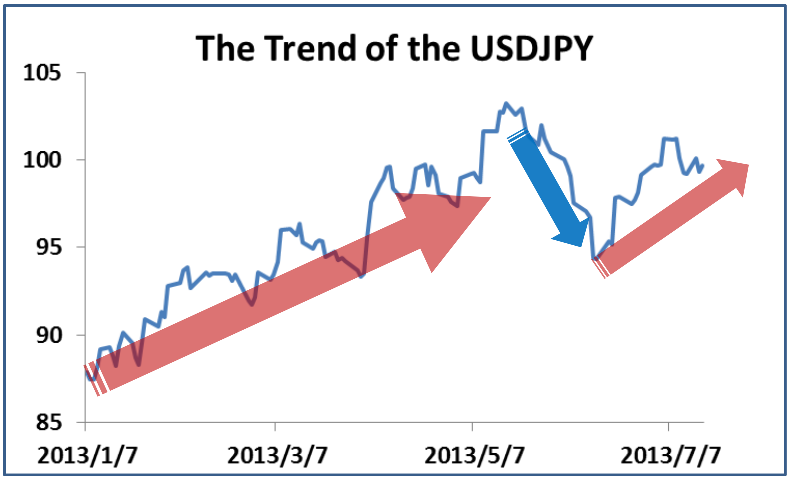
Wei-Lun Wu
VEE Time Series Project
Fall 2013

How to Predict the USDJPY

Introduction

Before the more advanced analysis, we have to describe a more background about the USDJPY. In the first half year, the announcement of Japan government's intervention affects the behavior of the investors, so the Yen is very weak. The USFJPY gets up from 87.99 to 99.66. Within 6 months, the Yen depreciates more than 13.39%. In this document, we try to find a way to capture this up trend of the USDJPY based on the time serious model.

Graph 1：The Trend of the USDJPY from 2013/1/2 to 2013/7/17



Data

- Y_t：The Last Price of the USDJPY
- t：The data frequency, and it represents the Daily in our project
- Y_{t-i}：The Last Price of the Previous i Day

The related data was downloaded from the Bloomberg and you can find the more detail information in the appendix.

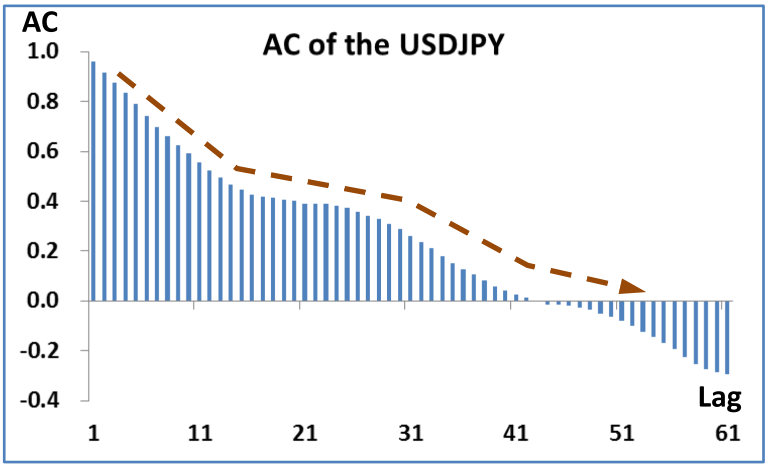
Analysis of the Data

For the general ARMA model, before we set the time series model, we should determine the suitable lag operators. In this project we determine the lag operators based on the Autocorrelation (denotes as AC) and Partial Autocorrelation (denotes as PAC). There are two approaches to be used to determine the suitable lag through the AC and PAC. One is through observing the graph, and another is through the Quenouill and Bartlett's formula. Here we adopt the first method and draw the AC and PAC with the different lag operators in the Graph 1 and Graph 2.

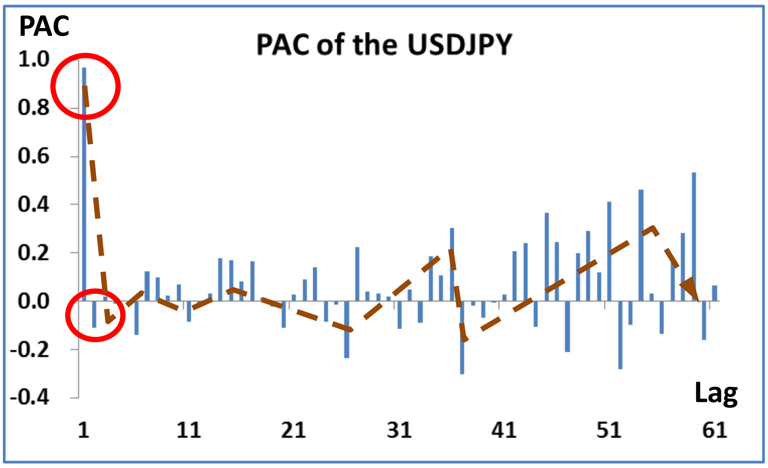
If the model exists the effect of the AR and then the curve of the AC should move very slowly toward the zero. The more time we spend to move toward the zero, the more lag operators we have. If the model exists the effect of the MA and then the PAC should be changing very serious. According to the Graph 1 and 2, we observe that the curve of the AC declines very slowly, and the AC needs about 43 days to achieve the zero. So we consider that AR impacts on the USDJPY. We recommend that the suitable lag of the AR might be 3, 4 or 5.

The PAC has a big initial point, but it goes up and down in a long time and increases its volatility after 30 days. We consider that the effect of the MA is either long or no in the USDJPY. However, the number of sample will decrease through the process of the selecting lag operator. If we choice too longer lag operators, the number of sample will reduce too much. So we prefer the short lag operator than long one and consider that there is no MA effect in the USDJPY.

Graph 2：Autocorrelation of the USDJPY in the Different Lag Operator



Graph 3：Partial Autocorrelation of the USDJPY in the Different Lag Operator



The Model

According to the previous section, we consider that the time serious model of the USDJPY should be a AR(p)[1] process and p might be 3,4 and 5. We set up each model and analyze their statistical properties. We try to find the best one and predict the future price of the USDJPY. In this document, we use the Excel Regression Analysis tool to get the related answer.

Model #1- Includes the Three Time Lag Operators

$$\mathrm{Y}_t = \alpha + \beta_1 \mathrm{Y}_{t-1} + \beta_2 \mathrm{Y}_{t-2} + \beta_3 \mathrm{Y}_{t-3} + \varepsilon_t$$

Table 2：The Regression Statistics of the Model #1

Regression Statistics								
Multiple R	0.979755							
R Square	0.959921							
Adjusted R	0.958919							
Standard E	0.828045							
observatio	124							
ANOVA								
	df	SS	MS	F	ignificance F			
Regressior	3	1970.623	656.8744	958.0187	1.29E-83			
Resiudual	120	82.27911	0.685659					
Total	123	2052.902						
	Coefficeien	:andard Err	t Stat	P-value	Lower 95%	Upper 95%	_owe 95.0%	Jpper 95.0%
Intercept	3.255574	1.738275	1.872877	0.063519	-0.18609	6.697236	-0.18609	6.697236
Y-1	1.080456	0.090973	11.87665	5.48E-22	0.900335	1.260576	0.900335	1.260576
Y-2	-0.13408	0.133318	-1.00572	0.316575	-0.39804	0.12988	-0.39804	0.12988
Y-3	0.020634	0.089798	0.229781	0.818653	-0.15716	0.198427	-0.15716	0.198427

1. We use the last price of the previous 3 day in the Model #1.

2. The R Square is equal to 0.960, and significance F is smaller than 5% significant level. It means that all explanatory variables are significantly and have explanatory power.

3. The effect of each explanatory variable has the significant difference based on the P-value. The highest value is higher than 0.819 but the minimum is less than 0.000.

4. This result is very reasonable. In general, if we want to guess the price in the tomorrow, the major method is based on the modified the price in the today.

[1] *In general it should be rewrote as ARMA(p, q). p represents the lag of the AR, and q represents the lag of the MA. In the USDJPY, there is no effect on the MA, so we get the abbreviation as ARMA(p,0) which is the same with AR(p).*

Model #2- Includes the Four Time Lag Operators

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-3} + \beta_4 Y_{t-4} + \varepsilon_t$$

Table 3：The Regression Statistics of the Model #2

Regression Statistics								
Multiple R	0.979187							
R Square	0.958807							
Adjusted R	0.957411							
Standard E	0.831936							
observatio	123							
ANOVA								
	df	SS	MS	F	ignificance F			
Regressior	4	1900.951	475.2378	686.6432	1.08E-80			
Resiudual	118	81.66987	0.692118					
Total	122	1982.621						
	Coefficeien	:andard Err	t Stat	P-value	Lower 95%	Upper 95%	.owe 95.0%	Jpper 95.0%
Intercept	3.559422	1.77627	2.003874	0.047374	0.041923	7.076922	0.041923	7.076922
Y-1	1.088762	0.091836	11.85547	7.85E-22	0.906901	1.270622	0.906901	1.270622
Y-2	-0.14955	0.135466	-1.10401	0.271837	-0.41781	0.118704	-0.41781	0.118704
Y-3	0.035811	0.135763	0.263774	0.792414	-0.23304	0.304658	-0.23304	0.304658
Y-4	-0.01112	0.090632	-0.1227	0.902554	-0.1906	0.168356	-0.1906	0.168356

1. We increase the last price of the previous day4 into the Model #1 and we get a new model, Model #2. So we use the last price of the previous 4 day in the Model #2.

2. The R Square is equal to 0.959, and significance F is smaller than 5% significant level. It means that all explanatory variables are significantly and have explanatory power.

3. The effect of each explanatory variable has the significant difference based on the P-value. The highest value is higher than 0.902554 but the minimum is less than 0.000. The spread is bigger than the Model #1.

4. This result is very reasonable. In general, if we want to guess the price in the tomorrow, the major method is based on the modified the price in the today. If the data is far away from the today, and it has poor ability of the explanation in the USDJPY.

Model #3- Includes the Five Time Lag Operators

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-3} + \beta_4 Y_{t-4} + \beta_5 Y_{t-5} + \varepsilon_t$$

Table 4：The Regression Statistics of the Model #3

Regression Statistics								
Multiple R	0.978477							
R Square	0.957417							
Adjusted R	0.955581							
Standard E	0.835971							
observatio	122							
ANOVA								
	df	SS	MS	F	ignificance F			
Regressior	5	1822.637	364.5274	521.6122	1.01E-77			
Resiudual	116	81.06631	0.698847					
Total	121	1903.703						
	Coefficeien	tandard Err	t Stat	P-value	Lower 95%	Upper 95%	owe 95.0%	Jpper 95.0%
Intercept	3.882593	1.820169	2.133094	0.035024	0.277518	7.487668	0.277518	7.487668
Y-1	1.081375	0.092636	11.67337	2.69E-21	0.897897	1.264852	0.897897	1.264852
Y-2	-0.13342	0.137305	-0.9717	0.33322	-0.40537	0.13853	-0.40537	0.13853
Y-3	0.021116	0.138181	0.152811	0.878813	-0.25257	0.294801	-0.25257	0.294801
Y-4	-0.01209	0.136952	-0.0883	0.929788	-0.28334	0.259157	-0.28334	0.259157
Y-5	0.003625	0.091118	0.039784	0.968334	-0.17685	0.184096	-0.17685	0.184096

1. We increase the last price of the previous day5 into the Model #2 and we get a new model, Model #3. So we use the last price of the previous 5 day in the Model #3.

2. The R Square is equal to 0.957, and significance F is smaller than 5% significant level. It means that all explanatory variables are significantly and have explanatory power.

3. The effect of each explanatory variable has the significant difference based on the P-value. The highest value is higher than 0.968 but the minimum is less than 0.000. The spread is bigger than the Model #1 and Model #2.

4. This result is very reasonable. In general, if we want to guess the price in the tomorrow, the major method is based on the modified the price in the today. If the data is far away from the today, and it has poor ability of the explanation in the USDJPY.

Durbin-Watson Test (D-W Test)

In this document we use the D-W test to check the one lag serious problem. Notice that D-W test use a special way to reject the null hypothesis or not, and we show this principle in the Graph 4. Besides, the D-W test has a different critical region/acceptance region based on the number of sample and explanatory variables. We show the formulas, hypothesis and related results as follows.

$$\begin{cases} H_0 : \phi = 0 \\ H_1 : \phi \neq 0 \end{cases}$$

$$\hat{\phi} = \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=2}^{n} e_{t-1}^2} \text{ and } d \approx 2(1 - \hat{\phi})$$

e_t ： It represents the residual

Table 5 ： The Related Results of the D-W Test

Model	$\hat{\phi}$	d	Number of the Sample	d_l	d_u	$4 - d_u$	$4 - d_l$	Reject Null Hypothesis
#1	0.01	1.983	100	1.486	1.604	2.514	2.396	No
			150	1.584	1.665	2.416	2.335	No
#2	-0.01	2.015	100	1.461	1.625	2.539	2.375	No
			150	1.571	1.679	2.429	2.321	No
#3	0.00	2.007	100	1.441	1.647	2.559	2.353	No
			150	1.557	1.693	2.443	2.307	No

Graph 4 ： The Principle of the D-W Test

Based on the results of the Table 5, we consider that the total models are no one lag serious problem. Notice that we don't recalculate the new D-W test for the USDJPY but use the parts of the original data from the paper of the Durbin and Watson. Although the critical and acceptance region of the D-W test will be changed based on the number of the sample and explanatory variables, but we make sure that they fall in the range between 100 to 150 (#1 is 124, #2 is 123 and #3 is 122). We use two standards to test each model and consider that each model rejects the null hypothesis in any standard.

Box-Pierce Q Test

In the previous section we make sure that each model has no one lag serious problem. And then we use the Box-Pierce Q test to check the null hypothesis that whole autocorrelation coefficients are zero, because we want to make sure that the residuals are still white noise. Notice that Q test follows Chi-Square distribution and the freedom is the number of the lag which we set in the null hypothesis. We show the formulas, hypothesis and related results as follows.

$$\begin{cases} H_0: Over\ \boldsymbol{m}\ or\ more\ lag\ relationship \\ H_1: No\ more\ \boldsymbol{m}\ lag\ relationship \end{cases}$$

$$\hat{\rho}_j = \frac{\sum_{t=j+1}^{n} e_t e_{t-j}}{\sum_{t=1}^{n} e_t^2} \text{ and } Q = n\sum_{j=1}^{m} \hat{\rho}_j^2$$

e_t：It represents the residual

Table 6：The Related Results of the Box-Pierce Q Test

	Model #1				Model #2				Model #3			
Lag	$\hat{\rho}_j$	Q	X^2	Reject Null Hypothesis	$\hat{\rho}_j$	Q	X^2	Reject Null Hypothesis	$\hat{\rho}_j$	Q	X^2	Reject Null Hypothesis
1	0.01	0.01	3.84	No	-0.01	0.01	3.84	No	0.00	0.00	3.84	No
2	0.00	0.01	5.99	No	0.00	0.01	5.99	No	0.00	0.00	5.99	No
3	0.01	0.02	7.81	No	0.01	0.01	7.81	No	0.01	0.03	7.81	No
4	-0.03	0.15	9.49	No	-0.03	0.09	9.49	No	-0.02	0.08	9.49	No
5	0.13	2.33	11.07	No	0.13	2.33	11.07	No	0.13	2.18	11.07	No
6	-0.09	3.45	12.59	No	-0.10	3.54	12.59	No	-0.11	3.53	12.59	No
7	-0.11	4.83	14.07	No	-0.11	5.05	14.07	No	-0.10	4.80	14.07	No
8	-0.02	4.87	15.51	No	-0.01	5.05	15.51	No	0.00	4.80	15.51	No
9	-0.10	6.23	16.92	No	-0.10	6.27	16.92	No	-0.11	6.17	16.92	No
10	0.08	7.03	18.31	No	0.08	7.01	18.31	No	0.07	6.73	18.31	No
11	0.00	7.03	19.68	No	-0.01	7.03	19.68	No	-0.02	6.76	19.68	No
12	-0.03	7.13	21.03	No	-0.03	7.15	21.03	No	-0.02	6.81	21.03	No
13	-0.13	9.12	22.36	No	-0.11	8.74	22.36	No	-0.11	8.19	22.36	No
14	-0.19	13.69	23.68	No	-0.18	12.80	23.68	No	-0.19	12.51	23.68	No
15	-0.03	13.79	25.00	No	-0.03	12.93	25.00	No	-0.03	12.59	25.00	No
16	-0.12	15.70	26.30	No	-0.12	14.61	26.30	No	-0.12	14.38	26.30	No
17	-0.03	15.85	27.59	No	-0.04	14.78	27.59	No	-0.02	14.45	27.59	No
18	0.02	15.92	28.87	No	0.04	14.97	28.87	No	0.04	14.63	28.87	No
19	0.07	16.55	30.14	No	0.07	15.60	30.14	No	0.08	15.38	30.14	No
20	0.08	17.26	31.41	No	0.08	16.39	31.41	No	0.08	16.18	31.41	No

Based on the results of the Table 6, we consider that the residuals are still white noise. We don't care about the problem of the Autocorrelation in each model.

Conclusion

Table 7：The Summary about the Previous Model

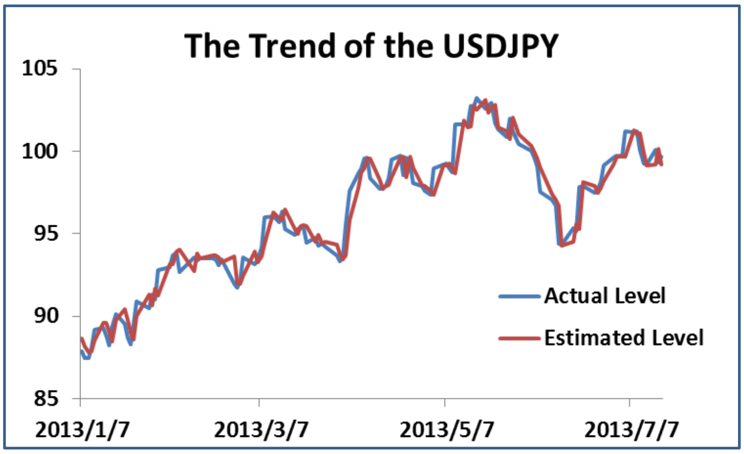
Model	Explanatory Variables	Increase	Adjusted R Square	MSE	F
#1	3	-	0.959	0.686	958.019
#2	4	Previous day 4	0.957	0.692	686.643
#3	5	Previous day 5	0.956	0.699	521.612

According to the results of the Table 7, we consider that the Model #1 is better than others. Because the Model#1 has the highest Adjusted R Square, the lowest Mean Square Error (denotes as MSE) and the largest F-value. So we recommend that the best predicted function is AR(3) and show the formula and Graph as follows：

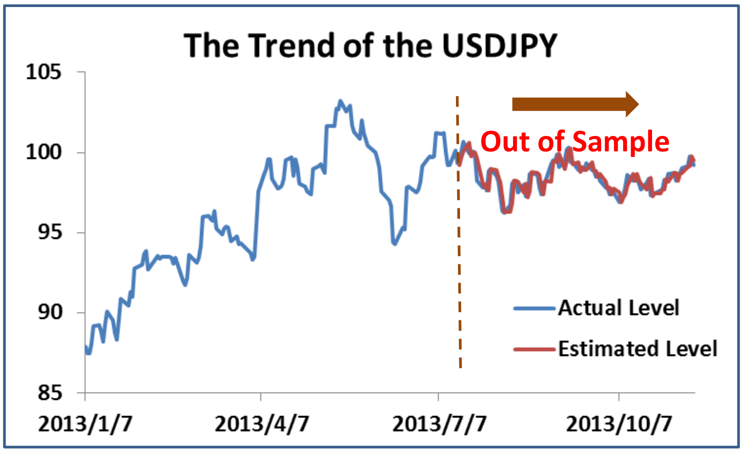
$$\hat{Y}_t = 3.256 + 1.080Y_{t-1} - 0.134Y_{t-2} + 0.021Y_{t-3}$$

Graph 5：Compare with the Actual and Estimated Level

(A) In the Sample



(B) Out of Sample



1. In overall, the fitting of this model is very good. We observe the Graph 5(A) and find that the estimated levels are match with the actual ones in the sample.

2. In addition to the internal fitting, we are more concerned about the prediction out of the sample. We observe the Graph 5(B) and find that the estimated levels exactly predict the actual ones out of sample.

3. Finally, we consider that the AR(3) is a good model for predicting the USDJPY. However, this model only considers about the price itself and we recommend that it might be more reasonable based on increasing some financing variables, just like the spread of US and Japan, the performance of Dow Jones index and the performance of Nikki index.

APPENDIX

The Last Price of the USDJPY from 2013/01/02 to 2013/7/17

Date	USDJPY	Date	USDJPY	Date	USDJPY	Date	USDJPY	Date	USDJPY
2013/1/2	87.04	2013/2/13	93.54	2013/3/26	94.29	2013/5/3	98.99	2013/6/13	94.38
2013/1/3	86.90	2013/2/14	93.38	2013/3/27	94.37	2013/5/6	99.18	2013/6/14	94.31
2013/1/4	88.15	2013/2/15	93.50	2013/3/28	94.22	2013/5/7	99.26	2013/6/18	95.34
2013/1/7	87.89	2013/2/19	93.51	2013/4/1	93.71	2013/5/8	98.98	2013/6/19	95.19
2013/1/8	87.46	2013/2/20	93.43	2013/4/2	93.31	2013/5/9	98.71	2013/6/20	97.81
2013/1/9	87.47	2013/2/21	93.10	2013/4/3	93.50	2013/5/10	101.62	2013/6/21	97.90
2013/1/10	88.10	2013/2/22	93.42	2013/4/4	95.50	2013/5/13	101.62	2013/6/25	97.48
2013/1/11	89.18	2013/2/26	91.95	2013/4/5	97.57	2013/5/14	101.62	2013/6/26	97.71
2013/1/14	89.27	2013/2/27	91.73	2013/4/8	98.66	2013/5/15	102.74	2013/6/27	98.15
2013/1/15	88.81	2013/2/28	92.15	2013/4/9	98.95	2013/5/16	102.67	2013/6/28	99.14
2013/1/16	88.21	2013/3/1	93.59	2013/4/10	99.55	2013/5/17	103.21	2013/7/1	99.62
2013/1/17	89.33	2013/3/5	93.16	2013/4/11	99.59	2013/5/20	102.55	2013/7/2	99.75
2013/1/18	90.1	2013/3/6	93.42	2013/4/12	98.37	2013/5/21	102.73	2013/7/3	99.69
2013/1/21	89.54	2013/3/7	94.15	2013/4/15	97.74	2013/5/22	102.93	2013/7/4	99.73
2013/1/22	88.73	2013/3/8	96.07	2013/4/16	97.83	2013/5/23	101.69	2013/7/5	101.20
2013/1/23	88.3	2013/3/11	96	2013/4/17	97.92	2013/5/24	101.31	2013/7/8	101.16
2013/1/24	89.56	2013/3/12	95.89	2013/4/18	98.37	2013/5/27	100.87	2013/7/9	101.19
2013/1/25	90.91	2013/3/13	95.72	2013/4/19	99.52	2013/5/28	102.00	2013/7/10	100.08
2013/1/29	90.46	2013/3/14	96.34	2013/4/22	99.72	2013/5/29	101.22	2013/7/11	99.24
2013/1/30	91.31	2013/3/15	95.28	2013/4/23	98.57	2013/5/31	100.45	2013/7/12	99.22
2013/1/31	91.03	2013/3/18	94.91	2013/4/24	99.59	2013/6/4	100.02	2013/7/15	100.10
2013/2/1	92.77	2013/3/19	95.29	2013/4/25	99.12	2013/6/5	99.60	2013/7/16	99.33
2013/2/5	93.00	2013/3/20	95.4	2013/4/26	98.05	2013/6/6	99.06	2013/7/17	99.66
2013/2/6	93.68	2013/3/21	95.35	2013/4/29	97.89	2013/6/7	97.56		
2013/2/7	92.68	2013/3/22	94.46	2013/4/30	97.58	2013/6/11	97.04		
2013/2/8	93.87	2013/3/25	94.76	2013/5/2	97.36	2013/6/12	96.68		