รายงาน

เรื่อง การศึกษาตัวแบบที่เหมาะสมในการพยากรณ์ข้อมูลอนุกรมเวลา

จัดทำโดย

2010511104009 นาย พัชร โสพสโชคชัย

2010511104017 กฤษณะ สุขวี

2010511104022 ชนัญญา แววพานิช

2010511104025 อัครชัย แสนศิลป์ชัย

2010511104029 ภูบดี กลางถิ่น

2010511104032 เตชณัฐ สุวรรณกันทร

นำเสนอ

ผู้ช่วยศาสตราจารย์ บุญหญิง สมร่าง

รายงานนี้เป็นส่วนหนึ่งของวิชา SM319-1

มหาวิทยาลัยหอการค้าไทย

การศึกษาตัวแบบที่เหมาะสมในการพยากรณ์ข้อมูลอนุกรมเวลาชุดที่ 1 จะพิจารณาเทคนิคการพยากรณ์ 2 วิธี คือ

1. **วิธีพยากรณ์ของบอกซ์-เจนกินส์** เป็นวิธีที่ใช้สำหรับเลือกรูปแบบที่เหมาะสมให้กับอนุกรมเวลา โดยพิจารณาจากสหสัมพันธ์ระหว่าง Y ที่คาบเวลา t (Y_t) และที่คาบเวลาที่ผ่านมา (Y_{t-1}, Y_{t-2}, …) เมื่อได้ตัวแบบที่เหมาะสมแล้ว จะใช้ตัวแบบนี้ในการพยากรณ์ Y_{t+1}, Y_{t+2},… ในอนาคต อนุกรมเวลาที่จะกำหนดรูปแบบโดยวิธีบอกซ์-เจนกินส์ จะต้องเป็นอนุกรมเวลาที่อยู่ในสภาวะนิ่ง (Stationary data series) เท่านั้น ซึ่งหมายถึง คงที่ในค่าเฉลี่ย ความแปรปรวน และความแปรปรวนร่วม ไม่แปรผันตามเวลา

ดังนั้นขั้นตอนของวิธีพยากรณ์ของบอกซ์-เจนกินส์ที่สำคัญประกอบด้วย 5 ขั้นตอนได้แก่

(1) ตรวจสอบสภาวะนิ่งโดยพิจารณาจากกราฟของอนุกรมเวลา และจากกราฟฟังก์ชันอัตตสหสัมพันธ์ (Autocorrelation Function: ACF) แทนด้วย r_k

(2) ถ้าตรวจสอบแล้วพบว่าอนุกรมเวลาไม่อยู่ในสภาวะคงที่ในค่าเฉลี่ย จะทำการแปลงเป็นอนุกรมเวลาชุดใหม่ $\{W_t\}$ ที่มีลักษณะคงที่ในค่าเฉลี่ย โดยการหาผลต่างของอนุกรมเวลา ถ้ามีฤดูกาลจะแปลงอนุกรมเวลาด้วยการหาผลต่างฤดูกาล เป็นต้น

(3) กำหนดตัวแบบที่คาดว่าจะเหมาะสมให้กับอนุกรมเวลาโดยพิจารณาจากกราฟ ACF และ PACF (Partial Autocorrelation Function: PACF) แทนด้วย r_{kk}

(4) ประมาณค่าพารามิเตอร์ในสมการที่เลือกไว้ด้วยวิธีกำลังสองน้อยที่สุด (Ordinary least Squares)

(5) ตรวจสอบความเหมาะสมของตัวแบบจากกราฟ ACF และPACF ของส่วนตกค้าง (Residuals: $e_t = Y_t - \hat{Y}_t$)

ด้วยวิธีบอกซ์-เจนกินส์จะได้ตัวแบบอนุกรมเวลาที่เรียกว่า ตัวแบบ ARIMA (Autoregressive Integrated Moving Average) และตัวแบบที่ศึกษาในครั้งนี้เป็นตัวแบบ ARIMA(p,d,q)(P,D,Q) มีรูปแบบดังนี้

$$\phi_p(B)\Phi_P(B^S)(1-B)^d(1\text{-}B^S)^D Y_t = \delta + \theta_q(B)\Theta_Q(B^S)\varepsilon_t$$

โดยที่

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$$

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q$$

$$\Phi_P(B^S) = 1 - \Phi_S B^S - \Phi_{2S} B^{2S} - \ldots - \Phi_{PS} B^{PS}$$

$$\Theta_Q(B^S) = 1 - \Theta_S B^S - \Theta_{2S} B^{2S} - \ldots - \Theta_{QS} B^{QS}$$

$\phi_1, \ldots, \phi_p$ คือ สัมประสิทธิ์การถดถอย (Autoregressive Coefficients)

$\theta_1, \ldots, \theta_q$ คือ สัมประสิทธิ์ค่าเฉลี่ยเคลื่อนที่ (Moving Average Coefficients)

δ คือ ค่าคงที่

B คือ ตัวดำเนินการถอยหลังเวลา (Backward Shift Operator) นั่นคือ $B^m Y_t = Y_{t-m}$

d คือ จำนวนครั้งของการทำผลต่างเพื่อให้อนุกรมเวลา $\{Y_t\}$ เป็นอนุกรมเวลาที่อยู่ในสภาวะคงที่ในค่าเฉลี่ย

D คือ จำนวนครั้งของการทำผลต่างฤดูกาล

p คือ อันดับของตัวแบบการถดถอย

P คือ อันดับของตัวแบบการถดถอยฤดูกาล

q คือ อันดับของตัวแบบค่าเฉลี่ยเคลื่อนที่

Q คือ อันดับของตัวแบบค่าเฉลี่ยเคลื่อนที่ฤดูกาล

S คือ จำนวนฤดูกาล

a_t คือ ตัวแปรสุ่มอิสระและมีการแจกแจงปกติ ซึ่งมีค่าเฉลี่ยเท่ากับศูนย์และความแปรปรวนคงที่ ให้เท่ากับ σ_ε^2 เรียก a_t ว่าความคลาดเคลื่อนสุ่ม หรือ กระตุกสุ่ม (Random Shocks)

2. **เทคนิคการปรับให้เรียบ** เป็นวิธีการพยากรณ์ข้อมูลโดยการขจัดความผันแปรที่ผิดปกติออก ทำให้เห็นองค์ประกอบอื่นของอนุกรมเวลา เพื่อจะสามารถพยากรณ์ค่าอนุกรมเวลาในอนาคตได้ เทคนิคการปรับให้เรียบมีหลายวิธี ขึ้นกับลักษณะของข้อมูลอนุกรมเวลา

ถ้าอนุกรมเวลามีทั้งแนวโน้มและฤดูกาล จะใช้**วิธีการทำให้เรียบแบบเอ็กซ์โพเนนเชียลด้วยวิธีของวินเตอร์** (Winter's three parameter trend and seasonality method หรือ Winter's method) โดยมีตัวแบบดังนี้

ตัวแบบเชิงคูณ $Y_t = (\beta_0 + \beta_1 t)S_t\varepsilon_t$

สมการพยากรณ์ $\hat{Y}_{t+m} = (a_t + b_t m)\hat{S}_t \quad ; t \le p$

$\hat{Y}_{t+m} = (a_t + b_t m)\hat{S}_{t-p+m} \quad ; t>p$

โดยที่

$$a_t = \frac{\alpha Y_t}{\hat{S}_t} + (1-\alpha)(a_{t-1} + b_{t-1}) \quad ; t \le p$$

$$a_t = \frac{\alpha Y_t}{\hat{S}_{t-p}} + (1-\alpha)(a_{t-1} + b_{t-1}) \quad ; t>p$$

$$b_t = \gamma(a_t - a_{t-1}) + (1-\gamma)b_{t-1}$$

และ

$$\hat{S}_t = \frac{\delta Y_t}{a_t} + (1-\delta)\hat{S}_{t-p} \quad ; t>p$$

เมื่อ p คือ จำนวนฤดูกาล

m คือ คาบเวลาที่ต้องการพยากรณ์ไปข้างหน้า

a_t คือ ค่าประมาณระดับอนุกรมเวลา โดย α $(0 < \alpha < 1)$ แทน ค่าคงที่ปรับให้เรียบสำหรับปรับระดับของอนุกรมเวลา

b_t คือ ค่าประมาณแนวโน้ม โดย γ $(0 < \gamma < 1)$ แทน ค่าคงที่ปรับแนวโน้มหรือความชันของอนุกรมเวลา

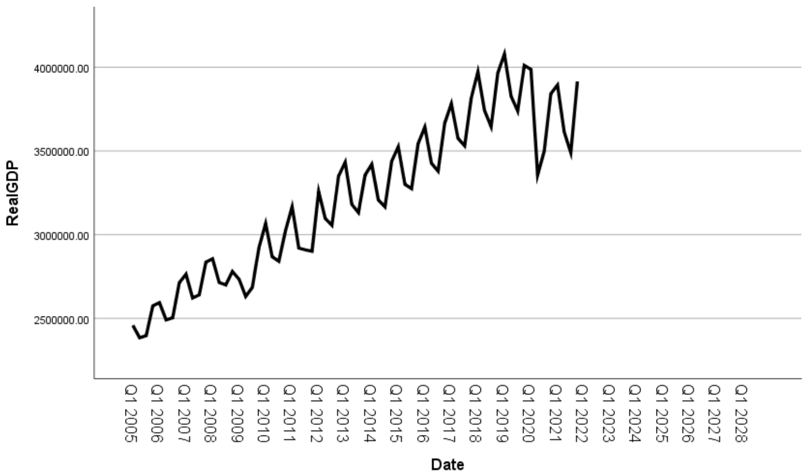
$\hat{S}_t$ คือ ค่าประมาณฤดูกาล โดย $\delta(0 < \delta < 1)$ แทน ค่าคงที่ปรับให้เรียบสำหรับฤดูกาลของอนุกรมเวลา

สรุปผลการศึกษา

ข้อมูลที่ศึกษาชุดที่ 1 คือ อนุกรมเวลา Real GDP ตั้งแต่ไตรมาศที่ 1 ค.ศ.2005 ถึงไตรมาสที่ 4 ค.ศ.2021 จำนวน 68 ไตรมาส และทำการวิเคราะห์โดยใช้โปรแกรมสำเร็จรูป SPSS

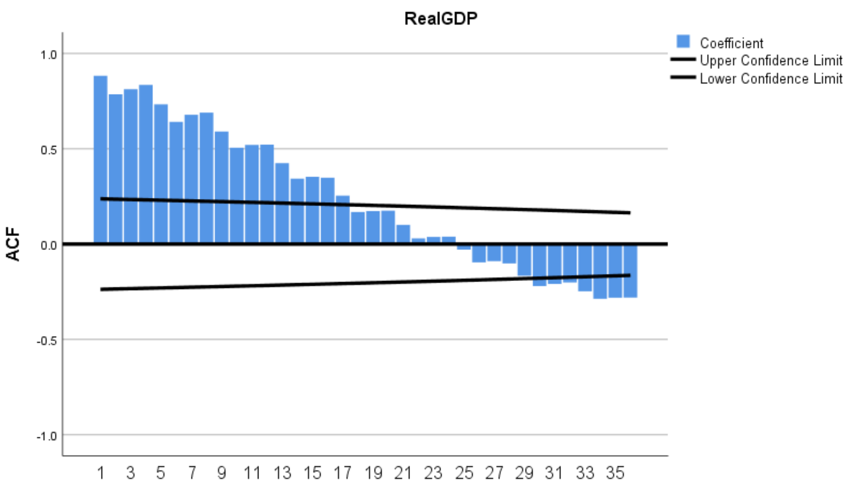
1. วิธีบอกซ์-เจนกินส์ ผลการศึกษาในแต่ละขั้นตอนเป็นดังนี้

(1) ตรวจสอบสภาวะนิ่ง จาก Sequence Chart ของอนุกรมเวลา พบว่าการเคลื่อนไหวของอนุกรมเวลาของ Real GDP มีความไม่คงที่ในค่าเฉลี่ย และมีอิทธิพลของฤดูกาล ดังรูปที่ 1

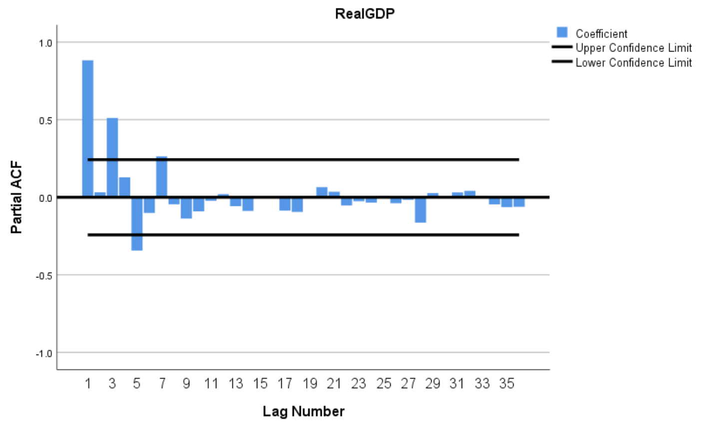


รูปที่ 1 การเคลื่อนไหวของอนุกรมเวลา

(2) เมื่อพิจารณาจากกราฟ ACF พบว่าการเคลื่อนไหวของ r_k มีลักษณะลดลงช้า และมีลักษณะเป็นวงจร แสดงว่าอนุกรมเวลาอยู่ในสภาวะไม่คงที่ (Non Stationary) ในค่าเฉลี่ย และมีอิทธิพลของฤดูกาล ดังรูปที่ 2

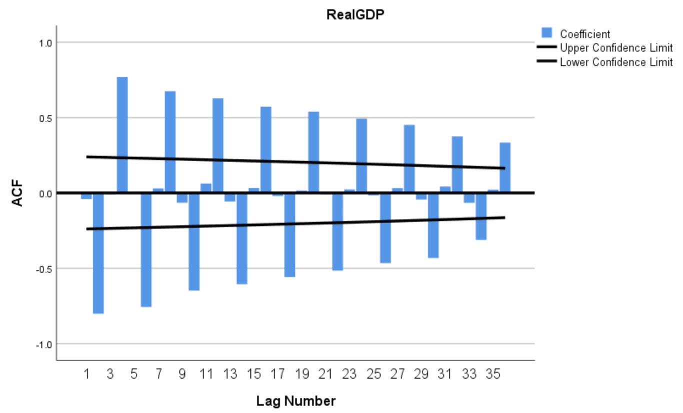


รูปที่ 2 กราฟ ACF ของอนุกรมเวลา

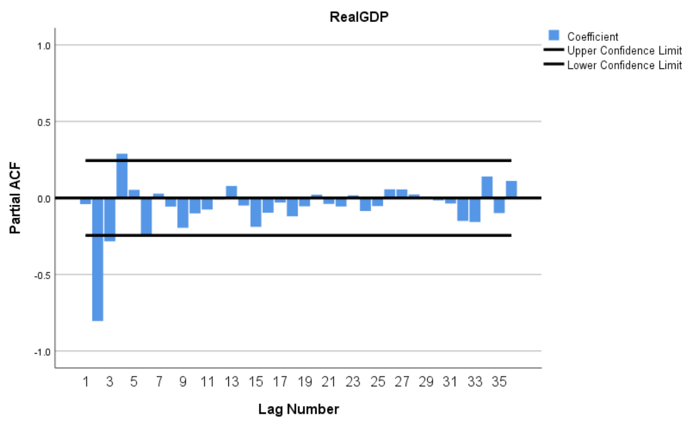


รูปที่ 3 กราฟ PACF ของอนุกรมเวลา

หลังจากหาผลต่าง 1 ครั้ง จากกราฟ ACF r_k ยังมีลักษณะการเคลื่อนไหวที่มีลักษณะคล้ายกันและค่อยๆ ลดลงช้าๆ แสดงว่าอนุกรมเวลายังอยู่ในสภาวะที่มีอิทธิพลของฤดูกาลอยู่ ดังรูปที่ 4 จึงต้องแปลงข้อมูลโดยการหาผลต่างของฤดูกาล 1 ครั้ง

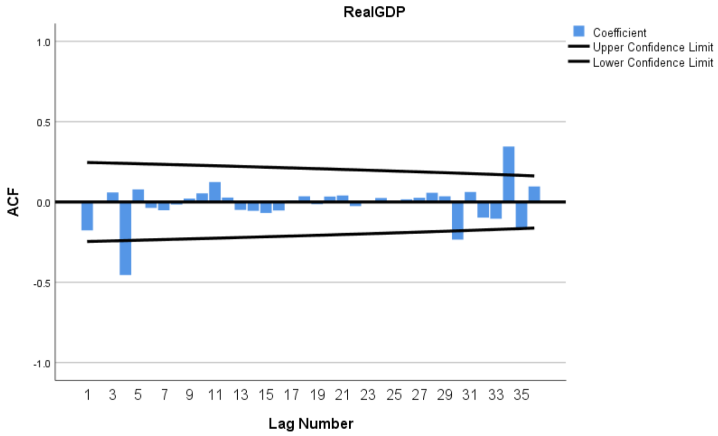


รูปที่ 4 กราฟ ACF ของอนุกรมเวลาเมื่อหาผลต่างอนุกรมเวลา 1 ครั้ง

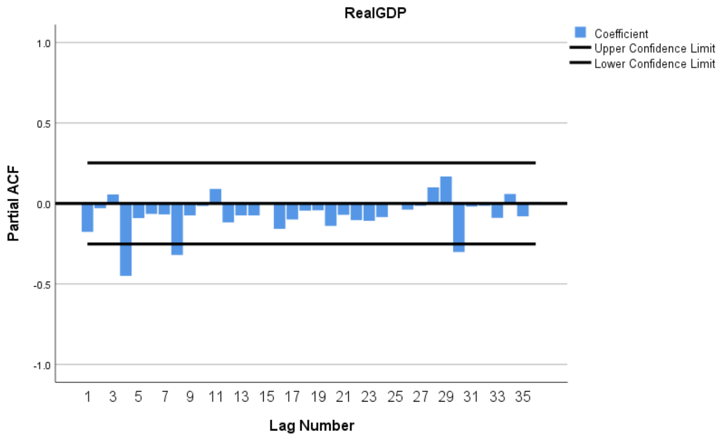


รูปที่ 5 กราฟ PACF ของอนุกรมเวลาเมื่อหาผลต่างอนุกรมเวลา 1 ครั้ง

กราฟ ACF และ PACF ของอนุกรมเวลาชุดใหม่ หลังจากหาผลต่างฤดูกาล 1 ครั้ง และผลต่างอนุกรมเวลา 1 ครั้ง ได้กราฟ ACF และ PACF ดังรูปที่ 6 และ 7



รูปที่ 6 กราฟ ACF เมื่อหาผลต่างฤดูกาล 1 ครั้ง และผลต่างอนุกรมเวลา 1 ครั้ง



รูปที่ 7 กราฟ PACF เมื่อหาผลต่างฤดูกาล 1 ครั้ง และผลต่างอนุกรมเวลา 1 ครั้ง

(3) กำหนดตัวแบบที่เหมาะสม โดยพิจารณาจากกราฟ ACF และ PACF ของอนุกรมเวลาชุดใหม่ พบว่า r_k มีค่าสูงที่ $k = 4$ ค่าเดียว และมีค่าเข้าใกล้ศูนย์ ที่ $k = 8, 12, 16, \ldots$ ส่วน r_{kk} มีค่าสูงที่ $k = 4$ และมีค่าลดลงเร็วๆ ที่ $k = 8, 12, 16, \ldots$ แสดงถึง ตัวแบบฤดูกาล SMA(1)

ดังนั้นตัวแบบที่เหมาะสมคือ ARIMA(0,1,0)(0,1,1)

(4) จากตัวแบบที่กำหนด ARIMA(0,1,0)(0,1,1) ของอนุกรมเวลา W_t ประมาณค่าารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุด ได้ค่าพารามิเตอร์ดังผลลัพธ์ข้างล่างนี้

Time Series Modeler

Model Description

			Model Type
Model ID	Real GDP	Model_1	ARIMA(0,1,0)(0,1,1)

ARIMA Model Parameters

					Estimate	SE	t	Sig.
RealGDP-Model_1	RealGDP	Natural Logarithm	Constant		-.001	.001	-.666	.508
			Difference		1			
			Seasonal Difference		1			
			MA, Seasonal	Lag 1	.772	.118	6.540	.000

ตารางที่ 1 ค่าพารามิเตอร์ของตัวแบบ ARIMA(0,1,0)(0,1,1) ที่ประมาณได้ และแปลงข้อมูลด้วย Natural Logarithm

ตัวแบบพยากรณ์ดังกล่าวให้ค่าความผิดพลาดของการพยากรณ์ MAPE = 1.812 ดังตาราง Model Summary

Model Summary

Model Fit

Fit Statistic	Mean	SE	Minimum	Maximum	Percentile 5	10	25	50	75	90	95
Stationary R-squared	.379	.	.379	.379	.379	.379	.379	.379	.379	.379	.379
R-squared	.957	.	.957	.957	.957	.957	.957	.957	.957	.957	.957
RMSE	92715.572	.	92715.572	92715.572	92715.572	92715.572	92715.572	92715.572	92715.572	92715.572	92715.572
MAPE	1.812	.	1.812	1.812	1.812	1.812	1.812	1.812	1.812	1.812	1.812
MaxAPE	11.475	.	11.475	11.475	11.475	11.475	11.475	11.475	11.475	11.475	11.475
MAE	59064.706	.	59064.706	59064.706	59064.706	59064.706	59064.706	59064.706	59064.706	59064.706	59064.706
MaxAE	385065.856	.	385065.856	385065.856	385065.856	385065.856	385065.856	385065.856	385065.856	385065.856	385065.856
Normalized BIC	23.006	.	23.006	23.006	23.006	23.006	23.006	23.006	23.006	23.006	23.006

ตารางที่ 2 ค่าความผิดพลาดของการพยากรณ์ของตัวแบบ ARIMA(0,1,0)(0,1,1) และแปลงข้อมูลด้วย Natural Logarithm

(5) ตรวจสอบความเหมาะสมของตัวแบบ โดยตรวจสอบว่าค่าความคลาดเคลื่อนจากตัวแบบมีค่าเฉลี่ยเท่ากับศูนย์ มีการแจกแจงแบบปกติ และค่าความคลาดเคลื่อนแต่ละค่าไม่มีอัตตสัมพันธ์กัน (เป็นอิสระกัน)

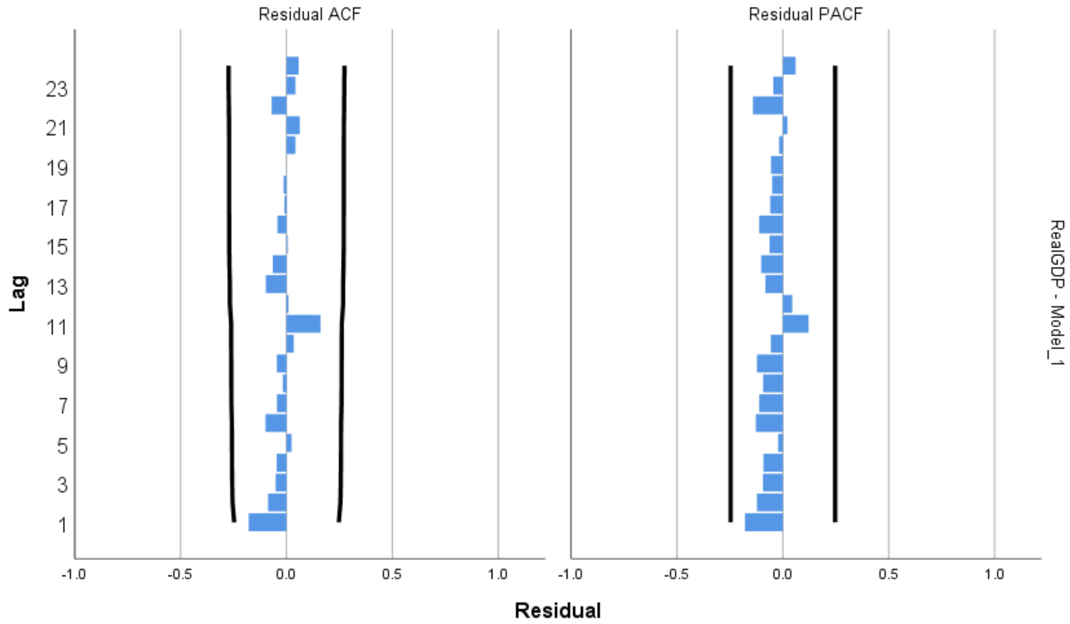
จากตารางที่ 3 Model Statistics ค่า P-value ของการทดสอบ Ljung-Box มีค่า = 0.977 แสดงว่าความคลาดเคลื่อนจากตัวแบบ ไม่มีอัตตสหสัมพันธ์กัน (เป็นอิสระกัน ไม่มีปัญหา Autocorrelation)

Model Statistics

Model	Number of Predictors	Model Fit statistics: Stationary R-squared	Ljung-Box Q(18): Statistics	DF	Sig.	Number of Outliers
RealGDP-Model_1	0	.379	7.470	17	.977	0

ตารางที่ 3 แสดงค่าสถิติ Ljung-Box และค่า P-value

เมื่อตรวจสอบจากกราฟ ACF และ PACF -ของส่วนตกค้าง (Residual) พบว่า r_k และ r_{kk} ของส่วนตกค้างตกอยู่ในขอบเขตความเชื่อมั่นร้อยละ 99 ทุกค่า lag ดังรูปที่ 8 แสดงว่าค่าความคลาดเคลื่อนไม่เกิดปัญหา Autocorrelation



รูปที่ 8 กราฟ ACF และ PACF ของส่วนตกค้าง

และเมื่อตรวจสอบค่า P-value ของค่าสถิติทดสอบ Kolmogolov-Sirminov มีค่าเท่ากับ 0.010 แสดงว่าค่าความคลาดเคลื่อนสุ่มมีการแจกแจงแบบไม่ปกติ ดังตารางที่ 4

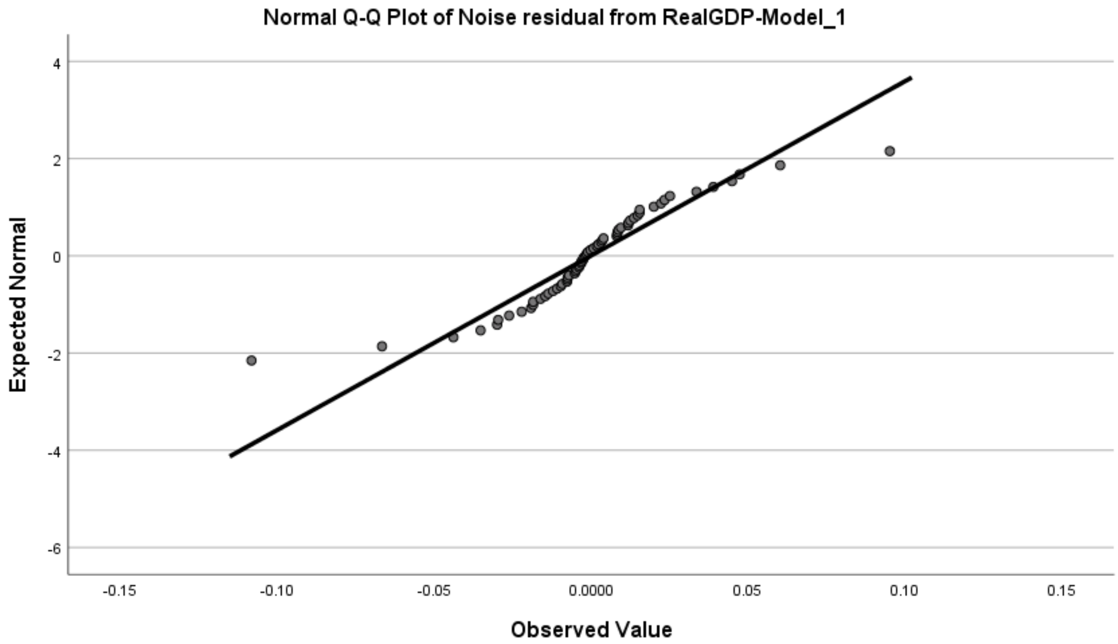
Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Noise residual from RealGDP-Model_1	.129	63	.010	.905	63	.000

a. Lilliefors Significance Correction

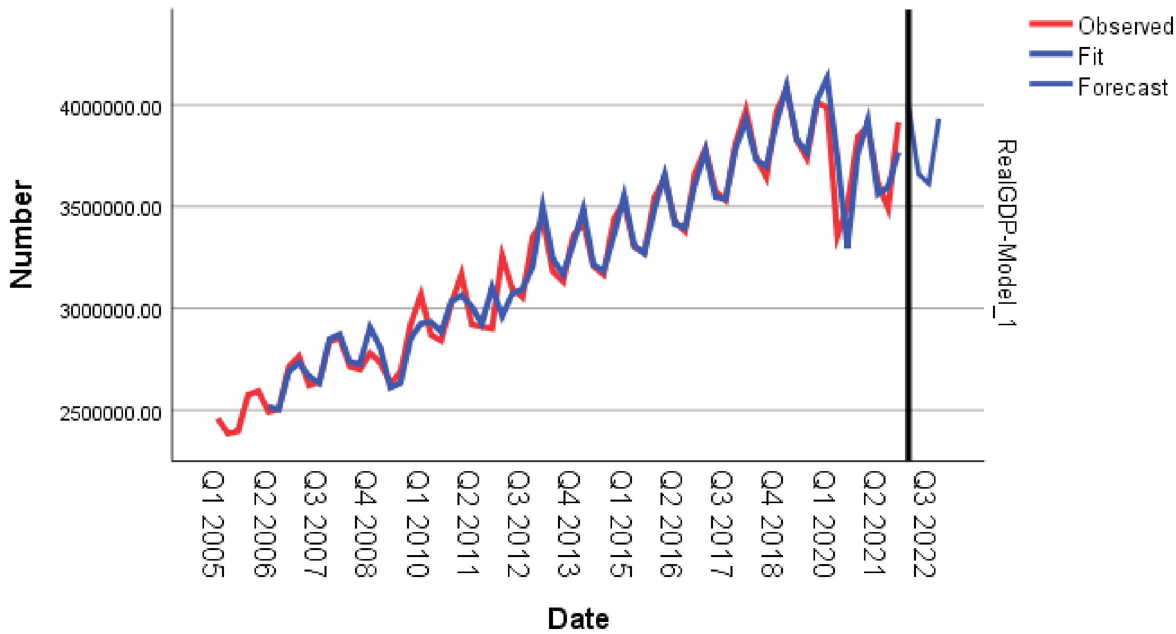
ตารางที่ 4 แสดงค่าสถิติ Kolmogorov-Smirnov และค่า P-value

เมื่อตรวจสอบ Normal Q-Q Plot ของส่วนตกค้างที่ได้จากตัวแบบ พบว่าส่วนใหญ่มีการเคลื่อนไหวไม่อยู่รอบเส้นตรงกลาง ดังรูปที่ 9



รูปที่ 9 แสดงความแปรปรวนของส่วนตกค้างที่ได้จากตัวแบบ

ดังนั้นนั่นคือ ตัวแบบ ARIMA(0,1,0)(0,1,1) ที่เลือกยังไม่มีความเหมาะสม จากความคลาดเคลื่อนสุ่มมีการแจกแจงแบบบไม่ปกติ ถึงแม้จะลองแก้ไขด้วยวิธี Take Natural Logarithm แล้วก็ตาม และเมื่อใช้ตัวแบบดังกล่าวในการพยากรณ์ Real GDP ตั้งแต่ไตรมาสที่ 1 ค.ศ.2005 ถึงไตรมาสที่ 4 ค.ศ.2022 ได้ค่าพยากรณ์ดังรูปที่ 10



รูปที่ 10 แสดงการเปรียบเทียบค่าอนุกรมเวลาและค่าพยากรณ์จากตัวแบบ ARIMA(0,1,0)(0,1,1)

2. วิธีพยากรณ์ของวินเตอร์

ผลลัพธ์จากโปรแกรมสำเร็จรูป SPSS การพยากรณ์ด้วยวิธีเทคนิคการปรับให้เรียบแบบวินเตอร์ ค่า α และ γ และ δ ที่ให้ค่า MSE ต่ำสุด คือ $\alpha = 0.771$ $\gamma = 0.001$ *และ* $\delta = 0.499$ ดังตารางที่ 5

Time Series Modeler

Model Description

			Model Type
Model ID	RealGDP	Model_1	Winters' Multiplicative

Exponential Smoothing Model Parameters

Model			Estimate	SE	t	Sig.
RealGDP-Model_1	No Transformation	Alpha (Level)	.771	.123	6.251	.000
		Gamma (Trend)	.001	.064	.016	.988
		Delta (Season)	.499	.344	1.452	.151

ตารางที่ 5 ค่าพารามิเตอร์ที่ประมาณได้ด้วยวิธีพยากรณ์แบบวินเตอร์

ตัวแบบพยากรณ์ดังกล่าวให้ค่าความผิดพลาดของการพยากรณ์ MAPE = 1.607 ดังตาราง Model Summary

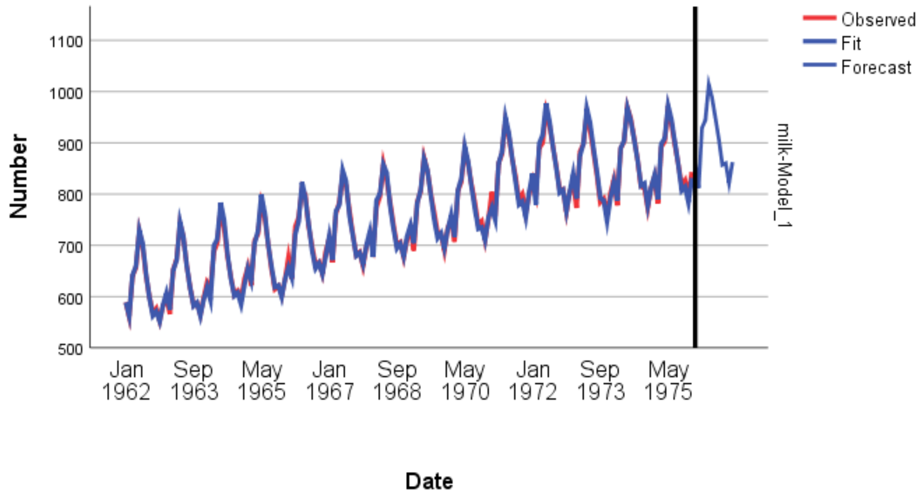
Model Summary

Model Fit

Fit Statistic	Mean	SE	Minimum	Maximum	Percentile 5	10	25	50	75	90	95
Stationary R-squared	.447	.	.447	.447	.447	.447	.447	.447	.447	.447	.447
R-squared	.969	.	.969	.969	.969	.969	.969	.969	.969	.969	.969
RMSE	84409.166	.	84409.166	84409.166	84409.166	84409.166	84409.166	84409.166	84409.166	84409.166	84409.166
MAPE	1.607	.	1.607	1.607	1.607	1.607	1.607	1.607	1.607	1.607	1.607
MaxAPE	12.158	.	12.158	12.158	12.158	12.158	12.158	12.158	12.158	12.158	12.158
MAE	51606.304	.	51606.304	51606.304	51606.304	51606.304	51606.304	51606.304	51606.304	51606.304	51606.304
MaxAE	407977.225	.	407977.225	407977.225	407977.225	407977.225	407977.225	407977.225	407977.225	407977.225	407977.225
Normalized BIC	22.873	.	22.873	22.873	22.873	22.873	22.873	22.873	22.873	22.873	22.873

ตารางที่ 6 ค่าความผิดพลาดของการพยากรณ์ด้วยวิธีพยากรณ์แบบวินเตอร์

เมื่อใช้วิธีพยากรณ์แบบวินเตอร์ดังกล่าวในการพยากรณ์ Real GDP ตั้งแต่ไตรมาสที่ 1 ค.ศ.2005 ถึงไตรมาสที่ 4 ค.ศ.2022 ได้ค่าพยากรณ์ดังรูปที่ 11



รูปที่ 11 แสดงการเปรียบเทียบค่าอนุกรมเวลาและค่าพยากรณ์ที่ได้จากวิธีพยากรณ์แบบวินเตอร์

การศึกษาตัวแบบที่เหมาะสมในการพยากรณ์ข้อมูลอนุกรเวลาชุดที่ 2 จะพิจารณาเทคนิคการพยากรณ์ 2 วิธี คือ

1. **วิธีพยากรณ์ของบอกซ์-เจนกินส์** เป็นวิธีที่ใช้สำหรับเลือกรูปแบบที่เหมาะสมให้กับอนุกรมเวลา โดยพิจารณาจากสหสัมพันธ์ระหว่าง Y ที่คาบเวลา t (Y_t) และที่คาบเวลาที่ผ่านมา (Y_{t-1}, Y_{t-2}, …) เมื่อได้ตัวแบบที่เหมาะสมแล้ว จะใช้ตัวแบบนี้ในการพยากรณ์ Y_{t+1}, Y_{t+2},… ในอนาคต อนุกรมเวลาที่จะกำหนดรูปแบบโดยวิธีบอกซ์-เจนกินส์ จะต้องเป็นอนุกรมเวลาที่อยู่ในสภาวะนิ่ง (Stationary data series) เท่านั้น ซึ่งหมายถึง คงที่ในค่าเฉลี่ย ความแปรปรวน และความแปรปรวนร่วม ไม่แปรผันตามเวลา

ดังนั้นขั้นตอนของวิธีพยากรณ์ของบอกซ์-เจนกินส์ที่สำคัญประกอบด้วย 5 ขั้นตอนได้แก่

(6) ตรวจสอบสภาวะนิ่งโดยพิจารณาจากกราฟของอนุกรมเวลา และจากกราฟฟังก์ชันอัตตสหสัมพันธ์ (Autocorrelation Function: ACF) แทนด้วย r_k

(7) ถ้าตรวจสอบแล้วพบว่าอนุกรมเวลาไม่อยู่ในสภาวะคงที่ในค่าเฉลี่ย จะทำการแปลงเป็นอนุกรมเวลาชุดใหม่ $\{W_t\}$ ที่มีลักษณะคงที่ในค่าเฉลี่ย โดยการหาผลต่างของอนุกรมเวลา ถ้ามีฤดูกาลจะแปลงอนุกรมเวลาด้วยการหาผลต่างฤดูกาล เป็นต้น

(8) กำหนดตัวแบบที่คาดว่าจะเหมาะสมให้กับอนุกรมเวลาโดยพิจารณาจากกราฟ ACF และ PACF (Partial Autocorrelation Function: PACF) แทนด้วย r_{kk}

(9) ประมาณค่าพารามิเตอร์ในสมการที่เลือกไว้ด้วยวิธีกำลังสองน้อยที่สุด (Ordinary least Squares)

(10)ตรวจสอบความเหมาะสมของตัวแบบจากกราฟ ACF และPACF ของส่วนตกค้าง (Residuals: $e_t = Y_t - \hat{Y}_t$)

ด้วยวิธีบอกซ์-เจนกินส์จะได้ตัวแบบอนุกรมเวลาที่เรียกว่า ตัวแบบ ARIMA (Autoregressive Integrated Moving Average) และตัวแบบที่ศึกษาในครั้งนี้เป็นตัวแบบ ARIMA(p,d,q)(P,D,Q) มีรูปแบบดังนี้

$$\phi_p(B)\Phi_P(B^S)(1-B)^d(1\text{-}B^S)^D Y_t = \delta + \theta_q(B)\Theta_Q(B^S)\varepsilon_t$$

โดยที่

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$$

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$$

$$\Phi_P(B^S) = 1 - \Phi_S B^S - \Phi_{2S} B^{2S} - \dots - \Phi_{PS} B^{PS}$$

$$\Theta_Q(B^S) = 1 - \Theta_S B^S - \Theta_{2S} B^{2S} - \dots - \Theta_{QS} B^{QS}$$

$\phi_1, \dots, \phi_p$ คือ สัมประสิทธิ์การถดถอย (Autoregressive Coefficients)

$\theta_1, \dots, \theta_q$ คือ สัมประสิทธิ์ค่าเฉลี่ยเคลื่อนที่ (Moving Average Coefficients)

δ คือ ค่าคงที่

B คือ ตัวดำเนินการถอยหลังเวลา (Backward Shift Operator) นั่นคือ $B^m Y_t = Y_{t-m}$

d คือ จำนวนครั้งของการทำผลต่างเพื่อให้อนุกรมเวลา $\{Y_t\}$ เป็นอนุกรมเวลาที่อยู่ในสภาวะคงที่ในค่าเฉลี่ย

D คือ จำนวนครั้งของการทำผลต่างฤดูกาล

p คือ อันดับของตัวแบบการถดถอย

P คือ อันดับของตัวแบบการถดถอยฤดูกาล

q คือ อันดับของตัวแบบค่าเฉลี่ยเคลื่อนที่

Q คือ อันดับของตัวแบบค่าเฉลี่ยเคลื่อนที่ฤดูกาล

S คือ จำนวนฤดูกาล

a_t คือ ตัวแปรสุ่มอิสระและมีการแจกแจงปกติ ซึ่งมีค่าเฉลี่ยเท่ากับศูนย์และความแปรปรวนคงที่ ให้เท่ากับ σ_ε^2 เรียก a_t ว่าความคลาดเคลื่อนสุ่ม หรือ กระตุกสุ่ม (Random Shocks)

2. **เทคนิคการปรับให้เรียบ** เป็นวิธีการพยากรณ์ข้อมูลโดยการขจัดความผันแปรที่ผิดปกติออก ทำให้เห็นองค์ประกอบอื่นของอนุกรมเวลา เพื่อจะสามารถพยากรณ์ค่าอนุกรมเวลาในอนาคตได้ เทคนิคการปรับให้เรียบมีหลายวิธี ขึ้นกับลักษณะของข้อมูลอนุกรมเวลา

ถ้าอนุกรมเวลามีทั้งแนวโน้มและฤดูกาล จะใช้**วิธีการทำให้เรียบแบบเอ็กซ์โพเนนเชียลซ้ำ 2 ครั้ง แบบ Brown** (Brown's one parameter linear model หรือ linear exponential smoothing) โดยมีตัวแบบดังนี้

ตัวแบบ $Y_t = \beta_0 + \beta_1 X + \varepsilon_t$

สมการพยากรณ์ $\hat{Y}_{t+m} = a_t + b_t m$

โดย $a_t = 2E'_t - E''_t$ และ $b_t = \dfrac{\alpha}{(1-\alpha)}(E'_t - E''_t)$

โดยที่ $E'_t = \alpha Y_t + (1-\alpha)E'_{t-1}$ และ $E''_t = \alpha E'_t + (1-\alpha)E''_{t-1}$

ค่าเริ่มต้นนิยมกำหนด $E'_1 = E''_1 = Y_1$

E'_t เป็นค่าที่ได้จากการทำให้เรียบแบบ Exponential อย่างง่ายครั้งที่ 1

E''_t เป็นค่าที่ได้จากการทำให้เรียบแบบ Exponential อย่างง่ายครั้งที่ 2

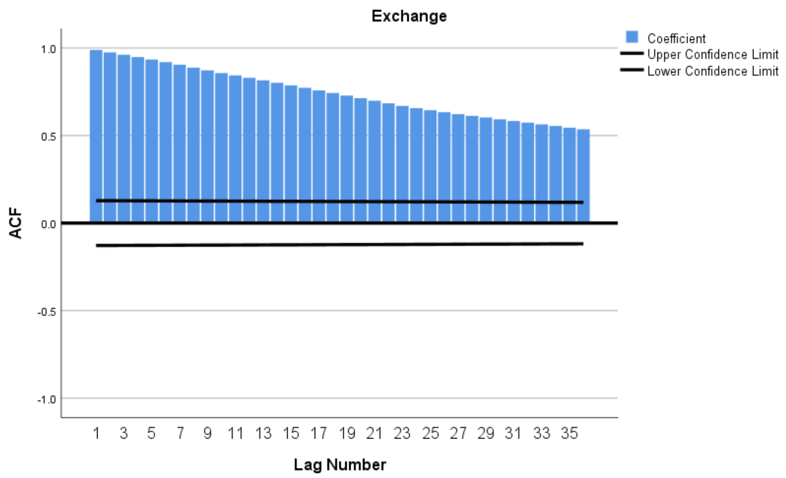
ข้อมูลที่ศึกษาชุดที่ 2 คือ อนุกรมเวลาของ Exchange หรืออัตราแลกเปลี่ยนเงินตราต่างประเทศ จำนวน 241 วัน และทำการวิเคราะห์โดยใช้โปรแกรมสำเร็จรูป SPSS

1. วิธีบอกซ์-เจนกินส์ ผลการศึกษาในแต่ละขั้นตอนเป็นดังนี้

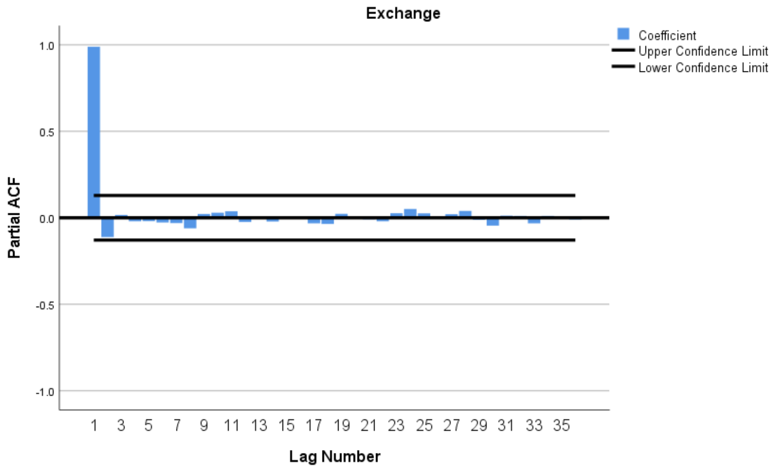
(1) ตรวจสอบสภาวะนิ่ง จาก Sequence Chart ของอนุกรมเวลา พบว่าการเคลื่อนไหวของอนุกรมเวลาของ Exchange มีความไม่คงที่ในค่าเฉลี่ย ดังรูปที่ 1

รูปที่ 1 การเคลื่อนไหวของอนุกรมเวลา

(2) เมื่อพิจารณาจากกราฟ ACF พบว่าการเคลื่อนไหวของ r_k มีลักษณะลดลงช้าๆ แสดงว่าอนุกรมเวลาอยู่ในสภาวะไม่คงที่ (Non Stationary) ในค่าเฉลี่ย ดังรูปที่ 2

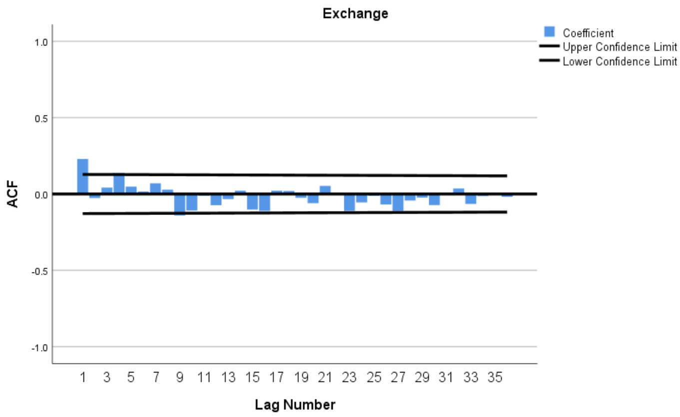


รูปที่ 2 กราฟ ACF ของอนุกรมเวลา

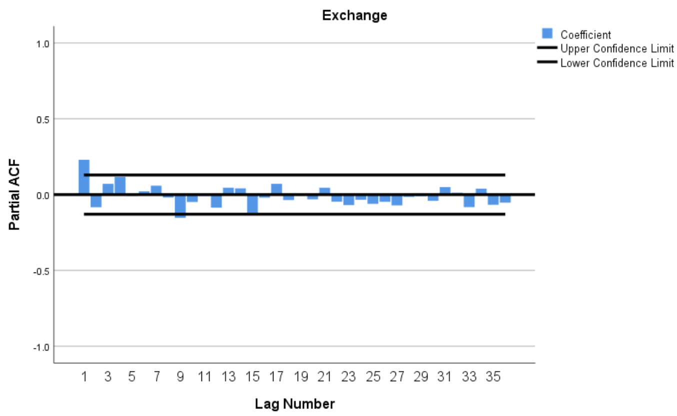


รูปที่ 3 กราฟ PACF ของอนุกรมเวลา

จากกราฟ ACF r_k หลังจากหาผลต่างอนุกรมเวลา 1 ครั้ง ได้กราฟ ACF และ PACF ดังรูปที่ 4 และ รูปที่ 5 ตามลำดับ



รูปที่ 4 กราฟ ACF ของอนุกรมเวลาเมื่อหาผลต่างอนุกรมเวลา 1 ครั้ง



รูปที่ 5 กราฟ PACF ของอนุกรมเวลาเมื่อหาผลต่างอนุกรมเวลา 1 ครั้ง

(3) กำหนดตัวแบบที่เหมาะสม โดยพิจารณาจากกราฟ ACF และ PACF ของอนุกรมเวลาชุดใหม่ พบว่า r_k มีค่าสูงที่ k = 1 ค่าเดียว และมีค่าเข้าใกล้ศูนย์ ที่ k = 2, 3, 4, … ส่วน r_{kk} มีค่าสูงที่ k = 1 และมีค่าลดลงเร็วๆ ที่ k = 2, 3, 4, … แสดงถึง ตัวแบบฤดูกาล AR(1) และ MA(1)
จากการทดสอบแต่ละตัวแบบแล้ว จะได้ตัวแบบที่เหมาะสมที่สุดคือ ARIMA(1,1,0)

(4) จากตัวแบบที่กำหนด ARIMA(1,1,0) ของอนุกรมเวลา W_t ประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุด ได้ค่าพารามิเตอร์ดังผลลัพธ์ข้างล่างนี้

Time Series Modeler

Model Description

			Model Type
Model ID	Exchange	Model_1	ARIMA(1,1,0)

ARIMA Model Parameters

					Estimate	SE	t	Sig.
Exchange-Model_1	Exchange	No Transformation	Constant		.014	.009	1.631	.104
			AR	Lag 1	.231	.063	3.639	.000
			Difference		1			

ตารางที่ 1 ค่าพารามิเตอร์ของตัวแบบ ARIMA(1,1,0) ที่ประมาณได้

ตัวแบบพยากรณ์ดังกล่าวให้ค่าความผิดพลาดของการพยากรณ์ MAPE = 0.241 ดังตาราง Model Summary

Model Summary

Model Fit

Fit Statistic	Mean	SE	Minimum	Maximum	Percentile 5	10	25	50	75	90	95
Stationary R-squared	.053	.	.053	.053	.053	.053	.053	.053	.053	.053	.053
R-squared	.993	.	.993	.993	.993	.993	.993	.993	.993	.993	.993
RMSE	.104	.	.104	.104	.104	.104	.104	.104	.104	.104	.104
MAPE	.241	.	.241	.241	.241	.241	.241	.241	.241	.241	.241
MaxAPE	1.008	.	1.008	1.008	1.008	1.008	1.008	1.008	1.008	1.008	1.008
MAE	.078	.	.078	.078	.078	.078	.078	.078	.078	.078	.078
MaxAE	.315	.	.315	.315	.315	.315	.315	.315	.315	.315	.315
Normalized BIC	-4.473	.	-4.473	-4.473	-4.473	-4.473	-4.473	-4.473	-4.473	-4.473	-4.473

ตารางที่ 2 ค่าความผิดพลาดของการพยากรณ์ของตัวแบบ ARIMA(1,1,0)

(5) ตรวจสอบความเหมาะสมของตัวแบบ โดยตรวจสอบว่าค่าความคลาดเคลื่อนจากตัวแบบมีค่าเฉลี่ยเท่ากับศูนย์ มีการแจกแจงแบบปกติ และค่าความคลาดเคลื่อนแต่ละค่าไม่มีอัตตสัมพันธ์กัน (เป็นอิสระกัน)

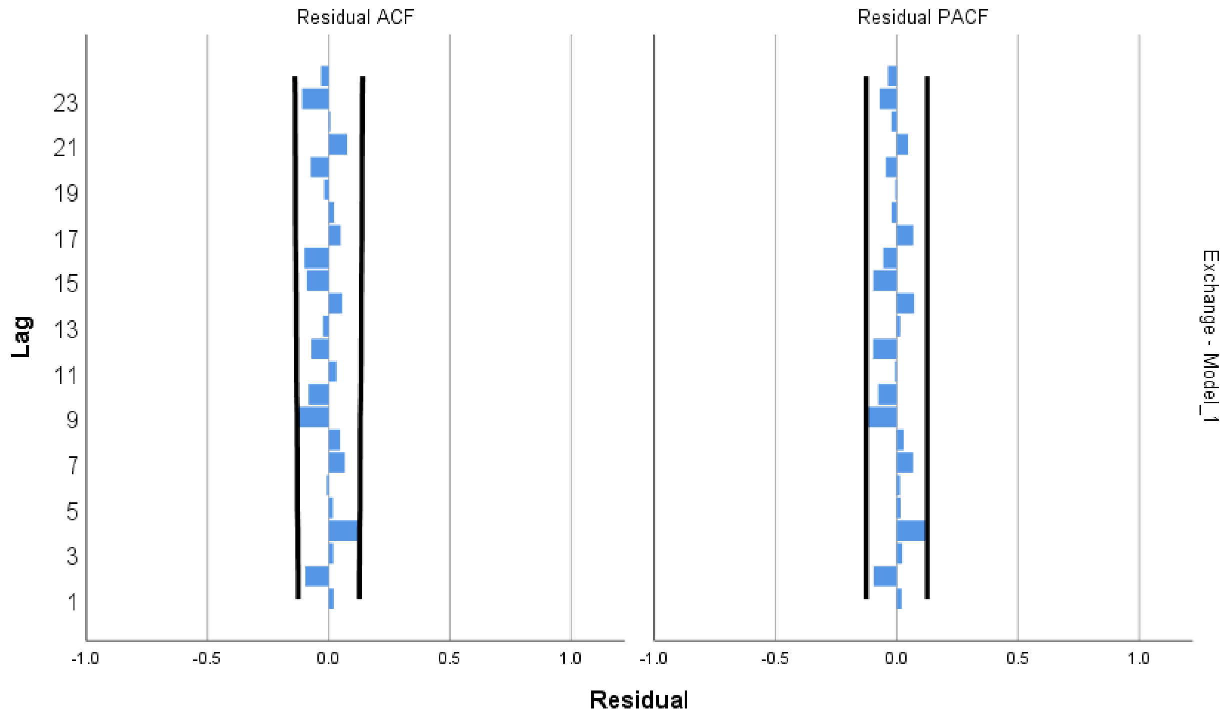
จากตารางที่ 3 Model Statistics ค่า P-value ของการทดสอบ Ljung-Box มีค่า = 0.153 แสดงว่าความคลาดเคลื่อนจากตัวแบบ ไม่มีอัตตสหสัมพันธ์กัน (เป็นอิสระกัน ไม่มีปัญหา Autocorrelation)

Model Statistics

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
Exchange-Model_1	0	.053	22.894	17	.153	0

ตารางที่ 3 แสดงค่าสถิติ Ljung-Box และค่า P-value

เมื่อตรวจสอบจากกราฟ ACF และ PACF -ของส่วนตกค้าง (Residual) พบว่า r_k และ r_{kk} ของส่วนตกค้างตกอยู่ในขอบเขตความเชื่อมั่นร้อยละ 99 ทุกค่า lag ดังรูปที่ 6 แสดงว่าค่าความคลาดเคลื่อนไม่เกิดปัญหา Autocorrelation



รูปที่ 6 กราฟ ACF และ PACF ของส่วนตกค้าง

และเมื่อตรวจสอบค่า P-value ของค่าสถิติทดสอบ Kolmogolov-Sirminov มีค่าเท่ากับ 0.200 แสดงว่าค่าความคลาดเคลื่อนสุ่มมีการแจกแจงแบบปกติ ดังตารางที่ 4

Tests of Normality

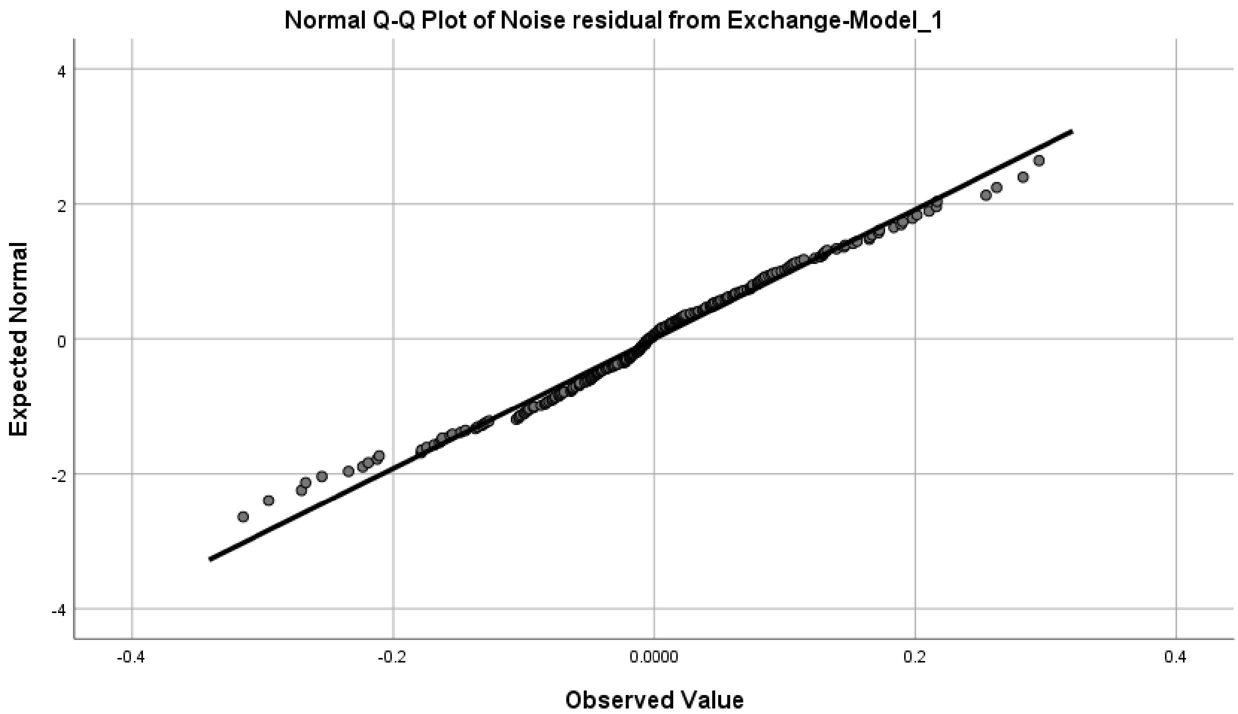
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Noise residual from Exchange-Model_1	.053	240	.200*	.988	240	.035

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

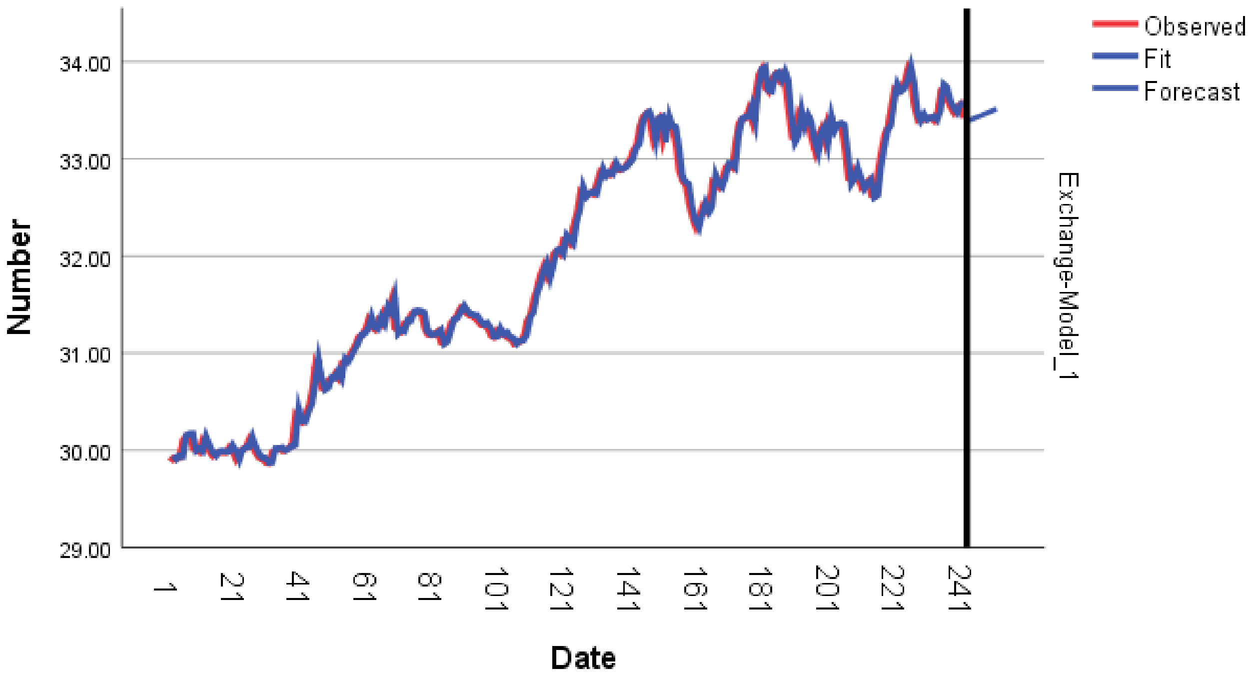
ตารางที่ 4 แสดงค่าสถิติ Kolmogorov-Smirnov และค่า P-value

เมื่อตรวจสอบ Normal Q-Q Plot ของส่วนตกค้างที่ได้จากตัวแบบ พบว่าส่วนใหญ่มีการเคลื่อนไหวอยู่รอบเส้นตรงกลาง ดังรูปที่ 7



รูปที่ 7 แสดงการเคลื่อนไหวความแปรปรวนของส่วนตกค้างที่ได้จากตัวแบบ

นั่นคือ ตัวแบบ ARIMA(1,1,0) ที่เลือกมีความเหมาะสมแล้ว และเมื่อใช้ตัวแบบดังกล่าวในการพยากรณ์ Exchange ตั้งแต่วันที่ 1 ถึง 251 ได้ค่าพยากรณ์ดังรูปที่ 8



รูปที่ 8 แสดงการเปรียบเทียบค่าอนุกรมเวลาและค่าพยากรณ์จากตัวแบบ ARIMA(1,1,0)

2. วิธีพยากรณ์ของบราวน์

ผลลัพธ์จากโปรแกรมสำเร็จรูป SPSS การพยากรณ์ด้วยวิธีเทคนิคการปรับให้เรียบแบบเอ็กซ์โพเนนเชียลซ้ำสองครั้งแบบบราวน์ค่า α ที่ให้ค่า MSE ต่ำสุด คือ $\alpha = 0.592$ ดังตารางที่ 5

Time Series Modeler

Model Description

			Model Type
Model ID	Exchange	Model_1	Brown

Exponential Smoothing Model Parameters

Model			Estimate	SE	t	Sig.
Exchange-Model_1	No Transformation	Alpha (Level and Trend)	.592	.029	20.084	.000

ตารางที่ 5 ค่าพารามิเตอร์ที่ประมาณได้ด้วยวิธีพยากรณ์แบบเอ็กซ์โพเนนเชียลซ้ำสองครั้งแบบบราวน์

ตัวแบบพยากรณ์ดังกล่าวให้ค่าความผิดพลาดของการพยากรณ์ MAPE = 0.268 ดังตารางที่ 6

Model Summary

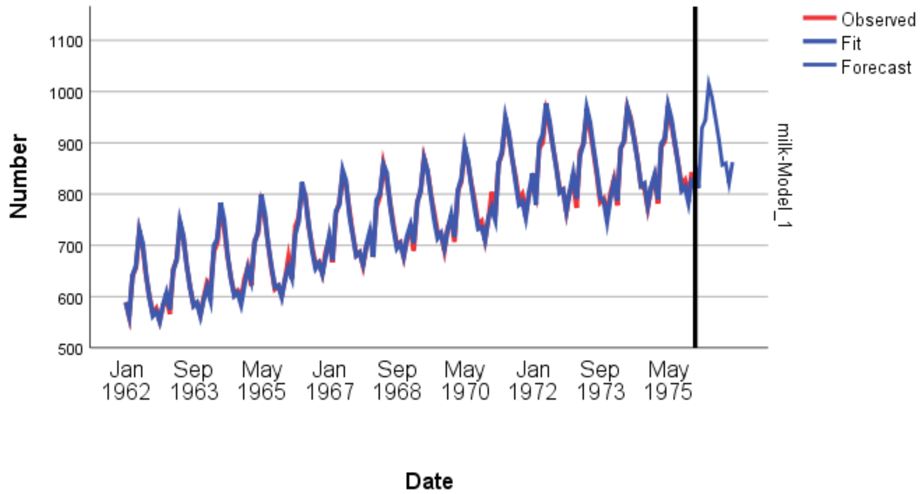
Model Summary

Model Fit

Fit Statistic	Mean	SE	Minimum	Maximum	Percentile 5	10	25	50	75	90	95
Stationary R-squared	.240	.	.240	.240	.240	.240	.240	.240	.240	.240	.240
R-squared	.992	.	.992	.992	.992	.992	.992	.992	.992	.992	.992
RMSE	.115	.	.115	.115	.115	.115	.115	.115	.115	.115	.115
MAPE	.268	.	.268	.268	.268	.268	.268	.268	.268	.268	.268
MaxAPE	1.107	.	1.107	1.107	1.107	1.107	1.107	1.107	1.107	1.107	1.107
MAE	.087	.	.087	.087	.087	.087	.087	.087	.087	.087	.087
MaxAE	.370	.	.370	.370	.370	.370	.370	.370	.370	.370	.370
Normalized BIC	-4.297	.	-4.297	-4.297	-4.297	-4.297	-4.297	-4.297	-4.297	-4.297	-4.297

ตารางที่ 6 ค่าความผิดพลาดของการพยากรณ์ด้วยวิธีพยากรณ์แบบเอ็กซ์โพเนนเชียลซ้ำสองครั้งแบบบราวน์

เมื่อใช้วิธีพยากรณ์แบบเอ็กซ์โพเนนเชียลซ้ำสองครั้งแบบบราวน์ดังกล่าวในการพยากรณ์ Exchange ตั้งแต่วันที่ 1 ถึงวันที่ 251 ได้ค่าพยากรณ์ดังรูปที่ 9



รูปที่ 9 แสดงการเปรียบเทียบค่าอนุกรมเวลาและค่าพยากรณ์ที่ได้จากวิธีพยากรณ์แบบเอ็กซ์โพเนนเชียลซ้ำสองครั้งแบบบราวน์