



รายงาน

เรื่อง การสร้างโมเดลพยากรณ์ ชุดข้อมูลหุ้น Waste Management Inc (WM) ย้อนหลัง 10ปี

เสนอ

ผศ.ดร. กนกวรรณ ลีโรจนประภา

จัดทำโดย

นายภาณุภณ	สิริวรพาส	65050689
นายศิรัช	เมธาสิทธิ์	65050866
นายสรวิษฐ์	หงษ์เกิด	65050895
นายอภิสิทธิ์	เยียรยงค์	65050989

รายงานนี้เป็นส่วนหนึ่งของวิชา TIME SERIES AND FORECASTING

ภาคเรียนที่ 1 ปีการศึกษา 2567

สาขาวิชาสถิติประยุกต์และการวิเคราะห์ข้อมูล คณะวิทยาศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

สารบัญ

1.ขั้นตอนจัดเตรียมข้อมูล Waste Management Inc (WM)	1
2.การทดสอบแนวโน้ม อิทธิพลของฤดูกาลและ สเตชันนารี	2
2.1 Testing of Normality : Reject H_0 จึงสรุปได้ว่า อนุกรมเวลาไม่มีการแจกแจงปกติ	
2.2 Testing for Trend : Reject H_0 จึงสรุปได้ว่า อนุกรมเวลาราคาหุ้นของ WM มีแนวโน้ม	
2.3 Testing for Seasonal : Failed to Reject H_0 จึงสรุปได้ว่า อนุกรมเวลาไม่มีอิทธิพล ฤดูกาลเข้ามา เกี่ยวข้อง	
2.4 Testing for Stationary : Reject H_0 จึงสรุปได้ว่า ข้อมูลอนุกรมเวลาไม่เป็น Stationary	
3..วิธีบ็อกซ์และเจนกินส์ (Box-Jenkins)	4
3.1 Autocorrelation Function : Lag มีการลดลงอย่างช้าๆหมายถึงข้อมูลอาจจะมีแนวโน้มหรือฤดูกาล	
3.2 Partial Autocorrelations : Lag ที่ 1 มีค่าสูงกว่าขอบเขตความเชื่อมั่นและมีความสัมพันธ์โดยตรงกับ ค่าในอดีต	
4.การจัดข้อมูลที่ไม่สเตชันนารีด้วยการหาผลต่าง (Difference) 1 ครั้ง	5
5.ผลทดสอบโมเดล	6
6.1 ARAIMA (1,1,1) : AIC P = 1 = 66.98190, AIC P = 2 = 68.98190, MSE = 37.5446	
6. ทำการเปรียบเทียบทั้ง 8 Model	7
7.วิธีปรับให้เรียบเอ็กซ์โปเนนเชียลแบบดับเบิ้ล (Double Exponential Smoothing)	7
8.สรุปผล	8

1. ขั้นตอนจัดเตรียมข้อมูล Waste Management Inc (WM)

Reference from: <https://www.investing.com/equities/waste-managemnt-historical-data>

การทำนายข้อมูล Waste Management หรือ หุ้น WM เป็นกลยุทธ์อย่างหนึ่งที่ช่วยให้เราคาดการณ์ราคาของหุ้นในวันถัดไปได้ ทำให้สามารถซื้อหรือขายหุ้นในราคาที่เหมาะสม โดยเราต้องคำนึงถึงปัจจัยอย่างอื่นร่วมด้วยเช่นกัน อย่างสภาพเศรษฐกิจในธุรกิจ

ข้อมูลที่น่าสนใจทั้งหมด 2517 record และมี 7 columns ได้แก่

1. Date(วันที่) 3. Open(ราคาเปิดของหุ้น) 5. Low(ราคาหุ้นที่ต่ำที่สุด) 7. Change %(อัตราการเปลี่ยนแปลง)
2. Price(ราคาหุ้น) 4. High(ราคาหุ้นที่สูงที่สุด) 6. Vol.(จำนวนหุ้นที่ซื้อขาย)

	A	B	C	D	E	F	G
1	Date	Price	Open	High	Low	Vol.	Change %
2	1/2/2014	44.2	44.79	44.8	44.1	1.51M	-1.49%
3	1/3/2014	44.19	44.25	44.46	44	1.46M	-0.02%
4	1/6/2014	43.76	44.38	44.39	43.71	1.91M	-0.97%
5	1/7/2014	44.08	44.04	44.18	43.84	2.14M	0.73%
6	1/8/2014	43.99	44.09	44.14	43.78	1.76M	-0.20%
7	1/9/2014	43.73	44.1	44.17	43.6	1.64M	-0.59%
8	1/10/2014	43.61	44.06	44.14	43.43	2.02M	-0.27%
9	1/13/2014	43.03	43.34	43.47	42.93	2.47M	-1.33%
10	1/14/2014	43.3	43.09	43.42	43.05	1.79M	0.63%
11	1/15/2014	43.19	43.44	43.48	43.16	1.83M	-0.25%
12	1/16/2014	42.86	43.11	43.19	42.75	2.19M	-0.76%
13	1/17/2014	43.12	42.96	43.33	42.84	2.45M	0.61%
14	1/21/2014	42.91	43.41	43.45	42.75	1.95M	-0.49%
15	1/22/2014	42.75	43.04	43.05	42.68	1.74M	-0.37%
16	1/23/2014	42.31	42.49	42.51	41.91	3.06M	-1.03%
17	1/24/2014	41.64	42.05	42.12	41.61	3.04M	-1.58%
18	1/27/2014	41.54	41.61	41.84	41.37	2.46M	-0.24%
19	1/28/2014	41.44	41.63	41.78	41.23	2.14M	-0.24%
20	1/29/2014	41.07	41.15	41.46	40.86	3.04M	-0.89%
21	1/30/2014	41.79	41.37	41.92	41.31	1.86M	1.75%
22	1/31/2014	41.78	41.34	41.84	41.16	2.75M	-0.02%
23	2/3/2014	41	41.72	41.86	40.89	3.10M	-1.87%

ข้อมูลตั้งแต่วันที่ 2/1/2014 ถึง 29/12/2023 รวมทั้งหมด 10 ปี

จากนั้นทำการแปลงข้อมูลจากรายวัน เป็นไตรมาส โดยใช้ Pivot Table จะได้ตัวแปร t ที่เป็นเวลาทั้งหมด 40 ไตรมาส และตัวแปร

Yt คือ ราคาปิดเฉลี่ยของหุ้น โดยเราจะใช้ตัวแปรสองตัวนี้ในการทำนายข้อมูล

Year Average of Price		2018		t	Yt	t	Yt
2014	45.02	Qtr1	85.8	1	41.9	21	97.58
Qtr1	41.9	Qtr2	82.8	2	43.5	22	108.5
Qtr2	43.5	Qtr3	88.67	3	45.8	23	116.78
Qtr3	45.8	Qtr4	89.79	4	48.69	24	112.93
Qtr4	48.69	2019	109.09	5	53.2	25	114.7
2015	51.6	Qtr1	97.58	6	50.28	26	100.97
Qtr1	53.2	Qtr2	108.5	7	49.83	27	110.51
Qtr2	50.28	Qtr3	116.78	8	53.13	28	116.61
Qtr3	49.83	Qtr4	112.93	9	55.2	29	116.88
Qtr4	53.13	2020	110.71	10	60.42	30	138.24
2016	62.03	Qtr1	114.7	11	65.38	31	149.65
Qtr1	55.2	Qtr2	100.97	12	66.88	32	160.79
Qtr2	60.42	Qtr3	110.51	13	71.36	33	151.6
Qtr3	65.38	Qtr4	116.61	14	72.81	34	156.63
Qtr4	66.88	2021	141.69	15	75.99	35	165.07
2017	75.41	Qtr1	116.88	16	81.4	36	160.91
Qtr1	71.36	Qtr2	138.24	17	85.8	37	153.16
Qtr2	72.81	Qtr3	149.65	18	82.8	38	165.32
Qtr3	75.99	Qtr4	160.79	19	88.67	39	161.81
Qtr4	81.4	2022	158.61	20	89.79	40	167.68
2018	86.76	Qtr1	151.6				
Qtr1	85.8	Qtr2	156.63				
Qtr2	82.8	Qtr3	165.07				
Qtr3	88.67	Qtr4	160.91				
Qtr4	89.79	2023	162.01				
		Qtr1	153.16				
		Qtr2	165.32				
		Qtr3	161.81				
		Qtr4	167.68				
		Grand Total	100.24				

2.การทดสอบแนวโน้ม อิทธิพลของฤดูกาลและ สเตชันนารี

(Trend Seasonal and Stationary Analysis)

- Testing of Normality (ทดสอบการแจกแจงปกติ)

ทำการทดสอบการแจกแจงแบบปกติ โดยจะใช้วิธีแอนเดอร์สัน-ดาร์ลิ่ง (Anderson-Darling: AD) เนื่องจากจำนวนข้อมูลที่น่ามาทดสอบ มีจำนวนน้อยกว่าหรือเท่ากับ 50 ตัว

Anderson Daring test

สมมติฐานการทดสอบ

H_0 : อนุกรมเวลาที่มีการแจกแจงปกติ

H_1 : อนุกรมเวลาไม่มีการแจกแจงปกติ

```
Anderson-Darling normality test
data:  z
A = 1.0371, p-value = 0.008866
```

Run ผลลัพธ์ด้วยโปรแกรม R

สรุปผลการทดสอบสมมติฐาน

Reject H_0 เมื่อค่า Sig. หรือค่า p-value ของการทดสอบน้อยกว่าระดับนัยสำคัญ 0.05 จึงสรุปได้ว่า อนุกรมเวลาราคาปิดเฉลี่ยของหุ้น WM ไม่มีการแจกแจงปกติ โดยจะใช้วิธีการทดสอบในกลุ่มทดสอบแบบไม่ใช้พารามิเตอร์ (Non-parametric Test)

- Testing for Trend (ทดสอบแนวโน้มแบบไม่ใช้พารามิเตอร์)

ใช้วิธีทดสอบของ Daniel's Test เนื่องจากเป็นการทดสอบแนวโน้มแบบไม่ใช้พารามิเตอร์

Daniel's Test

สมมติฐานการทดสอบ

H_0 : อนุกรมเวลาราคาหุ้นปิดเฉลี่ยของ WM ไม่มีแนวโน้ม

H_1 : อนุกรมเวลาราคาหุ้นปิดเฉลี่ยของ WM มีแนวโน้ม

```
Spearman's rank correlation rho
data:  dataframe$y and dataframe$Trend
S = 110, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.9896811
```

Run ผลลัพธ์ด้วยโปรแกรม R

สรุปผลการทดสอบสมมติฐาน

Reject H_0 เมื่อค่า Sig. หรือค่า p-value ของการทดสอบน้อยกว่าระดับนัยสำคัญ 0.05 จึงสรุปได้ว่า อนุกรมเวลาราคาปิดเฉลี่ยของหุ้น WM มีแนวโน้ม

- Testing for Seasonal (ทดสอบฤดูกาลแบบไม่ใช่พารามิเตอร์)

ใช้วิธีทดสอบของ Kruskal-Wallis test เนื่องจากเป็นการทดสอบฤดูกาลแบบไม่ใช่พารามิเตอร์

Kruskal-Wallis test

สมมติฐานการทดสอบ

H_0 : อนุกรมเวลาไม่มีอิทธิพลฤดูกาลเข้ามาเกี่ยวข้อง

H_1 : อนุกรมเวลามีอิทธิพลฤดูกาลเข้ามาเกี่ยวข้อง

```
Kruskal-wallis rank sum test
data: detrend by Seasonal
Kruskal-wallis chi-squared =
2.0751, df = 3, p-value = 0.557
```

Run ผลลัพธ์ด้วยโปรแกรม R

สรุปผลการทดสอบสมมติฐาน

Failed to Reject H_0 เมื่อค่า Sig. หรือค่า p-value ของการทดสอบมากกว่าระดับนัยสำคัญ 0.05 จึงสรุปได้ว่า อนุกรมเวลาราคาปิดเฉลี่ยของหุ้น WM ไม่มีอิทธิพลฤดูกาลเข้ามาเกี่ยวข้อง

- Testing for Stationary (ทดสอบ Stationary)

ทำการทดสอบข้อมูลสเตชันนารี (Stationary) ด้วยวิธีการทดสอบของ KPSS test

KPSS test

สมมติฐานการทดสอบ

H_0 : ข้อมูลอนุกรมเวลาเป็น Stationary

H_1 : ข้อมูลอนุกรมเวลาไม่เป็น Stationary

```
#####
# KPSS Unit Root Test #
#####

Test is of type: mu with 3 lags.

Value of test-statistic is: 1.0755

Critical value for a significance level of:
10pct 5pct 2.5pct
critical values 0.347 0.463 0.574
1pct
critical values 0.739
```

Run ผลลัพธ์ด้วยโปรแกรม R

สรุปผลการทดสอบสมมติฐาน

Reject H_0 เนื่องจากค่าสถิติทดสอบ 1.0755 มากกว่าค่าบริเวณวิกฤตที่ 0.463 เมื่อกำหนดระดับนัยสำคัญ 0.05 จึงถือว่าค่าสถิติทดสอบอยู่ในบริเวณวิกฤต สรุปได้ว่า ข้อมูลอนุกรมเวลาไม่เป็น Stationary

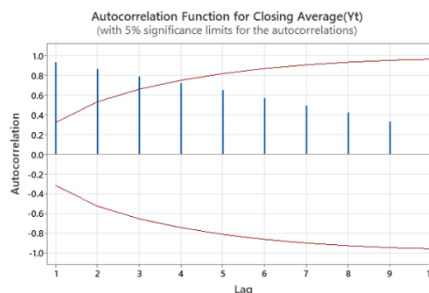
3.วิธีบ็อกซ์และเจนกินส์ (Box-Jenkins)

ตรวจสอบ Stationary ของข้อมูลอนุกรมเวลาจาก ACF และ PACF

- Autocorrelation Funtion

Autocorrelations

Lag	ACF	T	LBQ
1	0.927915	5.87	37.09
2	0.859218	3.29	69.73
3	0.783431	2.42	97.60
4	0.717391	1.95	121.61
5	0.645206	1.61	141.60
6	0.565343	1.32	157.39
7	0.491236	1.10	169.67
8	0.419436	0.91	178.91
9	0.328958	0.70	184.77
10	0.249435	0.53	188.26

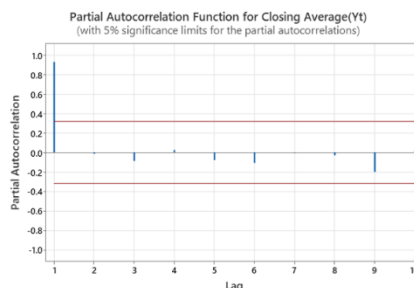


จะสังเกตได้ว่า Lag มีการลดลงอย่างช้าๆ นั้นหมายความว่าข้อมูลมีแนวโน้ม

- Partial Autocorrelations

Partial Autocorrelations

Lag	PACF	T
1	0.927915	5.87
2	-0.013012	-0.08
3	-0.087452	-0.55
4	0.025839	0.16
5	-0.077917	-0.49
6	-0.107917	-0.68
7	-0.002817	-0.02
8	-0.030772	-0.19
9	-0.198923	-1.26
10	0.018062	0.11



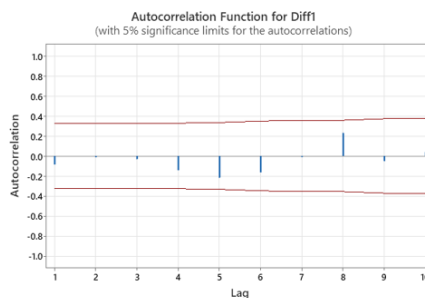
ค่า PACF ที่ Lag 1 มีค่าสูงและเกินขอบเขตความเชื่อมั่น หมายความว่า ค่าปัจจุบันมีความสัมพันธ์โดยตรงกับค่าในอดีตที่อยู่ใกล้ที่สุดอย่างมีนัยสำคัญ PACF นี้บ่งชี้ถึงอนุกรมเวลาที่มีความสัมพันธ์โดยตรงกับค่าในอดีตที่อยู่ใกล้ที่สุดเป็นหลักและมีความสัมพันธ์ดังกล่าว ลดลงอย่างรวดเร็วเมื่อเวลาผ่านไปในแต่ละไตรมาส

4.การขจัดข้อมูลที่ไม่สเตรชันนารีด้วยการหาผลต่าง (Difference) 1 ครั้ง

- Autocorrelation Funtion

Autocorrelations

Lag	ACF	T	LBQ
1	-0.086169	-0.54	0.31
2	-0.012664	-0.08	0.32
3	-0.032479	-0.20	0.37
4	-0.142340	-0.88	1.29
5	-0.217397	-1.32	3.51
6	-0.164822	-0.96	4.83
7	-0.011158	-0.06	4.84
8	0.230673	1.31	7.58
9	-0.051066	-0.28	7.72
10	0.038847	0.21	7.80

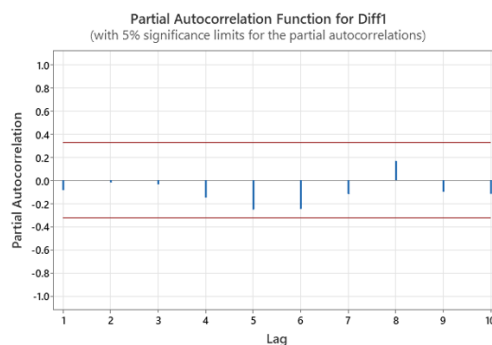


Autocorrelation Function เมื่อหาผลต่าง 1 ครั้ง ทำให้ค่า Lag ลดลงอย่างรวดเร็วและค่า Lag อยู่ในขอบเขตความเชื่อมั่นทุก Lag

- Partial Autocorrelations

Partial Autocorrelations

Lag	PACF	T
1	-0.086169	-0.54
2	-0.020240	-0.13
3	-0.035615	-0.22
4	-0.149947	-0.94
5	-0.253127	-1.58
6	-0.248092	-1.55
7	-0.119905	-0.75
8	0.165013	1.03
9	-0.099886	-0.62
10	-0.117328	-0.73



Partail Autocorrelation Function เมื่อหาผลต่าง 1 ครั้ง ทำให้ค่า Lag ลดลงอย่างรวดเร็วและค่า Lag อยู่ในขอบเขตความเชื่อมั่นทุก Lag

5.ผลทดสอบโมเดล

ในขั้นตอนถัดไปจะเป็นการพยากรณ์อนุกรมเวลา (Time Series) ด้วย AutoRegressive Intergrated Moving Average (ARIMA) โดยจะหา Model ที่ดีที่สุด โดยจะแสดง Model ที่ให้ค่า **Mean Square Error** และค่า **AIC** ที่น้อยที่สุด และผ่าน assumption เบื้องต้นมาแล้ว

ARAMA (1,1,1)

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
AR 1	0.693	0.189	3.67	0.001
MA 1	0.969	0.145	6.71	0.000
Constant	1.0499	0.0573	18.33	0.000

Differencing: 1 regular difference

Number of observations: Original series 40, after differencing 39

Residual Sums of Squares

DF	SS	MS
36	1351.60	37.5446

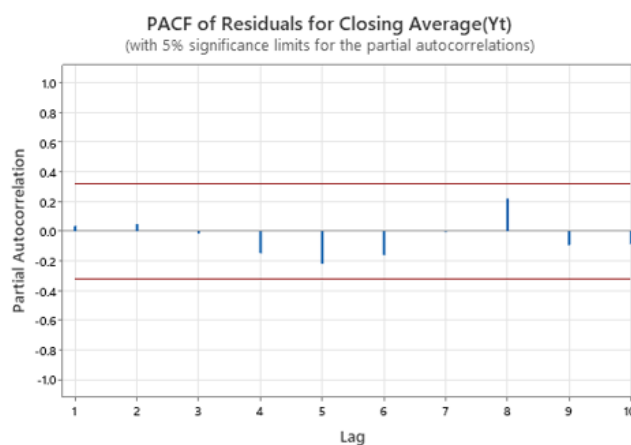
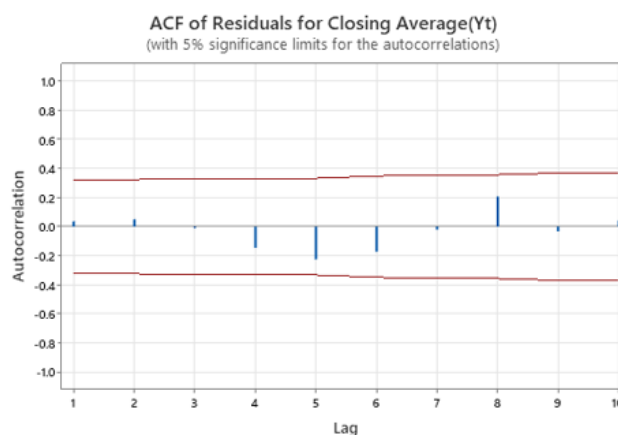
Back forecasts excluded

Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

Lag	12	24	36	48
Chi-Square	8.90	14.12	15.88	*
DF	9	21	33	*
P-Value	0.447	0.864	0.995	*

Forecasts from period 40

Period	Forecast	95% Limits		Actual
		Lower	Upper	
41	170.944	158.932	182.956	
42	174.256	159.426	189.086	
43	177.603	161.451	193.755	
44	180.973	164.122	197.825	



จากการ Run ผลลัพธ์ด้วย Minitab จะสรุปผลเบื้องต้นได้ว่า

1. ตัวแปร AR 1 และ MA 1 มีค่า P-Value ที่น้อยกว่า 0.005 จึง Reject H_0 สรุปได้ว่าตัวแปรตัวนี้เป็นตัวแปรที่เหมาะสม
2. มีค่า Mean Square Residual อยู่ที่ 37.5446
3. ในการทดสอบ Seasonal จะเห็นได้ว่า Lag ที่ 12, 24, 36 มีค่า P-Value ที่มากกว่า 0.05 จึง Failed to Reject H_0 สรุปได้ว่าอนุกรมเวลาไม่มี Seasonal
4. ค่า ACF และ PACF ของ Residual อยู่ภายในขอบเขตที่ระดับนัยสำคัญ 0.05 จึงสรุปได้ว่าโมเดลตัวนี้สามารถใช้ได้
5. โดยสามารถพยากรณ์ราคาหุ้นของ Quarter ที่ 41 ได้ 170.944 และอยู่ในช่วง (158.932, 182.956) ที่ระดับความเชื่อมั่น 95%

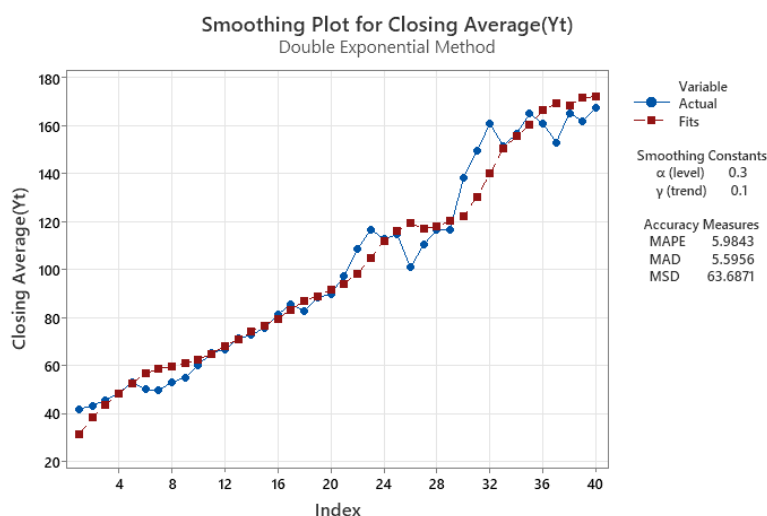
6.ทำการเปรียบเทียบทั้ง 8 Model

Model	p = 1	p = 2	MSE
	AIC	AIC	
ARIMA(0,1,1)	69.097	71.097	42.405
ARIMA(1,1,1)	66.982	68.982	37.545
ARIMA(0,1,2)	69.555	71.555	43.539
ARIMA(1,1,0)	69.102	71.102	42.418
ARIMA(2,1,0)	69.571	71.571	43.579
ARIMA(1,1,2)	67.986	69.986	39.780
ARIMA(2,1,1)	67.987	69.987	39.782
ARIMA(2,1,2)	68.006	70.006	39.825

p คือ ช่วงเวลาล่วงหน้า p ค่าเวลา (ไตรมาส)

จากการเปรียบเทียบทั้ง 8 Model จะสังเกตเห็นได้ว่า ARIMA(1,1,1) จะให้ค่า AIC และ MSE ที่ต่ำที่สุด

7.วิธีปรับให้เรียบเอ็กซ์โปเนนเชียลแบบดับเบิล (Double Exponential Smoothing)



เนื่องจากข้อมูลอนุกรมเวลามีแนวโน้มแต่ไม่มีอิทธิพลของฤดูกาล จึงเลือกใช้ Double Exponential Smoothing

กำหนดให้ $\alpha = 0.3$ และ $\beta = 0.1$ จะได้ผลลัพธ์ MSE = 63.6871

8.สรุปผล

จากการทดสอบ ARIMA ที่มี 8 Model และทดสอบ Double Exponential Smoothing โดยจะใช้ค่า MSE มาเปรียบเทียบโดยใช้ค่าที่น้อยที่สุด โดยจะเปรียบเทียบระหว่าง ARIMA(1,1,1) และ DES

ARIMA(1,1,1): MSE = 37.545

DES: MSE = 63.6871

เมื่อทำการเปรียบเทียบ Model ทั้งสองตัว จะสรุปได้ว่า ARIMA(1,1,1) เป็นโมเดลที่ดีที่สุด เนื่องจากมีค่า MSE ต่ำที่สุด

โดยเราสามารถพยากรณ์ล่วงหน้าได้ตามผลลัพธ์ที่แสดงดังต่อไปนี้

Forecasts from period 40

Period	Forecast	95% Limits		Actual
		Lower	Upper	
41	170.944	158.932	182.956	
42	174.256	159.426	189.086	
43	177.603	161.451	193.755	
44	180.973	164.122	197.825	

ARIMA(1,1,1)

เราจึงใช้วิธีทดสอบ Box-Jenkins ที่เป็น Model ARIMA(1,1,1) ในการทำนายราคาปิดเฉลี่ยของหุ้น เพื่อให้เราทราบได้ว่าในแต่ละช่วงไตรมาสช่วงไหนที่เหมาะสมกับการซื้อมากที่สุด หรือควรขายหุ้นที่สุด