ARIMA model.
  - This suggests that the time series components captured by ARIMA contributed significantly to accurate predictions.

- ANN (Artificial Neural Network):
  - Attempted to use ANN for prediction, but no big difference compared to linear regression.

Conclusion:
- The project demonstrated the effectiveness of various models in predicting exchange rates.
- Linear regression showcased a strong linear relationship, while time series models, especially ARIMA, excelled in forecasting accuracy.
- The decision to exclude certain variables based on correlation and VIF analysis contributed to model refinement.

Recommendations:
- Further exploration of advanced time series models or deep learning approaches like Long Short-Term Memory (LSTM) networks could be considered for even more accurate predictions.
- Continuous monitoring and updating of the model may be necessary to adapt to changing economic conditions.
- L2 regularization can be used as we saw that some features are highly correlated to others so it reduces multicollinearity.

This project provided valuable insights into exchange rate prediction methodologies and highlighted the significance of careful variable selection and model evaluation in achieving accurate forecasts.

Followup questions:

VIF (Variance Inflation Factor) Calculation and Analysis:

1. VIF Calculation:

- The VIF assesses the extent of multicollinearity among predictor variables in a regression model.
- For each predictor variable, the VIF is calculated as the ratio of the variance of the model with that variable to the variance of a model without that variable.

Mathematically:

$VIF_i = \frac{1}{1-R_i^2}$

- Where R_i^2 is the R^2 value of the regression model with the i-th variable as the dependent variable and all other variables as predictors.

2. Interpreting VIF Values:

- A VIF of 1 indicates no multicollinearity (perfectly uncorrelated predictors).
- Generally, a VIF below 5 is considered acceptable, while values above 10 may indicate problematic levels of multicollinearity.
- Higher VIF values suggest a higher correlation between the predictor variable and the other variables in the model.

3. Analysis and Action:

- **VIF < 5:** Variables with low VIF values are considered acceptable and contribute independently to the model.
- **5 < VIF < 10:** Moderate levels of multicollinearity. Some caution is needed, and consideration can be given to potential model simplification.
- **VIF > 10:** High multicollinearity. Variables with high VIF may need to be addressed by either removing them from the model or combining them with other related variables.

4. Handling Multicollinearity:

- **Variable Removal:** Identify and remove highly correlated variables if they do not contribute substantially to the model.
- **Variable Transformation:** Combine or transform variables to reduce correlation.
- **Principal Component Analysis (PCA):** Use PCA to create uncorrelated linear combinations of variables.

ARIMA

The ARIMA (AutoRegressive Integrated Moving Average) model is a time series forecasting model that combines autoregression (AR), differencing (I), and moving averages (MA). Let's delve into the mathematics behind ARIMA:

Autoregressive (AR) Component:

The AR(p) component models the relationship between the current value and its past values. It can be expressed as:

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \varepsilon_t$$

Here,

- X_t is the current value.
- $\phi_1, \phi_2, \ldots, \phi_p$ are the autoregressive coefficients.
- c is a constant term.
- ε_t is the white noise term.

Integrated (I) Component:

The I(d) component represents differencing, which is used to make the time series stationary. It involves subtracting the previous observation from the current one. The differenced series can be denoted as:

$$Y_t = X_t - X_{t-d}$$

Here,

- Y_t is the differenced series.
- d is the order of differencing.

Moving Average (MA) Component:

The MA(q) component models the relationship between the current value and a residual error from a moving average model. It can be expressed as:

$$X_t = c + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

Here,

- $\theta_1, \theta_2, \ldots, \theta_q$ are the moving average coefficients.

ARIMA(p, d, q) Model:

The ARIMA model combines these components:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

Here,

- Y_t is the differenced series.
- p is the order of the autoregressive component.
- d is the order of differencing.
- q is the order of the moving average component.

Stationarity:

ARIMA models assume that the time series is stationary after differencing. A stationary time series has constant mean and variance over time.

Parameter Estimation:

Parameters (ϕ's, θ's) are estimated using methods like Maximum Likelihood Estimation (MLE).

Forecasting:

Once the model is fitted, it can be used to forecast future values based on past observations and forecasted residuals.

In practice, tools like ACF (AutoCorrelation Function) and PACF (Partial AutoCorrelation Function) plots are often used to determine suitable values for p, d, and q during model selection.

Feature	ARIMA (p, d, q)	ARMA (p, q)
Components	AR + I + MA	AR + MA
Purpose	Time series forecasting with trend and seasonality	Time series forecasting without trend/seasonality
Differencing (I)	Includes differencing to make the series stationary	May not include differencing (stationary data)
Stationarity	Assumes differenced series is stationary	Assumes series is stationary
Typical Use Cases	Data with trends and seasonality	Stationary data without trends or seasonality
Model Equation (Simplified)	$Y_t = c + \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t$	$Y_t = c + \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t$
Example Library Functions	`statsmodels.tsa.arima.model.ARIMA`	`statsmodels.tsa.arima.model.ARMA`

What is R2 value?

$$R^2 = \frac{\text{Var(mean)} - \text{Var(fit)}}{\text{Var(mean)}}$$

$$R^2 = \frac{\text{SS(mean)} - \text{SS(fit)}}{\text{SS(mean)}}$$

SS= sum of squares

The Adjusted R-squared is a modified version of the R-squared (coefficient of determination) that adjusts for the number of predictors in a regression model. While R-squared measures the proportion of the variance in the dependent variable explained by the independent variables, Adjusted R-squared takes into account the number of predictors and penalizes the inclusion of irrelevant variables that do not significantly contribute to explaining the variation.

The formula for Adjusted R-squared is:

$$\text{Adjusted R-squared} = 1 - \left(\frac{(1-R^2)\times(n-1)}{(n-k-1)}\right)$$

where:

- R^2 is the R-squared value.
- n is the number of observations.
- k is the number of predictors.