

IE5202 Project 2 Report

Yang Xiaozhou, A0113538

November 19, 2017

1 Exploratory Data Analysis

The exchange rate from 1980 to 1995 seems to have experienced three distinct phases with a drastic drop around 1986. The three phases are shown in Fig 1. The time series always exhibit strong autocorrelation even beyond 100 lags as shown in Fig 2. SPACF plot shows the lag cuts off after order 1. This indicates an AR(1) model might be appropriate for the time series. Judging from the lag plots in Fig 3, historical values are very good indicators of the future value, at least in the near future (less than 100 days). Also, no obvious seasonal patterns could be observed.

2 Model Building

2.1 Exponential Smoothing Model

With exponentially weighted moving average, due to the strong autocorrelation, the optimal smoothing parameter, α , is found to be 1 for the smallest root mean square error (RMSE), see Fig 4. Similar results are shown when double exponential smoothing model is constructed with the parameter combination of ($\alpha = 1$, $\beta = 0.12$) achieving the least RMSE.

2.2 ARIMA Model

Several ARIMA models are built based on both original data and log-transformed data. 10-fold cross validation RMSEs of 1-step-ahead forecast are reported in Fig 5 for the original data and Fig 6 for log-transformed data. For both types of data, AR(1) and MA(1) models on 1st order differencing transformed data perform the best, with the lowest RMSE of 1.048.

Hence ARIMA(0,1,1) is chosen to model the time series. Since the original data is not stationary, a first order differencing transformation is applied to make it stationary. The resultant time series as shown in Fig 7 is stationary.

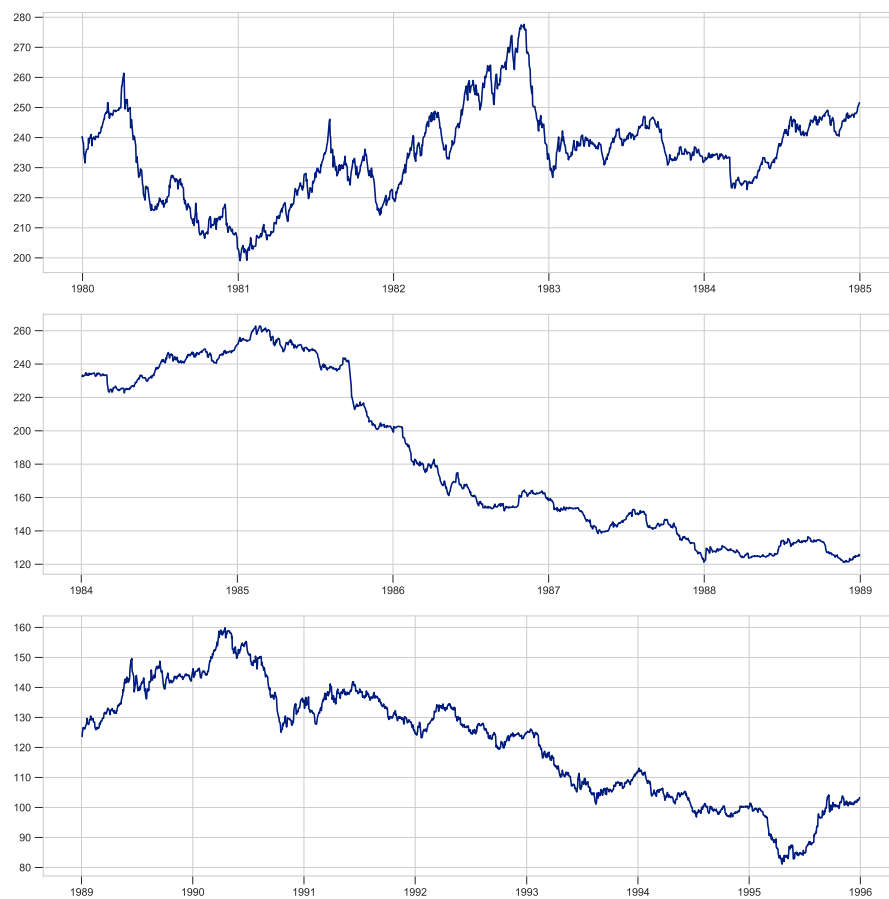


Figure 1: JPY/USD During Three Separate Periods.

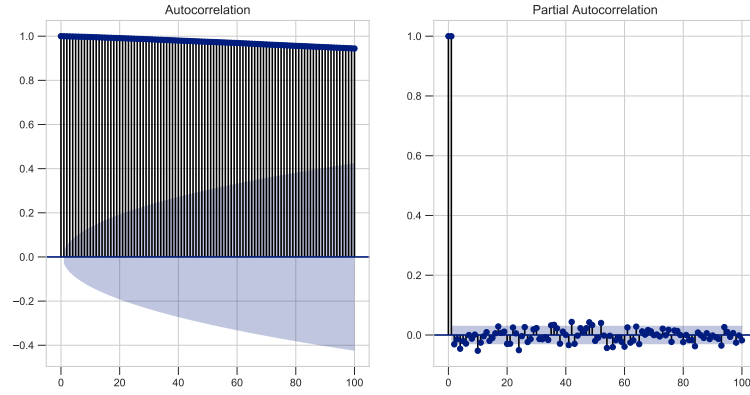


Figure 2: SACF and SPACF Plot for JPY/USD.

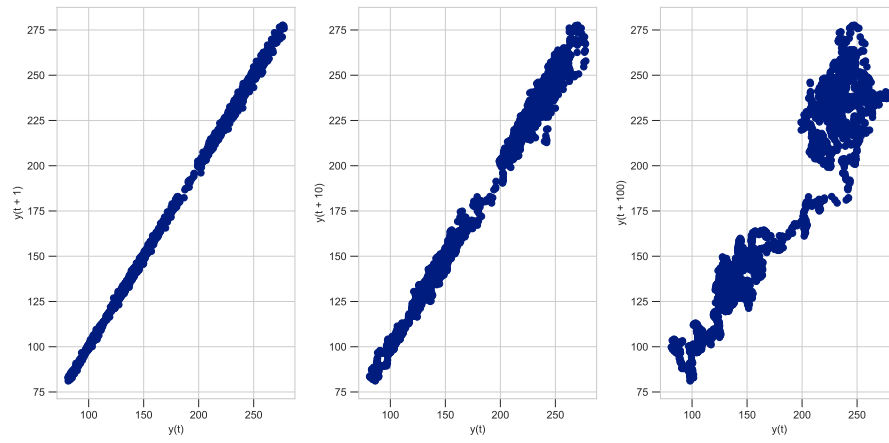


Figure 3: Lag Plots at Three Different Level: 1, 10, 100.

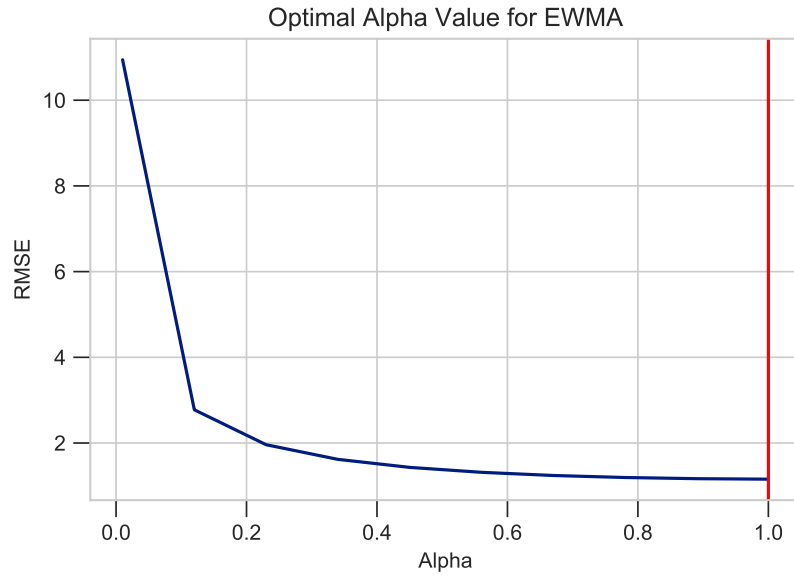


Figure 4: Optimal Alpha.

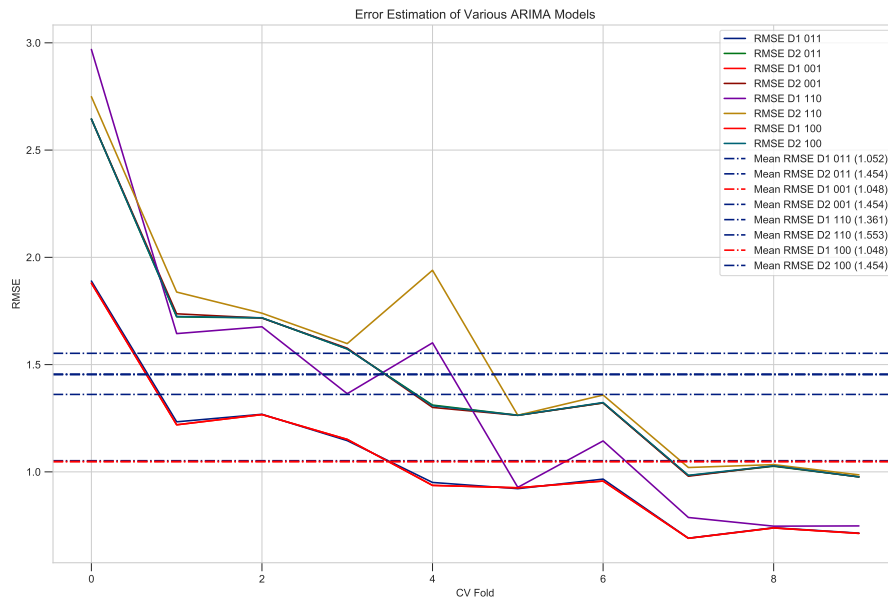


Figure 5: Error Estimation on Original Data.

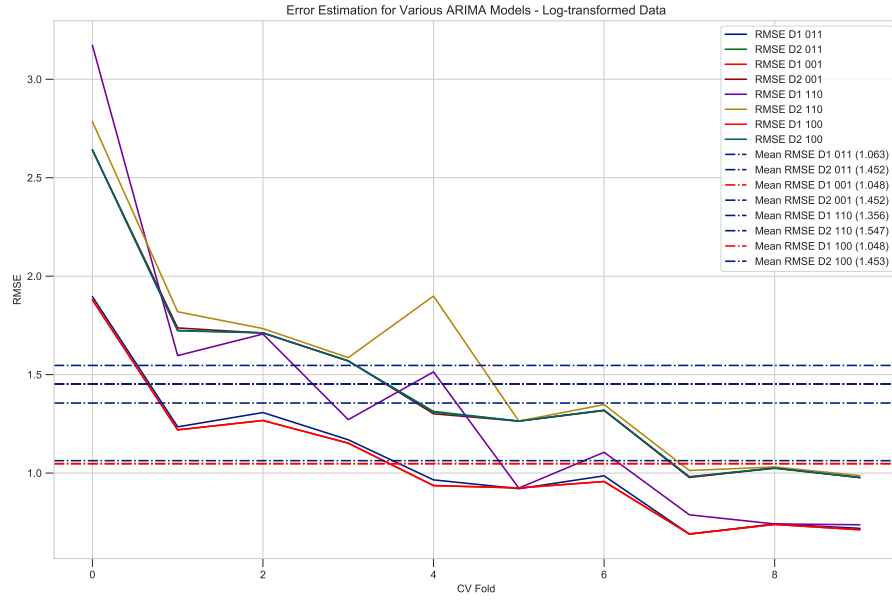


Figure 6: Error Estimation on Log-transformed Data.

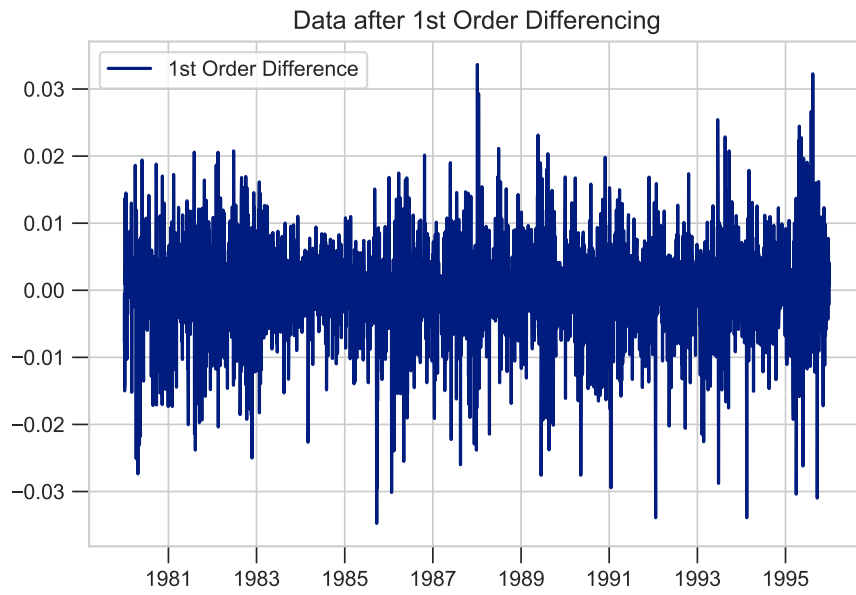


Figure 7: Data after 1st Order Differencing.

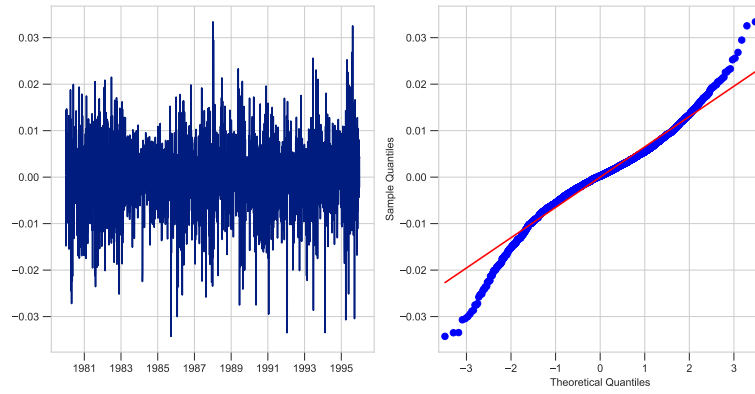


Figure 8: Residuals of ARIMA(0,0,1) Model.

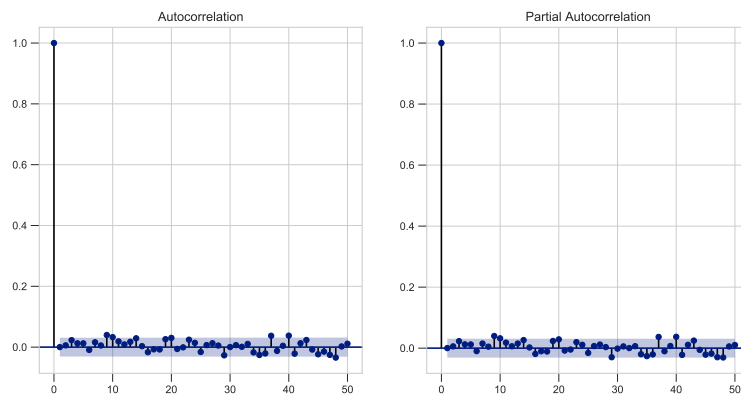


Figure 9: SACF and SPACF of Residuals from ARIMA(0,0,1) Model.

2.3 Regression on Time

A least square regression model is developed to model the time series, initially with only time-related information. These information are year, month, day and day of the week. Also, an indicator variable to indicate whether the year is before or after 1986 where the drastic drop happens. However, this regression model does not perform well by the 10-fold cross validation rmse with a high value of 369.

Regression on time model is not adequate enough to predict the time series, based on the 10-fold rmse, which is much higher than that of the MA(1) model on first order differenced data. Also, the error is heavily pulled up by the period where the exchange rate declined drastically during 1986. Excluding the error of that test, the average RMSE of the rest is around 24, which is still one-fold higher than the previous model.

However, with lagged data added to the predictors, the performance of the model improves tremendously.

Result: Using single currency information, regression on lag_1, lag2-lag1 and indicator variable performs the best, although considering uncertainties, they could be considered the same, for 10-fold cross validation in terms of rmse of 1-step-ahead forecasting.

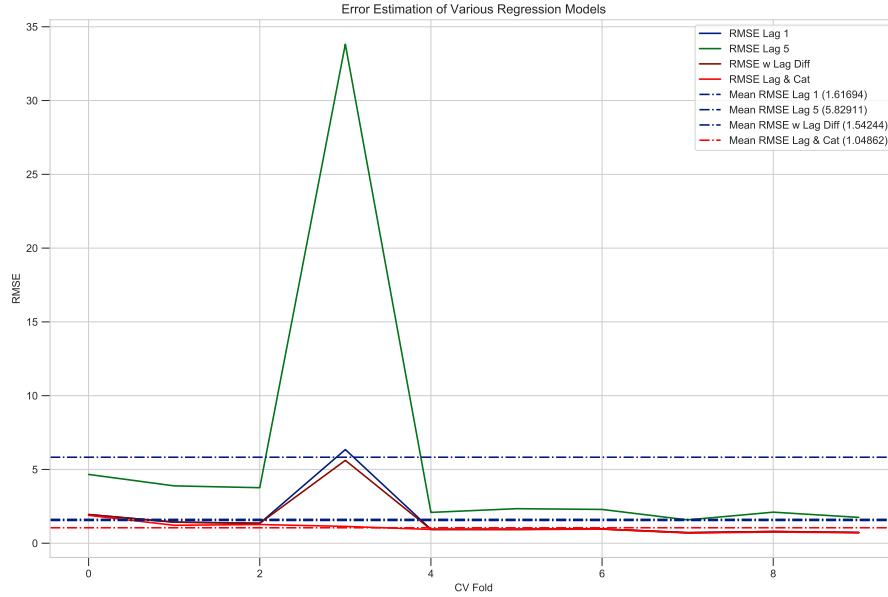


Figure 10: Error Estimation of Various Regression Models.