IE5202 Project 2 Report

Yang Xiaozhou, A0113538

December 5, 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 exhibits strong autocorrelation even beyong 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 Part 1

2.1 Exponential Smoothing Model

With exponentially weighted moving average, due to the strong autocorrelation, the optimal smoothing parameter, alpha, 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 orginal 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, shown in the two figures as the red horizontal dash lines. This is expected since the foreign exchange rate is a highly autocorrelated time series, hence immediate past values are actually great predictors of the next-day value.

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 much more

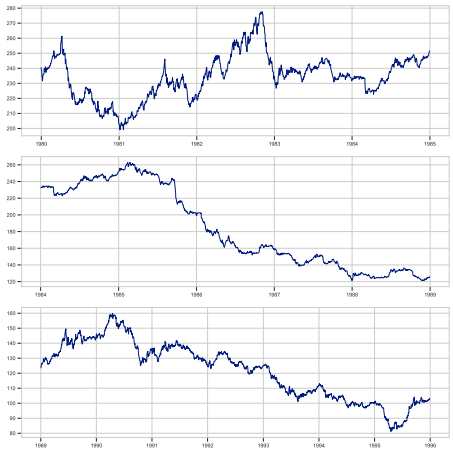

Figure 1: JPY/USD During Three Separate Periods.

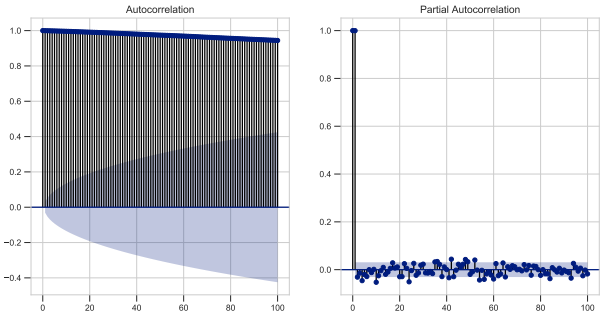

Figure 2: SACF and SPACF Plot for JPY/USD.

stationary than the original series. Residuals from ARIMA(0,1,1) model are uncorrelated and roughly follow a normal distribution with zero mean, as shown in Fig 8 and Fig 9. This indicates that ARIMA(0,0,1) is an appropriate model for the time series.

2.3 Regression on Time

A least square regression model is developed to model the time series, initially with only time-related information. These include variables of *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 as the 10-fold cross validation RMSE is a high value of 369.

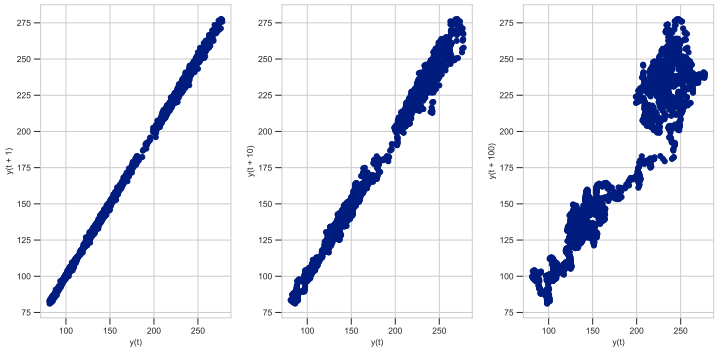

Figure 3: Lag Plots at Three Different Level: 1, 10, 100.

Hence, regression based solely on time information is not adequate enough to predict the time series, which is much higher than that of the ARIMA(0,1,1) model. Also, the error is heavily pulled up by the period where the exchange rate declined drastically during 1986. This is expected since the model cannot learn about this drastic drop based on time predictors alone, while ARIMA model allows the model to learn about the drop from the 1st order differencing. Even if the error of that period for regression is omitted, the average RMSE of the rest is still around 24, which is about one-fold higher than that of the ARIMA model.

However, once the regression model has lagged data as one of the predictors, the performance improves tremendously. This can be seen from the error estimation plot of Fig 10 where four different regression models' performance are compared:

1. Lag 1: Regression on time information and Lag 1 value
2. Lag 5: Regression on time information and Lag 5 value
3. w Lag Diff: Regression on time information, Lag 1 value and the difference between Lag 1 and Lag 2 value
4. Lag & Cat: Regression on only 1986 indicator variable, Lag 1 data and the difference between Lag 1 and Lag 2 data. This is the simplest version among all four models.

As seen from Fig 10, The fourth model performs the best, obtaining the lowest RMSE of 1.048 wihch is identical as the ARIMA(0,1,1) model. This is not surprising since both models utilized information on Lag 1 data and the difference between previous two values. The whole dataset was fitted with the best regression model and Table 1 reports the regression summary. Each of the model parameters are statistically significant and both R-scores are very high. The residuals of the model seem to be uncorrelated with each other (Fig 12) and follows a normal distribution with zero mean (Fig 11).

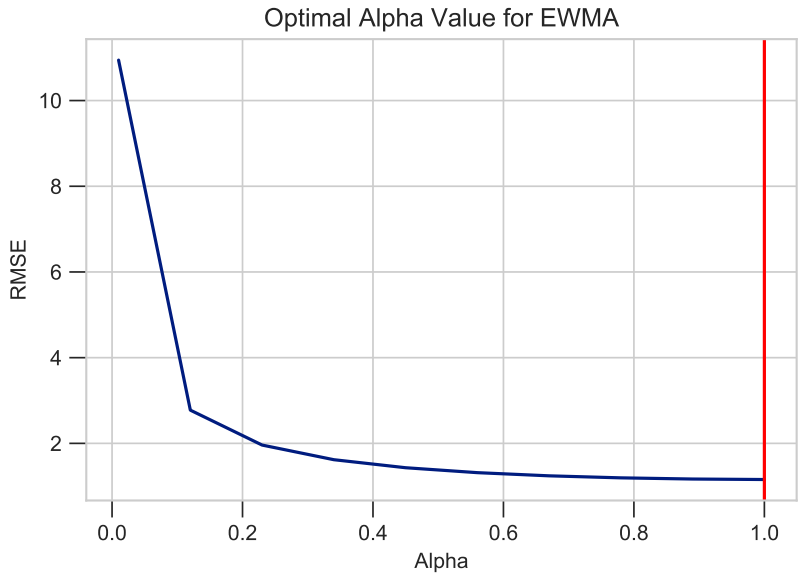

Figure 4: Optimal Alpha.

3 Part 2

Cross validation RMSE does not improve when information of other currencies are included in to the regression model. Hence they are not used when doing the prediction for the test data.

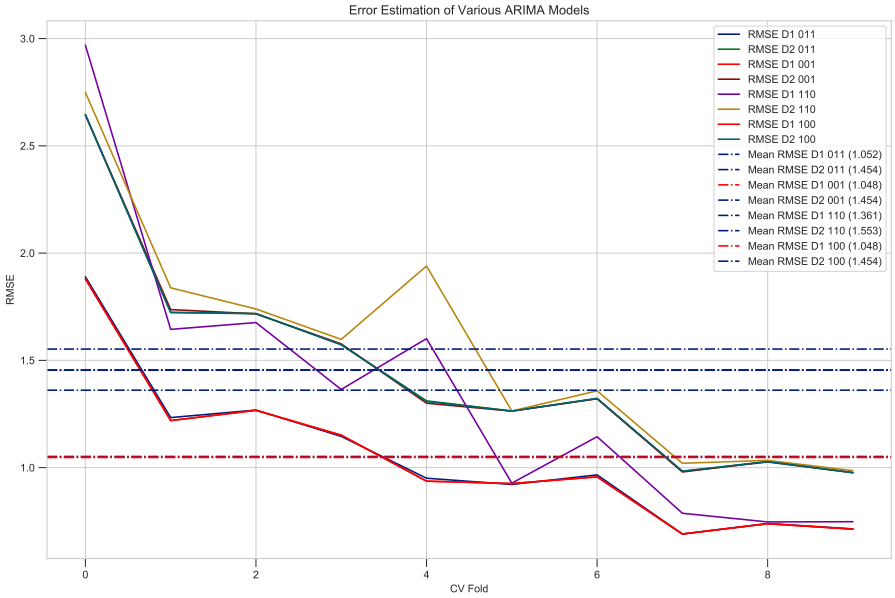

Figure 5: Error Estimation on Original Data.

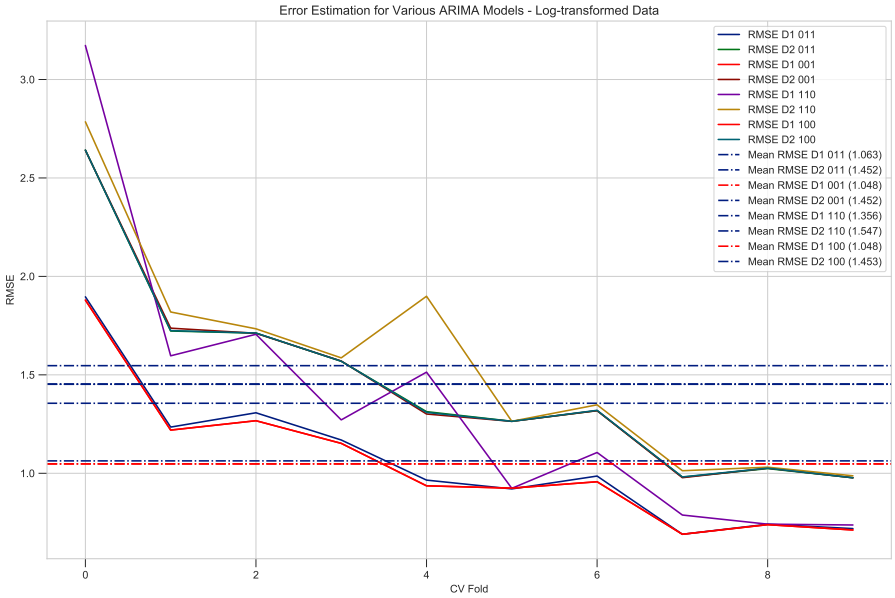

Figure 6: Error Estimation on Log-transformed Data.

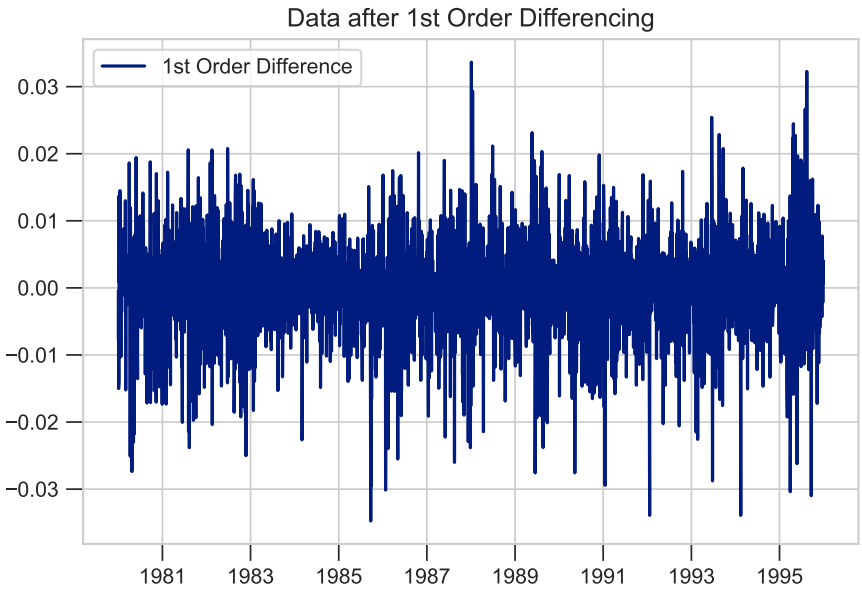

Figure 7: Data after 1st Order Differencing.

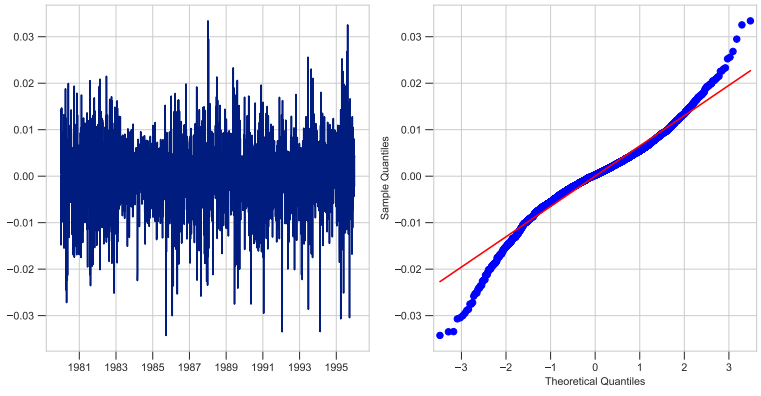

Figure 8: Residuals of ARIMA(0,1,1) Model.

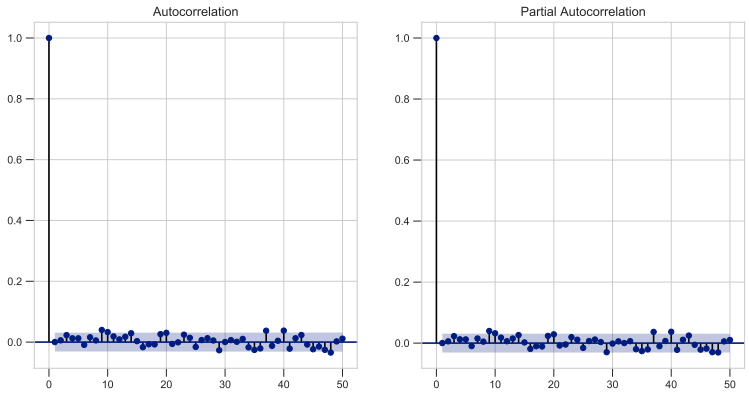

Figure 9: SACF and SPACF of Residuals from ARIMA(0,1,1) Model.

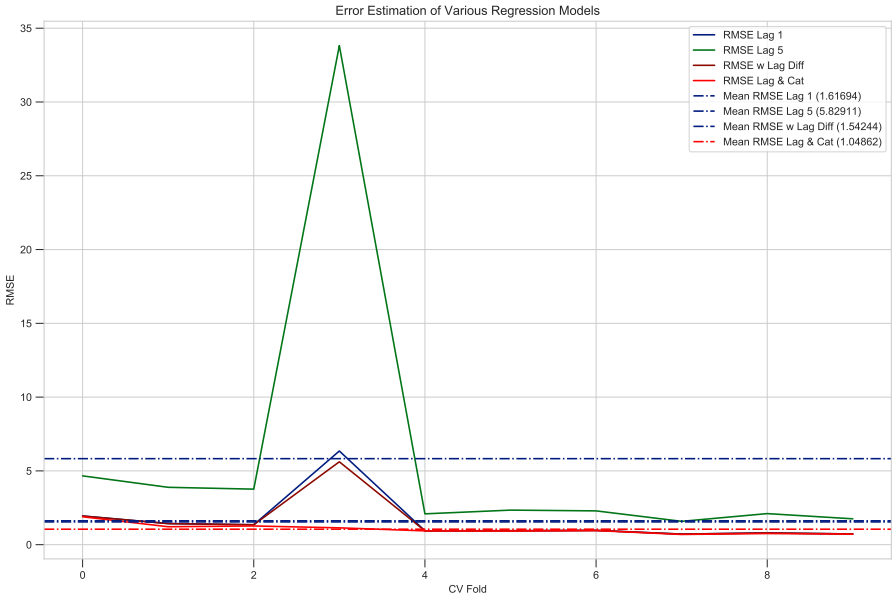

Figure 10: Error Estimation of Various Regression Models.

Dep. Variable:	JPY_USD	**R-squared:**	1.000
Model:	OLS	**Adj. R-squared:**	1.000
Method:	Least Squares	**F-statistic:**	3.032e+06
Date:	Sun, 19 Nov 2017	**Prob (F-statistic):**	0.00
Time:	20:57:14	**Log-Likelihood:**	-6274.6
No. Observations:	4017	**AIC:**	1.256e+04
Df Residuals:	4013	**BIC:**	1.258e+04
Df Model:	3		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	0.3044	0.119	2.551	0.011	0.070	0.538
C(bef_1986)[T.True]	0.2920	0.103	2.840	0.005	0.090	0.494
diff_lag_1_lag_2	0.0316	0.016	2.004	0.045	0.001	0.062
jpy_usd_lag_1	0.9974	0.001	1101.504	0.000	0.996	0.999

Omnibus:	570.043	**Durbin-Watson:**	2.001
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	3070.672
Skew:	-0.563	**Prob(JB):**	0.00
Kurtosis:	7.132	**Cond. No.**	1.49e+03

Table 1: OLS Regression Results

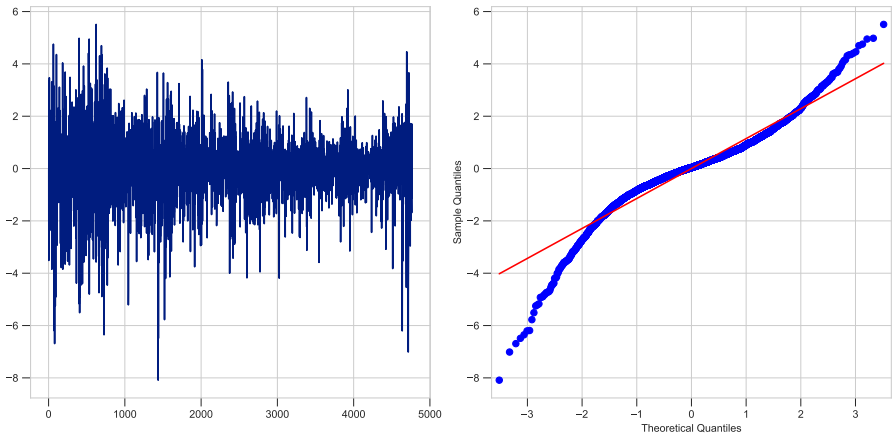

Figure 11: Residuals of Regression Model.

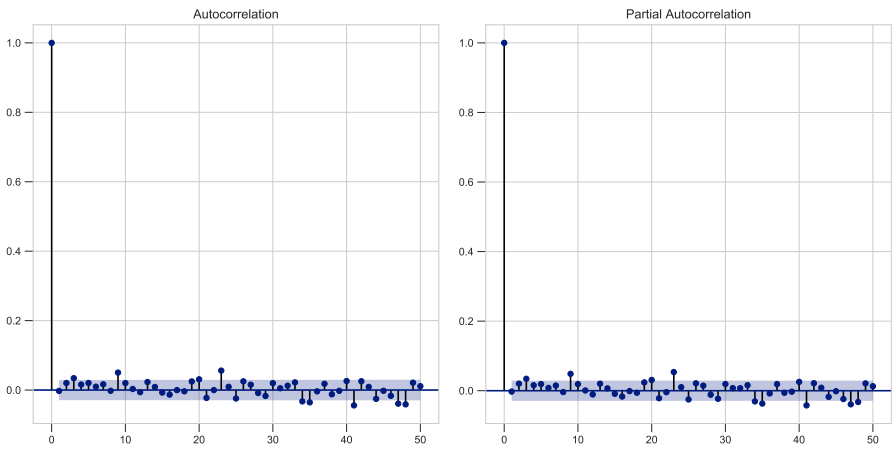

Figure 12: SACF and SPACF of Residuals from Regression Model.