Time series project

USD/INR Exchange Rate Prediction

Rutgers University

December 7,2021

1. Introduction

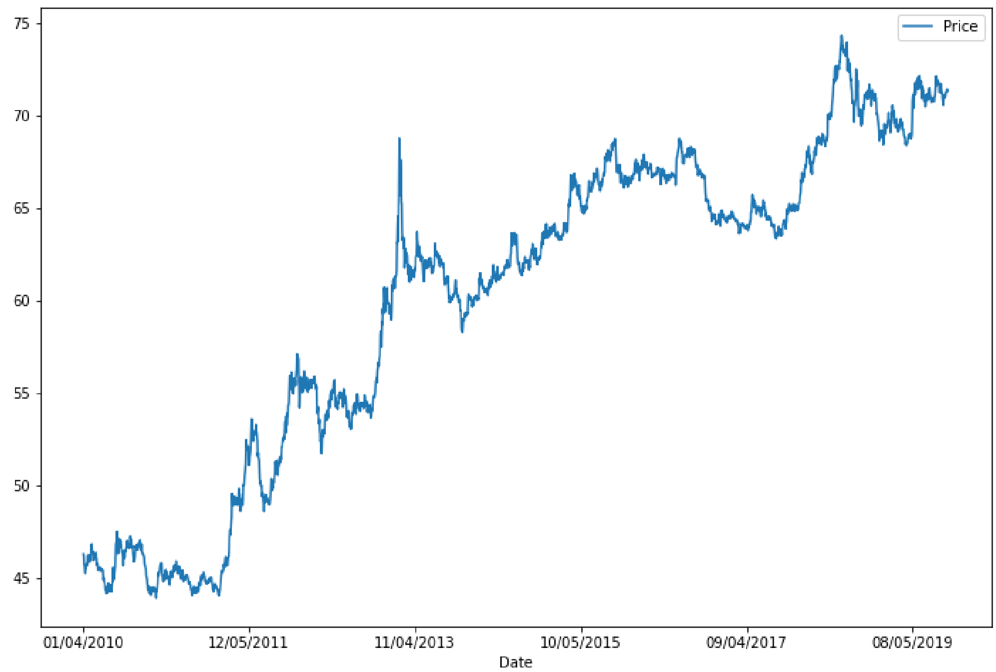
We analyze the time series data as the USD/INR Exchange rate from Jan 01, 2010, to Dec 31, 2019. There are a total of 2608 data corresponding to each day. The data is obtained from www.investing.com. The USD/INR pair tells the trader how many Indian Rupees (the quoted currency) are needed to purchase one U.S. dollar (the base currency). The Rupee is symbolized by ₹ and is the 20th most traded currency worldwide. This project expects to fit a forecasting model for the daily USD/INR exchange rate.

2. Data

The USD/INR exchange rate data has six variables and 2608 observations. The data does not have any missing values, and there is a strong correlation between 'Price' and all the other variables making it a univariate analysis.

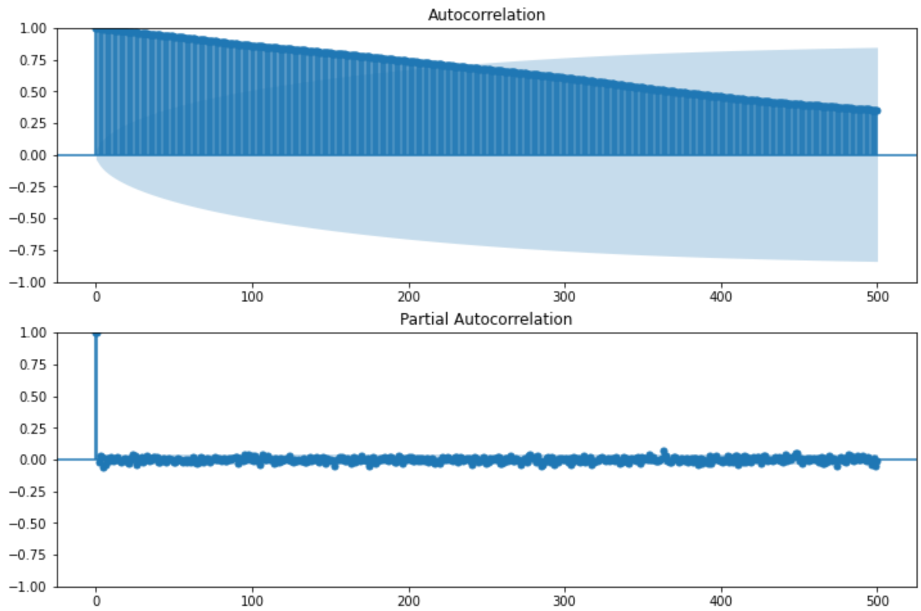
	Date	Price	Open	High	Low	Change %
0	Dec 31, 2019	71.35	71.295	71.385	71.225	0.06%
1	Dec 30, 2019	71.31	71.340	71.427	71.290	-0.18%
2	Dec 27, 2019	71.44	71.315	71.505	71.175	0.21%
3	Dec 26, 2019	71.29	71.270	71.348	71.225	0.01%
4	Dec 25, 2019	71.28	71.280	71.280	71.280	0.01%

In this section, we inspect the overall trend and seasonality of the data.

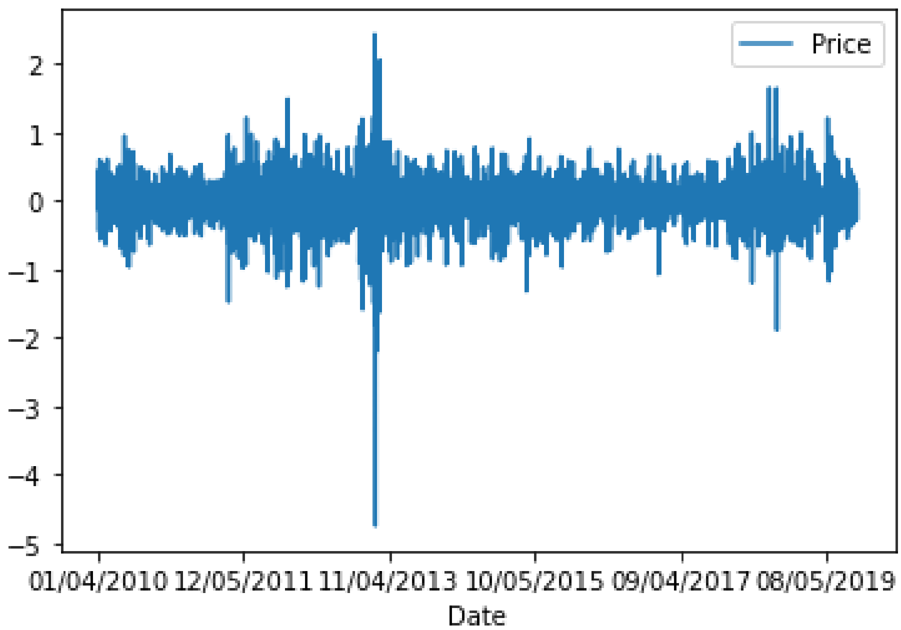


From the above plot, we can see that we generally see an upward trend meaning the conversion rate has increased over the last nine years. It is easy to notice that it increases fastly but decreases slowly. This indicates an overall increasing trend.

The second step consists of checking for stationarity. Stationarity means that the time series does not have a trend, has a constant variance, autocorrelation dependent only on lag, and no seasonal pattern. The autocorrelation function declines to near zero rapidly for a stationary time series. The first graph already suggests a non-stationary time series, and to know about the models to be used, we plot ACF and PACF.



We can easily see that the ACF shows exponential decay starting at lag 1, and the PACF shows a spike at lag 1, then use one autoregressive (p) parameter. (p=1, q=0)(extracted by the plot). So, we can conclude that the time series is not stationary, and by differencing, we obtain:



From the above plot, we see that the resulting series looks mean-stable. So, this is our transformed series after the differencing operation on the original data.

3. Models

3.1 Linear Regression Model

Linear regression is a linear model, e.g., assuming a linear relationship between the input variables (x) and the single output variable (y). More specifically, that y can be calculated from a linear combination of the input variables (x).

Simple linear regression is built to predict exchange rates with the lagged exchange rate. A new column is created, 'Lag_1,' and it has the lagged exchange rate values, i.e., exchange rates of the previous day. The input variable for this model is 'Lagged Price,' and the output variable is the actual exchange rate. This model predicts the current exchange rate based on the previous day's exchange rate. The data is divided into training and test datasets, and the model is built on the training data.

$$y = \beta0 + \beta1*x$$

The model gave an intercept value of 0.999 and a coefficient value of 0.062. Predictions are made using a linear regression model on the test data. The mean squared error for test data was estimated and equaled 0.073.

The above plots are plots of actual price in the test data and the predicted value from the regression model. It can be seen that the trend in the actual exchange rate is captured from the regression model.

3.2 ARMA model

Autoregressive moving average (ARMA) is a model of forecasting in which the methods of autoregression (AR) analysis and moving average (MA) are both applied to time-series data that is well behaved. The model's AR and MA parameters are 1 and 0, respectively.

ARMA Model Results

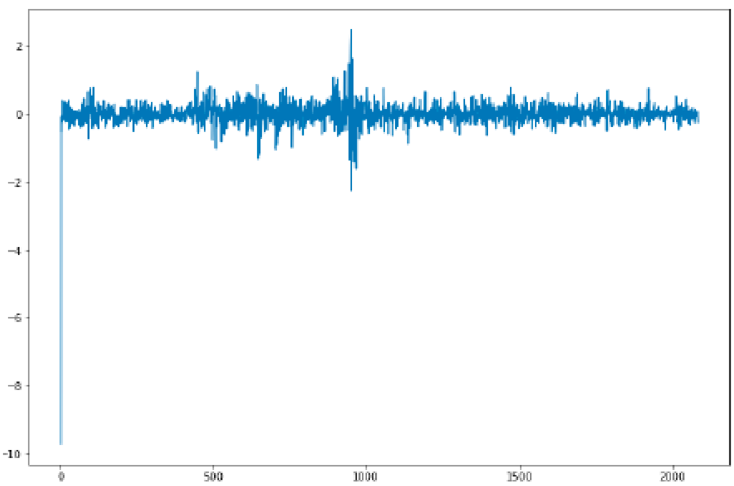
Dep. Variable:	Price	**No. Observations:**	2085
Model:	ARMA(1, 0)	**Log Likelihood**	-261.991
Method:	css-mle	**S.D. of innovations**	0.274
Date:	Wed, 08 Dec 2021	**AIC**	529.983
Time:	16:23:23	**BIC**	546.911
Sample:	0	**HQIC**	536.185

	coef	std err	z	P>|z|	[0.025	0.975]
const	55.9970	6.890	8.127	0.000	42.492	69.502
ar.L1.Price	0.9995	0.001	1926.275	0.000	0.998	1.000

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	1.0005	+0.0000j	1.0005	0.0000

From the model summary, we can see that the intercept or constant term is 55.997, and the coefficient of the AR term is 0.9995. The p-value of the model is much less than 0.05, suggesting that it is a good model with an AIC value of 529.983. The residual plot shows that the residual values are around mean 0.



The mean squared error value on train data is 0.12, but the mean squared error value on test data is very high, making the model unreliable.

3.3 First ARIMA models

Autoregressive integrated moving average (ARIMA) is a model similar to ARMA with an additional part, useful for subtracting an observation from observation at the previous time step to make the time series stationary. For our first model the parameters are: AR parameter p = 1, MA parameter q = 1 and differencing parameter d = 1.

ARIMA Model Results

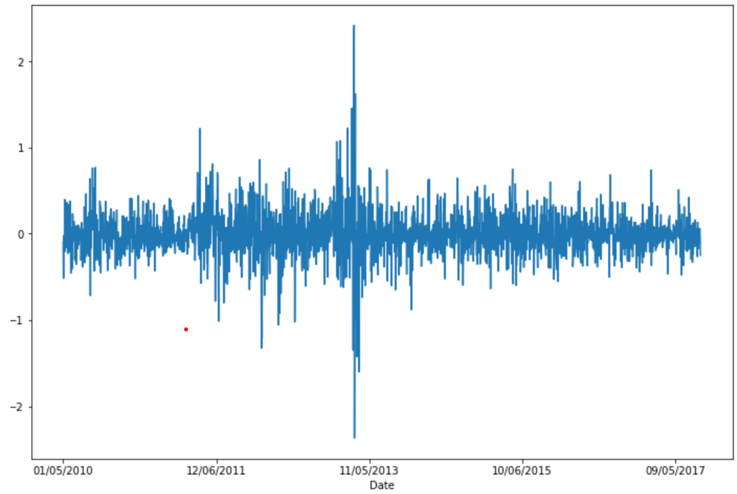
Dep. Variable:	D.Price	**No. Observations:**	2084
Model:	ARIMA(1, 1, 1)	**Log Likelihood**	-253.618
Method:	css-mle	**S.D. of innovations**	0.273
Date:	Wed, 08 Dec 2021	**AIC**	515.236
Time:	16:55:42	**BIC**	537.804
Sample:	1	**HQIC**	523.505

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.0084	0.006	1.352	0.176	-0.004	0.021
ar.L1.D.Price	-0.4186	0.154	-2.720	0.007	-0.720	-0.117
ma.L1.D.Price	0.4755	0.148	3.208	0.001	0.185	0.766

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	-2.3890	+0.0000j	2.3890	0.5000
MA.1	-2.1032	+0.0000j	2.1032	0.5000

From the model summary, we can see that the intercept or constant term is 0.0084, and the coefficient of A.R. and M.A. terms are -0.4186 and 0.4755, respectively. The p-values of the model are much less than 0.05, suggesting that it is a good model with an AIC value of 515.236. The residual plot shows that the residual values are mean stable.



The MSE for this model on test data is 15.432.

3.4 Second ARIMA models

For this second model the parameters are: AR parameter p = 2, MA parameter q = 2 and differencing parameter d = 1.

ARIMA Model Results

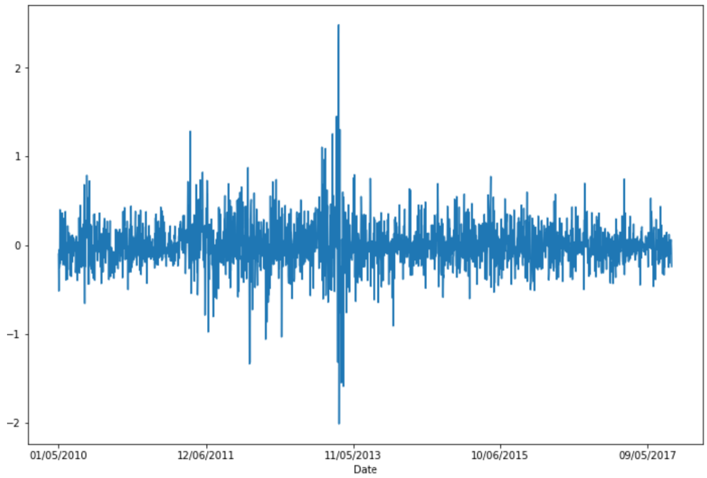
Dep. Variable:	D.Price	**No. Observations:**	2084
Model:	ARIMA(2, 1, 2)	**Log Likelihood**	-237.294
Method:	css-mle	**S.D. of innovations**	0.271
Date:	Tue, 07 Dec 2021	**AIC**	486.589
Time:	23:56:43	**BIC**	520.441
Sample:	01-05-2010	**HQIC**	498.993
	- 12-29-2017		

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
const	0.0085	0.006	1.532	0.126	-0.002	0.019
ar.L1.D.Price	0.2180	0.158	1.379	0.168	-0.092	0.528
ar.L2.D.Price	-0.5031	0.093	-5.433	0.000	-0.685	-0.322
ma.L1.D.Price	-0.1883	0.170	-1.109	0.267	-0.521	0.144
ma.L2.D.Price	0.3822	0.098	3.911	0.000	0.191	0.574

Roots

	Real	**Imaginary**	**Modulus**	**Frequency**
AR.1	0.2166	-1.3930j	1.4098	-0.2254
AR.2	0.2166	+1.3930j	1.4098	0.2254
MA.1	0.2463	-1.5986i	1.6175	-0.2257

From the model summary, we can see that the intercept or constant term is 0.0085 The coefficients of AR1 and AR2 terms are -0.4186 and 0.4755, while MA1 and MA2 are -0.1883 0.3822, respectively. The p-values of the model are much bigger than 0.05, suggesting that it may not be a good model with an AIC value of 486.589, which is less than the previous model. The residual plot shows that the residual values are mean stable.



The MSE for the test data is 15.05369, which is lower than the previous one.

4. Conclusion

Based on the developed models, the linear regression model is the best fit for predicting the future exchange rates since this model gave the least error and captured the trend of the time series data well. On the other hand, the time series ARMA model is not a good fit because of high mean squared error and should not be used for making exchange rate predictions. The ARIMA(1,1,1) model parameters were significant, and also, the AIC value was not high, making it a good model. On the other hand, the ARIMA(2,1,2) model's autoregressive and moving average parameters at lag 1 were insignificant, but the model's AIC value was the lowest.