



การพยากรณ์สัญญาณซื้อ-ขายล่วงหน้า จากการใช้อินดิเคเตอร์ Bollinger-Bands  
ร่วมกับตัวแบบ ARIMAX และตัวแบบการเรียนรู้ของเครื่อง

Forecasting buy-sell futures signals by using a Bollinger Bands indicator  
in combination with an ARIMAX model and a machine learning model.

โดย

นายชนวีร์ สุขเยาว์

นายเกริกพล รัตนภูมิ

รายงานนี้เป็นส่วนหนึ่งของวิชา ส.495 โครงการพิเศษ 2

ตามหลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาสถิติ

คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

ปีการศึกษา 2565

หัวข้อโครงการพิเศษ	การพยากรณ์สัญญาณซื้อ-ขายล่วงหน้า จากการใช้อินดิเคเตอร์ Bollinger-Bands ร่วมกับตัวแบบ ARIMAX และตัวแบบการเรียนรู้ของเครื่อง
คณะผู้จัดทำ	นายชนวีร์ สุขเยาว์ นายเกริกพล รัตนภูมิ
ชื่อปริญญา	วิทยาศาสตร์บัณฑิต
หลักสูตร/สาขา	หลักสูตรวิทยาศาสตร์บัณฑิต สาขาวิชาสถิติ
คณะ/มหาวิทยาลัย	คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์
อาจารย์ที่ปรึกษางานวิจัย	รองศาสตราจารย์ ดร.วิชัย วิทยาเกียรติเลิศ
ปีการศึกษา	2565

### บทคัดย่อ

งานวิจัยนี้ศึกษาตัวแบบพยากรณ์อนุกรมเวลาราคาดัชนีSET50 โดยแบ่งข้อมูลออกเป็นสองส่วน คือ ส่วนที่เป็นเชิงเส้นตรงและส่วนที่ไม่เป็นเชิงเส้นตรง ในงานวิจัยนี้ใช้ตัวแบบพยากรณ์อนุกรมเวลา ARIMAX พยากรณ์ข้อมูลในส่วนที่เป็นเชิงเส้นตรงและใช้ตัวแบบ ANN และ SVR พยากรณ์ข้อมูลในส่วนที่ไม่เป็นเชิงเส้นตรง การสร้างตัวแบบผสมนี้จะเรียกว่าตัวแบบ ARIMAX-ANN และ ARIMAX-SVR และได้ปรับปรุงตัวแบบแบบผสมโดยเชื่อมตัวแบบ ARIMAX-ANN และ ARIMAX-SVR กลายเป็นตัวแบบที่เรียกว่า Combined (ARIMA-ANN-SVM) ใช้เป็นตัวแบบพยากรณ์ราคาดัชนี SET50 โดยประเมินผลความแม่นยำของแต่ละตัวแบบด้วย RMSE, MSE, MAE และ MAPE ผลการทดลองพบว่า ตัวแบบ Combined (ARIMA-ANN-SVM) แม่นยำกว่ามากที่สุดเมื่อเทียบกับตัวแบบชนิดอื่นในงานวิจัย นอกจากนี้ ได้นำเสนอการนำเอาผลลัพธ์งานวิจัยไปประยุกต์ใช้ในการพยากรณ์สัญญาณซื้อ-ขายล่วงหน้า ร่วมกับเครื่องมือ Bollinger Band พบว่าให้ผลลัพธ์ที่เป็นสัญญาณซื้อ-ขายล่วงหน้าที่ใกล้เคียงสัญญาณซื้อ-ขายที่เกิดขึ้น ณ เวลาจริง

**คำสำคัญ** stock forecasting, hybrid models, ARIMAX, artificial neural network, support vector regression, Bollinger bands

Title	Forecasting buy-sell futures signals by using a Bollinger Bands indicator in combination with an ARIMAX model and a machine learning model.
Faculty of providers	Chonnavee Sukyao Kerkpol Rattanapoom
Degree	Bachelor of Science
Bachelor/Department	Bachelor of Science Program Department of Statistics
Faculty/University	Faculty of Science and Technology Thammasat University
Advisor	Assoc.Prof. Wichai Witayakiattilerd
Academic year	2023

### ABSTRACT

This research investigates the forecasting models for the SET50 index prices, dividing the data into linear and non-linear components. The ARIMAX time-series forecasting model is employed for the linear part, while ANN and SVR models are used for the non-linear part. These hybrid models are known as ARIMAX-ANN and ARIMAX-SVR. The study further refines the hybrid models by integrating ARIMAX-ANN and ARIMAX-SVR into a combined model, termed Combined (ARIMA-ANN-SVM), which is then applied to forecast the SET50 index prices. The accuracy of each model is assessed using RMSE, MSE, MAE, and MAPE. The experimental results indicate that the Combined (ARIMA-ANN-SVM) model provides the highest degree of accuracy compared to the other models in the study. The research outcomes are also applied to forecast buy-sell signals in conjunction with the Bollinger Band tool, demonstrating results that closely align with real-time buy-sell signals.

**Keyword** stock forecasting, hybrid models, ARIMAX, artificial neural network, support vector regression, Bollinger bands

### กิตติกรรมประกาศ

ผู้วิจัยขอกราบขอบพระคุณ รองศาสตราจารย์ ดร.วิชัย วิทยาเกียรติเลิศ อาจารย์ที่ปรึกษา เป็นอย่างสูงที่กรุณาให้คำแนะนำช่วยเหลือและร่วมติดตามการทดลองอย่างใกล้ชิดในทุกขั้นตอนทำให้โครงการพิเศษสำเร็จลุล่วงไปได้ด้วยดีและขอขอบพระคุณกรรมการสอบโครงการพิเศษซึ่งประกอบด้วย รองศาสตราจารย์ ดร.วิชัย วิทยาเกียรติเลิศ ผู้ช่วยศาสตราจารย์นิฉา แก้วหาวงษ์ และ ผู้ช่วยศาสตราจารย์.ดร.แสงดาว วงศ์สาย ที่ได้สละเวลา อันมีค่าเพื่อเป็นเพื่อกรรมการสอบโครงการพิเศษจนทำให้มีความสมบูรณ์ในที่สุด ผู้วิจัยขอกราบขอบพระคุณ คุณพ่อ คุณแม่ ที่เปรียบเสมือนทุกอย่างของชีวิต ที่ให้กำลังใจ ความรัก ให้คำปรึกษา และช่วยเหลือในทุกๆด้าน จนทำให้ผู้ทำวิจัยสามารถทำการวิจัยจนสำเร็จ ถ้าวิทยานิพนธ์เล่มนี้ผิดพลาดประการใดผู้จัดทำต้องขออภัยมา ณ ที่นี้ด้วย

นายชนวีร์ สุขเยาว์

นายนายเกริกพล รัตนภูมิ

## สารบัญ

	หน้า
บทคัดย่อ	ก
กิตติกรรมประกาศ	ค
สารบัญ	ง
สารบัญตาราง	ฉ
สารบัญรูป	ช
บทที่ 1 บทนำ	1
1.1 ที่มาของปัญหาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของการวิจัย	3
1.3 ประโยชน์ที่คาดว่าจะได้รับ	3
1.4 ขอบเขตการศึกษา	3
1.5 นิยามศัพท์	5
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	9
2.1 ทฤษฎีที่เกี่ยวข้องกับการพยากรณ์	9
2.2 ตัวแบบ Autoregressive Integrated Moving Average (ARIMA)	13
2.3 ตัวแบบ (Autoregressive Integrated Moving Average with Exogenous Variable (ARIMAX)	19
2.4 โครงข่ายประสาทเทียม (Artificial Neural Network: ANN)	20
2.5 ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM)	32
2.6 ซัพพอร์ตเวกเตอร์รีเกรสชัน (Support Vector Regression: SVR)	34

## สารบัญ(ต่อ)

	หน้า
2.7 ขั้นตอนการสร้างตัวแบบผสมที่มีกระบวนการทำต่อเนื่องกัน (Traditional hybrid)	38
2.8 ตัวแบบที่ทำการพยากรณ์แยกส่วนกัน (Hybrid combined)	39
2.9 แถบโบลลิงเจอร์ (Bollinger Bands : BB)	40
2.10 การวิเคราะห์ปัจจัย (Factor Analysis)	41
2.11 วรรณกรรมและงานวิจัยที่เกี่ยวข้อง	45
<b>บทที่ 3 วิธีการดำเนินงานวิจัย</b>	<b>49</b>
3.1 กลุ่มตัวอย่างและการเก็บรวบรวมข้อมูล	49
3.2 การสร้างตัวแบบและวิเคราะห์ข้อมูล	51
3.3 การคำนวณหาความคลาดเคลื่อน	56
3.4 การพยากรณ์สัญญาณการซื้อขายล่วงหน้าและการวัดประสิทธิภาพ	57
3.5 แผนผังการดำเนินงานวิจัย	58
<b>บทที่ 4 ผลการวิจัยและอภิปรายผล</b>	<b>60</b>
4.1 กลุ่มตัวอย่าง และการเก็บรวบรวมข้อมูล	60
4.2 การสร้างตัวแบบ และวิเคราะห์ข้อมูล	66
4.3 การคำนวณหาค่าความคลาดเคลื่อน และการคัดเลือกตัวแบบ	82
4.4 การพยากรณ์สัญญาณการซื้อขายล่วงหน้าและการวัดประสิทธิภาพ	84
<b>บทที่ 5 สรุปผลการวิจัยและข้อเสนอ</b>	<b>86</b>
5.1 การพยากรณ์สัญญาณการซื้อขายล่วงหน้าและการวัดประสิทธิภาพ	86
5.2 ข้อเสนอแนะและแนวทางในการทำวิจัยในขั้นถัดไป	87

สารบัญ(ต่อ)

บรรณานุกรม	89
ภาคผนวก	92
ภาคผนวก ก ผลลัพธ์การเปรียบเทียบตัวแบบ ARIMA ที่เหมาะสมที่สุดผ่านตัวชี้วัด Akaike Info Criterion (AIC) ที่ได้จาก Google Google Collaboratory run on Python version 3.9	93
ภาคผนวก ข ผลลัพธ์ตัวแบบ ARIMAX(p,d,q) ที่เหมาะสม ที่ได้จาก Google Collaboratory run on Python version 3.9	94

## สารบัญตาราง

	หน้า
ตารางที่ 2.1 ความสัมพันธ์ระหว่างค่า loading ที่มีนัยสำคัญทางสถิติที่ระดับ 0.05	
ต่อจำนวนตัวอย่าง	45
ตารางที่ 3.1 ตารางชี้แจงแหล่งที่มาของข้อมูลในการนำมาใช้พยากรณ์	50
ตารางที่ 4.1 ผลลัพธ์จากการทดสอบ Augmented Dickey-Fuller test ของข้อมูล SET50	
ก่อนการแก้ไขข้อมูล หรือ ณ ผลต่างลำดับที่ 0 (At Level)	60
ตารางที่ 4.2 ผลลัพธ์จากการทดสอบ Augmented Dickey-Fuller test ของข้อมูล SET50	
หลังการแก้ไขข้อมูล หรือ ณ ผลต่างลำดับที่ 1 (1 <sup>st</sup> Difference)	62
ตารางที่ 4.3 ผลลัพธ์จากการทดสอบ KMO and Bartlett's	64
ตารางที่ 4.4 เปรียบเทียบตัวแบบ ARIMA ที่เหมาะสมที่สุดผ่านตัวชี้วัด Akaike Info Criterion (AIC)	67
ตารางที่ 4.5 แสดงผลลัพธ์ของตัวแบบ <i>ARIMAX(2,0,3)</i>	69
ตารางที่ 4.6 แสดงผลลัพธ์ของตัวแบบ <i>ARIMAX(2,0,2)</i>	70
ตารางที่ 4.7 ผลการทดสอบ Ljung and Box เพื่อประเมิน White Noise ในตัวแบบ <i>ARIMAX</i>	
ที่นำมาพยากรณ์	71
ตารางที่ 4.8 ความคลาดเคลื่อนในการพยากรณ์ตัวแบบ ANN ในแต่ละพารามิเตอร์	74
ตารางที่ 4.9 แสดงผลลัพธ์ของตัวแบบ SVR	77
ตารางที่ 4.10 ความคลาดเคลื่อนในการพยากรณ์สำหรับตัวแบบผสมระหว่าง <i>ARIMAX</i> และ ANN	79
ตารางที่ 4.11 ความคลาดเคลื่อนในการพยากรณ์สำหรับตัวแบบผสมระหว่าง <i>ARIMAX</i> และ SVR	80
ตารางที่ 4.13 ความคลาดเคลื่อนในการพยากรณ์สำหรับตัวแบบรวม <i>ARIMAX-ANN-SVM</i>	81
ตารางที่ 4.14 แสดงค่าพารามิเตอร์ของตัวแบบ,ค่า RMSE,ค่า MSE,ค่า MAE และ ค่า MAPE	
ที่เหมาะสมกับข้อมูลดัชนี SET50	82



## สารบัญรูป

	หน้า
รูปที่ 2.1 กราฟแสดงลักษณะของแนวโน้มลดลง	10
รูปที่ 2.2 กราฟแสดงลักษณะของแนวโน้มเพิ่มขึ้น	11
รูปที่ 2.3 กราฟแสดงลักษณะของฤดูกาล	11
รูปที่ 2.4 กราฟแสดงลักษณะของวัฏจักรที่	12
รูปที่ 2.6 ACF (Stationary)	16
รูปที่ 2.7 ACF (Non-Stationary)	16
รูปที่ 2.8 หลักการทำงานของตัวแบบ ANN	22
รูปที่ 2.9 โครงสร้างของ Feedforward network	24
รูปที่ 2.10 Linear Transfer Function	25
รูปที่ 2.11 Log-Sigmoid Transfer Function	25
รูปที่ 2.12 Tan-Sigmoid Transfer Function	26
รูปที่ 2.13 แผนผังตัวแปรอิสระและตัวแปรตามสำหรับตัวแบบ ANN ในการพยากรณ์ข้อมูลอนุกรมเวลา	29
รูปที่ 2.14 แผนผังของตัวแบบ Feedforward network ในการพยากรณ์ข้อมูลอนุกรมเวลา	30
รูปที่ 2.15 เส้นตรงแบ่งกลุ่มข้อมูลที่มี 2 ตัวแปร สำหรับตัวแบบ SVM	32
รูปที่ 2.16 เส้นตรงที่ใช้แบ่งกลุ่มข้อมูลที่เป็นไปได้ สำหรับตัวแบบ SVM	33
รูปที่ 2.17 เส้นตรงที่ใช้แบ่งกลุ่มที่ดีที่สุด	33
รูปที่ 2.18 รูปแบบการแปลงจุด เพื่อให้สามารถหาเส้นตรงมาแบ่ง	34
รูปที่ 2.19 ขั้นตอนการสร้างตัวแบบ Traditional hybrid	39
รูปที่ 2.20 ขั้นตอนการสร้างตัวแบบ Hybrid Combined	40

สารบัญรูป(ต่อ)

	หน้า
รูปที่ 4.1 ข้อมูลอนุกรมเวลา SET50 Index ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2563 ถึงวันที่ 18 เมษายน พ.ศ. 2566	38
รูปที่ 4.2 แสดงการวิเคราะห์สหสัมพันธ์ระหว่างดัชนีราคาหุ้นตลาดหลักทรัพย์ แห่งประเทศไทย (SET50)และตัวแปรต่างๆ	63
รูปที่ 4.3 การวิเคราะห์สหสัมพันธ์ระหว่างดัชนีราคาหุ้นตลาดหลักทรัพย์แห่งประเทศไทย (SET50) และ ตัวแปรที่ผ่านการทำ PCA	65
รูปที่ 4.4 ผล Correlogram ของดัชนี SET50 ณ ผลต่างลำดับที่ 0 (At Level)	66
รูปที่ 4.5 ผล Correlogram ของดัชนี SET50 ณ ผลต่างลำดับที่ 1 ( $1^{st}$ 1 Difference)	67
รูปที่ 4.6 ACF ของค่าเศษเหลือ	71
รูปที่ 4.7 ACF Scatter plot ของค่าเศษเหลือ	72
รูปที่ 4.8 Histogram ของค่าเศษเหลือ	72
รูปที่ 4.9 กราฟแสดงผลการพยากรณ์ผ่านตัวแบบ ARIMAX (2,0,2) ทั้งข้อมูลชุดฝึกสอน และข้อมูลชุดทดสอบ	73
รูปที่ 4.10 กราฟแสดงผลการพยากรณ์ผ่านตัวแบบ ANN (6-50-50-1) ทั้งข้อมูลชุดฝึกสอน และข้อมูลชุดทดสอบ	76
รูปที่ 4.11 กราฟแสดงผลการพยากรณ์ผ่านตัวแบบ ARIMAX-ANN ทั้งข้อมูลชุดฝึกสอน และข้อมูลชุดทดสอบ	79
รูปที่ 4.12 กราฟแสดงผลการพยากรณ์ผ่านตัวแบบ ARIMAX-SVR ทั้งข้อมูลชุดฝึกสอน และข้อมูลชุดทดสอบ	80
รูปที่ 4.13 กราฟแสดงผลการพยากรณ์ผ่านตัวแบบรวม ARIMAX-ANN-SVM ทั้งข้อมูลชุดฝึกสอน และข้อมูลชุดทดสอบ	82

สารบัญรูป(ต่อ)

หน้า

รูปที่ 4.14 กราฟแสดงผลการเปรียบเทียบผลลัพธ์ระหว่างสัญญาณซื้อ-ขายของ Bollinger band ที่ได้จากข้อมูลพยากรณ์กับข้อมูลจริง	84
รูปที่ 4.15 กราฟแสดงผลการเปรียบเทียบผลลัพธ์ระหว่างสัญญาณซื้อ-ขายของ Bollinger band ที่ได้จากข้อมูลพยากรณ์กับสัญญาณซื้อ-ขายของเครื่องมือ Moving Average ที่ได้จาก ข้อมูลจริง	85

## บทที่ 1

### บทนำ

#### 1.1 ที่มาและความสำคัญของปัญหา

ปัจจุบันนี้การลงทุนในตลาดหลักทรัพย์ของประเทศไทยเป็นสิ่งที่น่าสนใจเนื่องจากการลงทุนในตลาดหลักทรัพย์สามารถให้ผลตอบแทนที่สูงกว่าการฝากเงินไว้กับธนาคาร แต่เนื่องจากราคาหลักทรัพย์มีความผันผวนไม่แน่นอน อันนำไปสู่ความเสี่ยงในการลงทุนเมื่อราคาตกต่ำลงมากกว่าราคาเริ่มแรกที่นักลงทุนเข้าซื้ออย่างไรก็ตาม การลดความเสี่ยงจากการสูญเสียเงินทุนที่หายไปในตลาดหลักทรัพย์สามารถหลีกเลี่ยงได้ หากนำผลการวิเคราะห์เชิงเทคนิค (Technical analysis) เข้ามาช่วยในการตัดสินใจลงทุนเข้าซื้อขายหลักทรัพย์ใดๆ ก็ตาม (Gary P. Brinson, L. Randolph Hood and Gilbert L. Beebower, 1986)

ในการวิจัยนี้มีการนำเครื่องมือทางเทคนิคมาใช้ คือ แถบโบลิงเจอร์ (Bollinger bands : BB) โดยมีพื้นฐานมาจาก เส้นค่าเฉลี่ยเคลื่อนที่ (Moving Average Indicator) เป็นการจับสัญญาณซื้อขายโดยการดูความเปลี่ยนแปลงของราคาในอดีต ซึ่งถูกสร้างจากระดับค่า Standard deviation level ที่อยู่เหนือและต่ำกว่า Moving average เพื่อใช้วัดค่าความผันผวนของราคา (ชลนิศา พรประสิทธิ์, 2559) ซึ่งค่าความผันผวนสามารถดูได้จากการเคลื่อนที่ของกรอบของตัวบ่งชี้ (Indicator) ว่ากว้างขึ้นหรือแคบลง Bollinger Bands ประกอบด้วยเส้นสามเส้น ได้แก่ Upper Band, Middle Band และ Lower Band เนื่องจากเส้นสามเส้นที่กล่าวไปจะใช้สูตรคำนวณตาม Simple Moving Average (SMA) โดยปกติใช้ 20 วัน หากเป็นหุ้นไทย แต่เมื่อทดสอบไปเรื่อยๆ จะเห็นเส้นค่าเฉลี่ยดังกล่าวส่งสัญญาณล่าช้ากว่าเหตุการณ์จริงที่เกิดขึ้น มาถึงปี พ.ศ.2543 พบว่าไม่ควรใช้ค่าเฉลี่ยเคลื่อนที่เกิน 14 วัน (สนธิ อังสนากุล, 2547) ทำให้ทางคณะผู้จัดทำต้องหาวิธีปรับปรุงแก้ไขตัวส่งสัญญาณการซื้อขายล่วงหน้าของตัวบ่งชี้ Bollinger Band เป็นระยะเวลาล่วงหน้าจากตัวแบบพยากรณ์ราคาหุ้น

การสร้างตัวแบบพยากรณ์ราคาหุ้นนั้น สามารถในการทำนายการเคลื่อนไหวของราคาเพื่อหาเวลาซื้อหรือขายที่เหมาะสม โดยเฉพาะอย่างยิ่งในด้านการเงินเชิงปริมาณ การสร้างแบบจำลองอนุกรมเวลาและการพยากรณ์ราคาหุ้นได้กลายเป็นสาขาการวิจัยที่สำคัญ (Jesper Groenendijk, 2021)

การพยากรณ์เป็นการคาดคะเนของเหตุการณ์ในอนาคต โดยอาศัยรูปแบบของการเกิดเหตุการณ์หรือการพยากรณ์ที่เก็บข้อมูลจากอดีต รวมถึงความรู้ความสามารถของผู้พยากรณ์ หากทราบเหตุการณ์ต่าง ๆ ในอนาคต มีความเป็นไปได้ที่จะเพิ่มความเชื่อมั่นให้กับการวางแผนดำเนินงานที่มีความถูกต้องและผิดพลาดน้อยที่สุด ซึ่งมี

ความสอดคล้องกับการคาดการณ์ราคาหุ้นและวิธีการพยากรณ์อนุกรมเวลา สำหรับการพยากรณ์อนุกรมเวลา (time series forecasting) มีอยู่ด้วยกันหลายวิธี วิธีที่นิยมใช้กันมาก คือ วิธีของ Box-Jenkins ด้วยตัวแบบ ARIMA (Autoregressive Integrated Moving Average model) ซึ่งเป็นวิธีหาตัวแบบพยากรณ์โดยอาศัยความสัมพันธ์ของข้อมูลในอดีตเพื่อหาตัวแบบที่แสดงพฤติกรรมของข้อมูล และใช้เป็นแนวทางในการพยากรณ์ค่าในอนาคต (บุญกอง ทะกลโยธิน, ยุพาภรณ์ อารีพงษ์, 2561)

เมื่อประมาณ 30 ปีก่อน ได้เกิดตัวแบบพยากรณ์ใหม่ คือ Artificial intelligent และ Machine learning ขึ้นมา เป็นลักษณะของการสร้างตัวแบบที่ไม่จำเป็นต้องมีสมมติฐานรองรับ ต่างจากตัวแบบทางสถิติแบบดั้งเดิม มีความยืดหยุ่นในการสร้างตัวแบบ (Zhang et al., 2001) หนึ่งในตัวแบบที่ได้รับความนิยม คือ Artificial neural network (ANN) ถูกนำมาใช้ในการสร้างตัวแบบพยากรณ์อนุกรมเวลา (Zhang et al., 2001, de Groot and Würtz, 1991, Beale et al., 2013) และผลตอบรับก็เป็นที่น่าพอใจมาก ต่อมาในปี 1995 ก็เกิดตัวแบบสำหรับการประมาณค่าขึ้นมาใหม่ คือ Support vector machine for Regression (SVR) (Vapnik V, 1995) เป็นตัวแบบที่มีประสิทธิภาพความแม่นยำในการพยากรณ์ที่สูง มีลักษณะคล้าย ANN คือตัวแบบจะเน้นจับลักษณะข้อมูลไม่เป็นเชิงเส้นตรงได้ดี

(Stock and Watson ,1998) พบว่าประสิทธิภาพโดยรวมของตัวแบบเดี่ยวเหล่านี้สามารถปรับปรุงได้เมื่อนำตัวแบบมารวมกัน ตัวแบบARIMA เพียงอย่างเดียวไม่เพียงพอในการสร้างแบบจำลองและคาดการณ์ข้อมูลอนุกรมเวลาของราคาของหุ้นที่ไม่ได้สัมพันธ์แบบเชิงเส้นได้ เหมือนกับตัวแบบไม่เชิงเส้น เช่น Neural Networks แต่ตัวแบบไม่เชิงเส้นเพียงอย่างเดียวก็ไม่สามารถ มีประสิทธิภาพได้ดีเทียบกันรูปแบบ ด้วยการรวมวิธีการเชิงเส้นและไม่เชิงเส้นเข้าด้วยกัน โครงสร้างที่ซับซ้อนสามารถสร้างแบบจำลองได้แม่นยำยิ่งขึ้น

ในการวิจัยนี้ได้ทำการศึกษาเครื่องมือทางเทคนิคมาใช้ คือ Bollinger Bands มาใช้ร่วมกับตัวแบบพยากรณ์ราคาหุ้น โดยเลือกค่าพยากรณ์แต่ละตัวแบบมีความแม่นยำสูงที่สุดที่เหมาะสมกับหุ้นแต่ละตัวมาหาค่าเฉลี่ยเคลื่อนที่ และส่วนเบี่ยงเบนมาตรฐานเป็นระยะเวลา 20 วัน โดยการพยากรณ์ราคาหุ้นล่วงหน้าจากนำข้อมูลรายชื่อหลักทรัพย์ที่ใช้คำนวณดัชนี SET50 (หุ้นในกลุ่มดัชนี SET50) ตั้งแต่ช่วง 1 มกราคม พ.ศ.2563 ถึง 18 เมษายน พ.ศ. 2566 นำมาทดสอบกับตัวแบบARIMAX ,ตัวแบบSVR และ ตัวแบบANN ซึ่งเรียกว่า ตัวแบบเดี่ยว (Individual model) กับ ตัวแบบARIMAX ผสมตัวแบบSVR (ARIMAX-SVR) ,ตัวแบบARIMAX ผสมตัวแบบANN (ARIMAX-ANN) ,ตัวแบบรวมเกิดจากตัวแบบARIMAX ผสมตัวแบบSVR และตัวแบบ ANN ซึ่งเรียกว่า ตัวแบบผสม (Hybrid model)

## 1.2 วัตถุประสงค์ของการศึกษา

1.2.1 เพื่อหาตัวแบบที่ดีที่สุดและเหมาะสมที่สุดสำหรับการพยากรณ์ราคาหุ้นในหุ้นดัชนี SET50

1.2.2 เพื่อพยากรณ์ตัวส่งสัญญาณการซื้อ-ขายล่วงหน้าของตัวบ่งชี้ Bollinger Bands ร่วมกับการใช้ตัวแบบเดี่ยว ได้แก่ ARIMAX, ANN, SVR และตัวแบบผสม ได้แก่ ARIMAX-ANN, ARIMAX-SVR, ARIMAX-ANN-SVR เพื่อใช้สำหรับการหาตัวแบบพยากรณ์ราคาหุ้นที่ดีที่สุดและเหมาะสมกับการเป็นตัวแทนข้อมูลพยากรณ์กับข้อมูลในดัชนีหุ้น SET50

## 1.3 ประโยชน์ที่คาดว่าจะได้รับ

1.3.1 ช่วยให้ทราบถึงตัวแบบพยากรณ์ที่ดีที่สุดที่ใช้สำหรับพยากรณ์ราคาตลาดหุ้นในหุ้นดัชนี SET50

1.3.2 ได้สัญญาณการซื้อ-ขายล่วงหน้าโดยใช้ตัวบ่งชี้ Bollinger Bands ที่มีความถูกต้องและแม่นยำ เพื่อใช้สำหรับวางแผนการซื้อ-ขายหุ้นให้กับนักลงทุน

## 1.4 ขอบเขตการศึกษา

ในงานวิจัยนี้เป็นการศึกษาการทำนายสัญญาณการซื้อ-ขายโดยใช้ตัวบ่งชี้ Bollinger Bands ร่วมกับการใช้ตัวแบบเดี่ยว และตัวแบบผสม โดยประมวลผลจากโปรแกรม IBM SPSS Statistics 25 และ Google Colaboratory runs on Python 3.9. ซึ่งมีขอบเขต ของการศึกษาดังนี้

1.4.1 ข้อสมมติฐานในการวิเคราะห์ทางเทคนิคสำหรับงานวิจัยนี้

1.4.1.1 ราคาเป็นผลรวมที่สะท้อนให้ทราบ ถึงข่าวสารในด้านต่างๆทั้งหมดแล้ว

1.4.1.2 ราคาจะเคลื่อนไหวอย่างมีแนวโน้ม และจะคงอยู่ในแนวโน้มนั้นๆในช่วงระยะเวลาหนึ่ง  
จนกว่าจะ เกิดการเปลี่ยนแปลงแนวโน้มใหม่

1.4.1.3 พฤติกรรมการลงทุนของนักลงทุนในตลาดหุ้น จะยังคงมีลักษณะที่คล้ายกับพฤติกรรม  
การลงทุนในอดีต

1.4.2 ตัวบ่งชี้ Bollinger Bands สำหรับงานวิจัยนี้จะเลือกใช้ Simple Moving Average (SMA) ระยะเวลาเฉลี่ยเคลื่อนที่ 20 วัน (MA20)

1.4.3 อนุกรมเวลา (Time Series) ในที่นี่เราจะเลือกใช้ตัวแบบ Box–Jenkins หรือ Auto Regressive Integrated Moving Average (ARIMA) และ Autoregressive Moving Average with Exogenous Variable (ARIMAX)

1.4.4 ปัจจัยอื่นที่เกี่ยวข้องนอกเหนือจากราคาและปริมาณการซื้อขาย ซึ่งในที่นี่จะใช้ปัจจัยที่มีความสัมพันธ์กับดัชนีราคา SET50 ที่อ้างอิงมาจากการงานวิจัยอื่น ได้แก่

1.4.4.1 อัตราแลกเปลี่ยนค่าเงินบาทสหรัฐ

1.4.4.2 ราคาน้ำมันดิบเวสต์เท็กซัส

1.4.4.3 ราคาทองคำแท่ง (GOLD)

1.4.4.4 ดัชนี Dow Jones Industrial Average ประเทศสหรัฐอเมริกา

1.4.4.5 ดัชนี FTSE 100 ประเทศอังกฤษ

1.4.4.6 ดัชนีตลาดหลักทรัพย์ ประเทศสิงคโปร์

1.4.4.7 ดัชนี PSEI ประเทศฟิลิปปินส์

1.4.4.8 ดัชนี Merval Buenos Aires ประเทศอาร์เจนตินา

1.4.4.9 ดัชนี Nikkie 225 ประเทศญี่ปุ่น

1.4.4.10 ดัชนี Composite ประเทศอินโดนีเซีย

1.4.4.11 ดัชนี Hang Seng

1.4.4.12 ดัชนีราคา SET50 Indexย้อนหลัง1วัน

1.4.5 การเรียนรู้ของเครื่อง (Machine Learning) สำหรับงานวิจัยนี้จะใช้ Artificial Neural Network (ANN) และ Support Vector Machine for Regression (SVR) รวมถึงตัวแบบผสม ARIMAX-ANN, ARIMAX-SVM และ ARIMAX-ANN-SVM ในการเปรียบเทียบเพื่อเลือกตัวแบบที่ดีที่สุด

#### 1.4.6 ข้อมูลจริง

1.4.6.1 ข้อมูลราคาหุ้นที่อยู่ในดัชนี SET50 โดยใช้ข้อมูลย้อนหลังตั้งแต่ปี 1 มกราคม พ.ศ.2563 ถึง 18 เมษายน พ.ศ. 2566

1.4.6.2 ข้อมูลปัจจัยอื่นนอกเหนือจากราคาและปริมาณการซื้อขายที่มีผลต่อหุ้นดัชนี SET50

### 1.5 นิยามศัพท์

1.5.1 การวิเคราะห์ทางเทคนิค (Technical analysis : TA) คือการวิเคราะห์กราฟของข้อมูลของหุ้น เพื่อที่จะคาดการณ์การเคลื่อนไหวทิศทางของราคาในอนาคต โดยอาศัยข้อมูลจากการเคลื่อนไหวของราคาในอดีตที่ผ่านมา อาจผิดพลาด หรือคลาดเคลื่อนกันได้ โดย Technical Analysis จะเป็นตัวช่วยให้เทรดเดอร์หรือนักลงทุน สามารถเทรดหรือลงทุนได้อย่างมีประสิทธิภาพมากยิ่งขึ้น

1.5.2 ตัวบ่งชี้ หรือ ดัชนีชี้วัด หรือ อินดิเคเตอร์ (Indicator) คือเครื่องมือประเภทหนึ่งที่จะนำค่าตัวเลขหรือข้อมูลของหุ้นและอนุพันธ์มาผ่านการคำนวณตามหลักทางคณิตศาสตร์เพื่อนำเสนอ แบ่งแยก หรือ จัดเรียงข้อมูลตัวเลขนั้นๆ ออกมาให้เราได้ศึกษาได้อย่างเข้าใจมากยิ่งขึ้น เช่น รูปแบบแนวโน้ม (Trend), ความผันผวน (Volatility) และโมเมนตัม (Momentum) เป็นต้น โดยที่ตัวเลขหรือข้อมูลนั้นๆ จะมีความเกี่ยวข้องโดยตรงกับราคา จำนวนการ หรือ ดัชนีอ้างอิงการซื้อขายของหุ้นในตลาด ณ ขอบเขตเวลาที่สนใจ

1.5.3 เส้นค่าเฉลี่ยเคลื่อนที่แบบปกติ (Simple moving average : SMA) คือตัวบ่งชี้ที่ใช้ข้อมูลย้อนหลังมาคำนวณค่าเฉลี่ยแล้วพล็อตออกมาเป็นกราฟโดยให้ความสำคัญของข้อมูลที่ใช้คำนวณเท่าๆ บางครั้ง Simple moving average (SMA) อาจถูกเรียกในชื่อว่า Moving average (MA) ได้ ซึ่งความหมายก็คือตัวบ่งชี้ตัวเดียวกัน

1.5.4 แถบโบลลิงเจอร์ (Bollinger bands : BB) เป็นตัวบ่งชี้ซึ่งบอกความผันผวนของราคาในช่วงเวลาหนึ่ง โดยใช้วัดเครื่องมือทางการเงิน โภคภัณฑ์ หรือสินทรัพย์อื่นๆ ซึ่งแล้วแต่ผู้ใช้งานจะเรียกใช้ ถูกสร้างโดย John Bollinger ในช่วงปี 1980s

1.5.5 ดัชนี SET 50 (SET50 Index : SET50) เป็นดัชนีราคาหุ้น ที่แสดงระดับและความเคลื่อนไหวของราคาหุ้นสามัญ 50 หลักทรัพย์ ที่มีมูลค่าตามราคาตลาดสูง และการซื้อขายมีสภาพคล่องอย่างสม่ำเสมอ โดยจะมีการพิจารณาเลือกหุ้นเพื่อใช้ในการคำนวณ SET50 Index ทุกๆ 6 เดือน



1.5.6 เศรษฐศาสตร์มหภาค(Macroeconomics) คือการศึกษาระบบเศรษฐกิจโดยรวม ไม่ว่าจะเป็นเรื่องการทำงานและการว่างงาน อัตราค่าจ้าง อัตราดอกเบี้ย อัตราแลกเปลี่ยนเงินตราต่างประเทศ เพื่อที่จะอธิบายปรากฏการณ์ทางเศรษฐกิจที่เกิดขึ้น เช่นภาวะเงินเฟ้อ เศรษฐกิจถดถอย หรือ การขาดดุลบัญชีเดินสะพัด เป็นต้น

1.5.7 การเรียนรู้ของเครื่อง (อังกฤษ: machine learning : ML) เป็นการศึกษาอัลกอริทึมของคอมพิวเตอร์ที่มีการพัฒนา การเรียนรู้ของเครื่องถูกมองว่าเป็นส่วนหนึ่งของปัญญาประดิษฐ์ โดยอัลกอริทึมสร้างแบบจำลองทางคณิตศาสตร์จากข้อมูลตัวอย่าง เพื่อที่จะคาดการณ์หรือตัดสินใจได้อย่างชัดเจน การเรียนรู้ของเครื่องมีเกี่ยวข้องอย่างมากกับสถิติศาสตร์ เนื่องจากทั้งสองสาขาศึกษาการวิเคราะห์ข้อมูลเพื่อการทำนายเช่นกัน นอกจากนี้ยังมีความสัมพันธ์กับสาขาการหาค่าเหมาะที่สุดในทางคณิตศาสตร์ที่แง่ของวิธีการ ทฤษฎี และการประยุกต์ใช้

1.5.8 ตัวแบบ ARIMA (Autoregressive integrated moving average : ARIMA) เป็นเทคนิคพยากรณ์ซึ่งได้รับการเสนอโดย Box and Jenkins ในปี ค.ศ.1970 การพยากรณ์ด้วยวิธี ARIMA เป็นการอาศัยพฤติกรรมของข้อมูลในอดีต เพื่อกำหนดรูปแบบในปัจจุบัน และอธิบายแนวโน้มหรือปรากฏการณ์ต่างๆ ของตัวข้อมูลเองในอนาคต

1.5.9 ตัวแบบ ARIMAX (Autoregressive Integrated Moving Average with Exogenous Variable: ARIMAX) เป็นแบบจำลองทางสถิติที่ใช้สำหรับการวิเคราะห์อนุกรมเวลาที่รวมเอาตัวแปรภายนอก (ตัวแปรอิสระที่ไม่ได้เป็นส่วนหนึ่งของอนุกรมเวลาที่กำลังวิเคราะห์) เข้าในแบบจำลอง ARIMA ใน ARIMAX ตัวแปรตามจะถดถอยตามค่าที่ล่าช้าของตัวเอง ค่าที่ล่าช้าของตัวแปรภายนอก และค่าที่มีข้อผิดพลาดแบบสุ่ม สิ่งนี้ทำให้แบบจำลองสามารถอธิบายผลกระทบของตัวแปรภายนอกที่มีต่อตัวแปรตามได้

1.5.10 ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine : SVM) เป็นเทคนิคหนึ่งที่มีความนิยมอย่างแพร่หลายในงานที่เกี่ยวข้องกับการจดจำรูปแบบตลอดจนการแก้ปัญหาการจัดกลุ่ม (classification problem) (Wang et al. 2009, Chen et al. 2009) โดยอาศัยหลักการของการหาสัมประสิทธิ์ของสมการเพื่อสร้างเส้นแบ่งแยกกลุ่มข้อมูลที่ถูกป้อนเข้าสู่กระบวนการสอนให้ระบบเรียนรู้ โดยเน้นไปยังเส้นแบ่งแยกกลุ่มข้อมูลได้ดีที่สุด (optimal separating hyperplane)

1.5.11 ซัพพอร์ตเวกเตอร์รีเกรสชัน (Support Vector Regression : SVR) เป็นเทคนิคที่ใช้วิธีการของซัพพอร์ตเวกเตอร์แมชชีน มาวิเคราะห์ความถดถอยระหว่างอินพุตเวกเตอร์ (Input vector) และตัวแปร

เอาต์พุต (Output variables) ซึ่งนำมาใช้กับการพยากรณ์อนุกรมเวลาได้ โดยเปลี่ยนการจำแนกคลาสด้วย SVM เป็นการทำนายค่าด้วย SVR โดยมีเป้าหมายคือต้องการค้นหาความสัมพันธ์เชิงเส้นระหว่างอินพุตเวกเตอร์

1.5.12 **โครงข่ายประสาทเทียม (Artificial neural networks: ANN)** เป็นการจำลองการทำงานของสมองรูปแบบมาใช้กับโครงข่ายประสาทเทียมจะใช้โหนด (Nodes) ทำหน้าที่คล้ายกับ ตัวเซลล์ประสาท โดยโหนดจะรับค่านำเข้าได้หลายค่า เช่น แต่ผลการกระตุ้น หรือค่าส่งออกที่ได้มีเพียงหนึ่ง (Blum and Li, 1991)

1.5.13 **การวิเคราะห์ปัจจัย (Factor Analysis)** เป็นเทคนิควิธีทางสถิติที่ใช้ในการจับกลุ่ม รวมกลุ่ม หรือรวมตัวแปรที่มีความสัมพันธ์กันไว้ในกลุ่มเดียวกัน ซึ่งความสัมพันธ์จะเป็นไปได้ทั้งทางบวกและทางลบ โดยตัวแปรภายในองค์ประกอบเดียวกันจะมีความสัมพันธ์กันสูง ส่วนตัวแปรต่างองค์ประกอบกันจะมีความสัมพันธ์กันน้อย หรือไม่มีความสัมพันธ์กันซึ่งเทคนิควิธีทางสถิติวิธีนี้สามารถใช้ได้ทั้งการพัฒนาทฤษฎีใหม่หรือการทดสอบนัยสำคัญของทฤษฎีเดิม

1.5.14 **ตัวแบบเดี่ยว (Individual model)** ในการวิจัยครั้งนี้ได้แก่ ตัวแบบเชิงเส้น ได้แก่ ตัวแบบ ARIMAX , ตัวแบบไม่เชิงเส้น ได้แก่ SVR (Support vector regression) และตัวแบบ ANN (Artificial neural network)

1.5.15 **ตัวแบบผสมที่มีกระบวนการทำต่อเนื่องกัน (Traditional hybrid)** คือ ตัวแบบที่เริ่มจากตัวแบบ Linear เช่นตัวแบบ ARIMA ทำการสร้างตัวแบบแล้วพยากรณ์ จะได้ค่าเศษเหลือจากการพยากรณ์ (Residuals) จากนั้นนำค่าเศษเหลือตัวนี้ทำการพยากรณ์ต่อด้วยตัวแบบ Nonlinear เช่น ANN, SVR จึงจะได้ค่าพยากรณ์รวมที่นำไปใช้ได้ ในการวิจัยครั้งนี้ได้ใช้ ตัวแบบ ARIMA-SVR (ตัวแบบ ARIMA ผสม ตัวแบบ SVR) ,ตัวแบบ ARIMA-ANN (ตัวแบบ ARIMA ผสม ตัวแบบ SVR)

1.5.16 **ตัวแบบที่ทำการพยากรณ์แยกส่วนกัน (Hybrid combined)** คือ ทำการพยากรณ์ด้วยตัวแบบ Linear เช่นตัวแบบ ARIMAX และ ตัวแบบ Nonlinear เช่น ตัวแบบ ANN และ SVR แยกจาก กัน สามารถทำพร้อมๆกันบนเครื่องคอมพิวเตอร์เดียวกันได้ เมื่อสร้างตัวแบบเสร็จก็นำค่าพยากรณ์ จากทั้ง 3 ตัวแบบมารวมกันในรูปแบบผลรวมเชิงเส้น แล้วจึงหาค่าน้ำหนัก (Weights) ที่ดีที่สุดแล้วทำให้ค่าพยากรณ์รวมมีความแม่นยำมากที่สุด ในการวิจัยครั้งนี้ได้ใช้ ตัวแบบARIMA ผสม ตัวแบบSVR และตัวแบบANN ซึ่งจะเรียกว่า ตัวแบบรวม (Combined model)

1.5.17 **ตัวแบบผสม (Hybrid model)** ได้แก่ ตัวแบบ ARIMAX-SVR (ตัวแบบ ARIMAX ผสม ตัวแบบ SVR) ,ตัวแบบ ARIMA-ANN (ตัวแบบ ARIMA ผสม ตัวแบบ ANN) และ ตัวแบบรวม

## บทที่ 2

### ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

#### 2.1 ทฤษฎีที่เกี่ยวข้องกับการพยากรณ์

(ณฐา คุปต์ชัยธีร ,2558) ได้ให้ความหมายไว้ว่า การพยากรณ์ (Forecasting) หมายถึง กระบวนการในการคาดเดาเหตุการณ์ในอนาคต หรือเหตุการณ์ที่ยังไม่เกิดขึ้น โดยการพยากรณ์ สามารถแบ่งได้ตามระยะเวลาในการพยากรณ์ออกเป็น 3 ประเภท ดังนี้

1. การพยากรณ์ระยะสั้น (Short-term Forecasting) มักจะอยู่ในช่วงเวลาไม่เกิน 1 ปี
2. การพยากรณ์ระยะปานกลาง (Intermediate Forecasting) มักจะอยู่ในช่วงเวลา 3 เดือนถึง 3 ปี
3. การพยากรณ์ระยะยาว (Long-term Forecasting) มักจะอยู่ในช่วงเวลามากกว่า 3 ปีขึ้นไป

(อนุสรณ์ บุญสง่า ,2559) ได้ให้ความหมายไว้ว่า การพยากรณ์ คือการคาดคะเน หรือทำนายเหตุการณ์ในอนาคตจากข้อมูลในอดีต ปัจจุบัน หรือประสบการณ์ การพยากรณ์เป็นทั้งศาสตร์และศิลป์ ตั้งแต่สมัยโบราณมนุษย์รู้จักการพยากรณ์เพื่อการดำรงชีวิต เช่น การพยากรณ์ดินฟ้าอากาศเพื่อการล่าสัตว์และเพาะปลูก ปัจจุบันการพยากรณ์ได้ถูกนำมาใช้ในการตัดสินใจสำหรับการดำเนินชีวิตประจำวันสำหรับแต่ละคน จนถึงการทำเนิกรกิจกรรมในองค์กรต่างๆ การพยากรณ์จะให้ค่าพยากรณ์ คือจำนวน หรือปริมาณที่ต้องการทราบในอนาคต ในธุรกิจค่าพยากรณ์ที่สำคัญ คือ ปริมาณความต้องการสินค้าหรือบริการในอนาคตที่ฝ่ายการตลาดจะทำการพยากรณ์ออกมาและฝ่ายผลิตจะนำไปใช้ในการวางแผนการผลิตต่อไป

##### 2.1.1 ขั้นตอนในการพยากรณ์

ขั้นตอนในการพยากรณ์พื้นฐานอยู่ทั้งหมด 6 ขั้นตอนดังนี้

**ขั้นตอนที่1:** กำหนดวัตถุประสงค์ในการพยากรณ์ว่าต้องการพยากรณ์เพื่ออะไร ต้องการผลลัพธ์จากการพยากรณ์เมื่อไร วัตถุประสงค์ของการพยากรณ์จะนำมาใช้เป็นแนวทางในการกำหนด

**ขั้นตอนที่2:** กำหนดระยะเวลาในการพยากรณ์เนื่องจากยิ่งพยากรณ์ในช่วงเวลานานขึ้นมากเท่าไร ความแม่นยำของการพยากรณ์จะลดลงเท่านั้น

**ขั้นตอนที่3:** เลือกวิธีการพยากรณ์ที่เหมาะสม เช่น การเลือกระหว่างการพยากรณ์เชิงปริมาณกับเชิงคุณภาพ รวมทั้งเทคนิคการวิเคราะห์ข้อมูล

**ขั้นตอนที่4:** รวบรวมและวิเคราะห์ข้อมูล

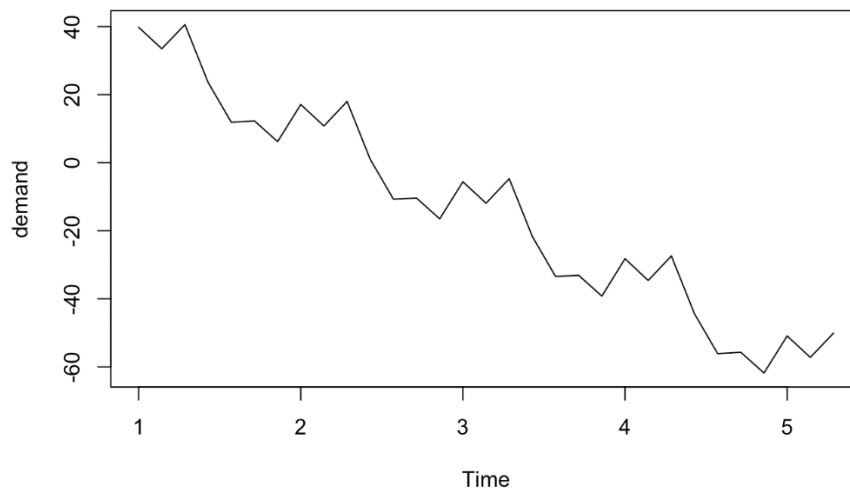
**ขั้นตอนที่5:** ทำการพยากรณ์ด้วยวิธีการพยากรณ์ที่เลือกไว้

**ขั้นตอนที่6:** ติดตามผลการพยากรณ์ คำนวณค่าความแม่นยำ เพื่อตรวจสอบวิธีการ

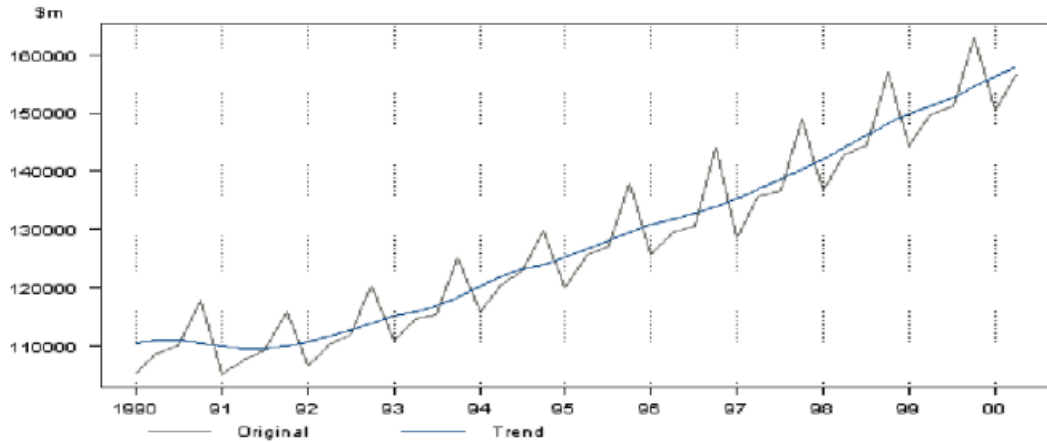
พยากรณ์สามารถใช้ได้ผลดีและตรงกับระดับความแม่นยำที่ต้องการ

### 2.1.2 การพยากรณ์แบบอนุกรมเวลา (Time Series Forecasting)

การเก็บข้อมูลความต้องการสินค้าหรือยอดขายระยะเวลาหนึ่งสม่ำเสมอโดยมีช่วงห่างในการเก็บข้อมูลที่เท่าๆกัน องค์ประกอบของอนุกรมเวลา (Time Series Components) เทคนิคการพยากรณ์แบบอนุกรมเวลา มีสมมติฐานว่าความต้องการสินค้ามีการเปลี่ยนแปลงตามระยะเวลาที่ผ่านไปดังนั้นการ คำนวณด้วยสมการทางคณิตศาสตร์ ตัวแปรที่มีผลต่อความต้องการสินค้าคือเวลา (Time) โดยทั่วไปความต้องการสินค้าจะมากหรือน้อยนั้นเกิดจากอิทธิพล 4 ประการได้แก่ แนวโน้ม (Trend) ฤดูกาล(Seasonal) วัฏจักร (Cycle) และความผิดปกติหรือความไม่แน่นอน (Irregular or Random) แนวโน้ม (Trend) หมายถึง ความต้องการสินค้ามีลักษณะเพิ่มขึ้นหรือลดลงอย่างค่อยเป็นค่อยไป โดยส่วนมากอิทธิพลของแนวโน้มมีสาเหตุมาจากการเปลี่ยนแปลงจำนวนประชากร การเปลี่ยนแปลงรายได้ของครอบครัวและลักษณะทางวัฒนธรรม ดังแสดงในรูป

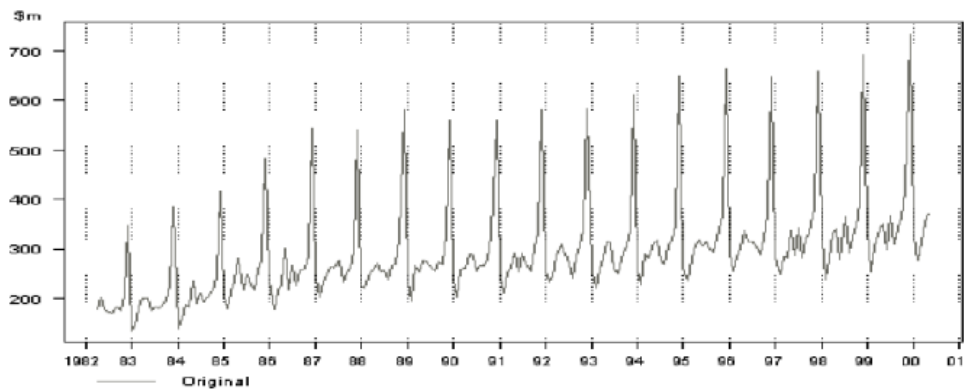


**รูปที่ 2.1** กราฟแสดงลักษณะของแนวโน้มลดลง ที่มา: เว็บไซต์สำนักงานสถิติแห่งชาติ



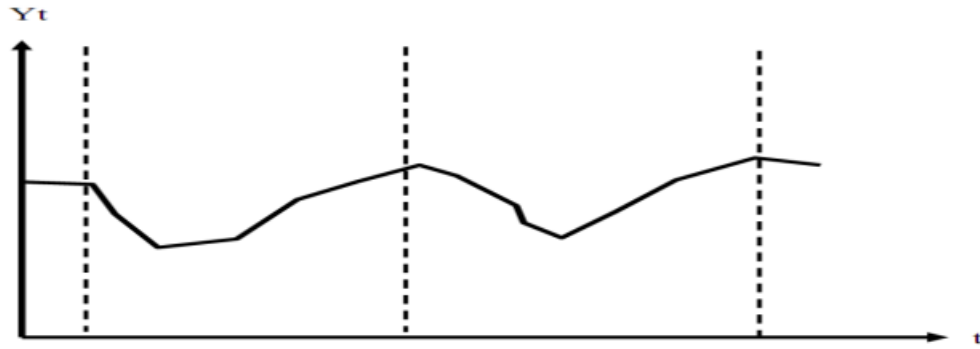
รูปที่ 2.2 กราฟแสดงลักษณะของแนวโน้มเพิ่มขึ้น ที่มา: เว็บไซต์สำนักงานสถิติแห่งชาติ

ฤดูกาล (Seasonal) หมายถึง ความต้องการสินค้ามีลักษณะเป็นรูป แบบซ้ำๆ กัน จากผลกระทบของ ฤดูกาลโดยสามารถสังเกตเห็นรูปแบบ Pattern ที่ชัดเจนและเกิดตามรูปแบบนั้นๆ ซ้ำในชวงเวลาถัดมา โดย ส่วนมากลักษณะความต้องการสินค้าแบบฤดูกาลมีอิทธิพลมาจากสภาพ ภูมิอากาศ เทศกาล และวันหยุดต่าง ๆ ดัง แสดงในรูป



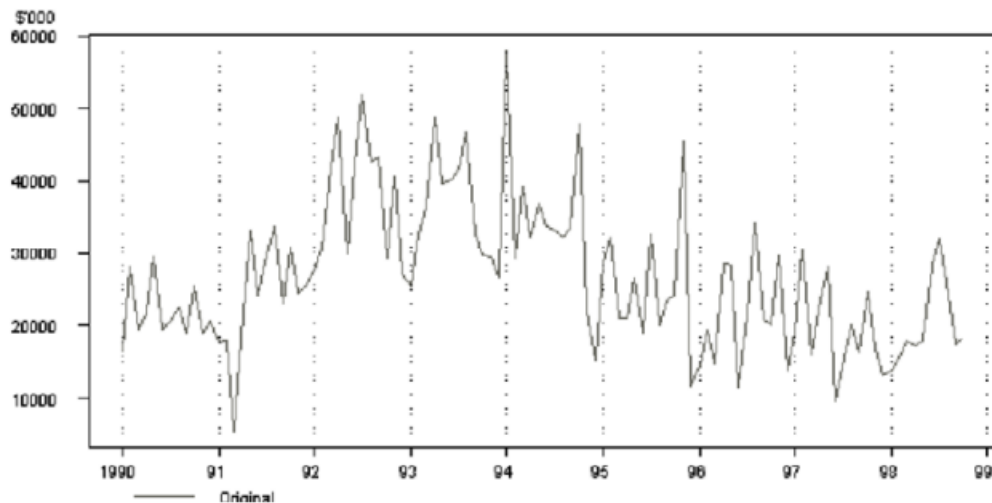
รูปที่ 2.3 กราฟแสดงลักษณะของฤดูกาล ที่มา: เว็บไซต์สำนักงานสถิติแห่งชาติ

วัฏจักร (Cycle) หมายถึง ความต้องการสินค้าที่มีลักษณะเพิ่มขึ้นหรือลดลง ซึ่งเป็นผลกระทบจากสภาวะ การเปลี่ยนแปลงด้วยปัจจัยด้านเศรษฐกิจหรือการเมืองสามารถสังเกตจากเส้นกราฟมีลักษณะขึ้น-ลงคล้ายคลื่น โดยช่วงระยะเวลาของการเกิดอิทธิพลของวัฏจักรครอบคลุมมากกว่า 1 ปีขึ้นไป ดังแสดงในรูป



รูปที่ 2.4 กราฟแสดงลักษณะของวัฏจักร ที่มา: เทคนิคการพยากรณ์, กรุงเทพฯ, โครงการส่งเสริมเอกสา  
วิชาการ สถาบันบัณฑิตพัฒนบริหารศาสตร์ โดยวิจิต หล่อจิระสุนท์, สมบูรณ์วัลย์ สัตยารักษ์วิทย์,  
จิราวัลย์ จิตรถเวช (2548)

ความผิดปกติหรือความไม่แน่นอน (Irregular or Random) หมายถึง ภัยจากธรรมชาติต่าง ๆ สงคราม  
การประท้วง สามารถก่อให้เกิดการเปลี่ยนแปลงความต้องการสินค้าได้ หากพบเหตุการณ์เช่นนี้ในการพยากรณ์  
มักจะตัดข้อมูลที่ผิดปกติทิ้งไปก่อนที่จะทำการพยากรณ์ต่อไปดังแสดงในรูป



รูปที่ 2.5 กราฟแสดงลักษณะของความผิดปกติ ที่มา: : เว็บไซต์สำนักงานสถิติแห่งชาติ

## 2.2 ตัวแบบ Autoregressive Integrated Moving Average (ARIMA)

การวิเคราะห์อนุกรมเวลาแบบ Box – Jenkins (Box – Jenkins, 1976) เป็นระเบียบวิธีทางสถิติ สำหรับหาตัวแบบพยากรณ์ เทคนิคนี้อาศัยความสัมพันธ์จากข้อมูลในอดีตเพื่อหาตัวแบบแสดงพฤติกรรมของข้อมูลและใช้เป็นแนวทางในการพยากรณ์พฤติกรรมในอนาคตเป็นเทคนิคที่มีประสิทธิภาพสูงสำหรับการพยากรณ์ในระยะสั้น ตัวแบบที่ใช้ในการวิเคราะห์อนุกรมเวลาแบบ Box-Jenkins เรียกตัวแบบ ARIMA ซึ่งย่อมาจากคำว่า Autoregressive Integrated Moving Average ตัวแบบ ARIMA มีข้อสมมติว่าค่าปัจจุบันของค่าสังเกตเป็นฟังก์ชันเชิงเส้นของค่าสังเกต และค่าความคลาดเคลื่อนสุ่มในอดีต ตัวแบบคือ

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

โดย  $y_t$  แทนค่าสังเกตของอนุกรมเวลาที่เวลา  $t$

$p$  แทนอันดับของ Autoregressive

$q$  แทนอันดับของ Moving Average

$\phi_i = (i=1, \dots, p)$  และ  $\theta_i = (i=1, \dots, q)$  แทนพารามิเตอร์ในตัวแบบ ส่วน  $p$  และ  $q$  เป็นจำนวนเต็ม ซึ่งแสดงอันดับของตัวแบบ

$\varepsilon_t$  แทนความคลาดเคลื่อนสุ่มที่เวลา  $t$  ซึ่งมีข้อสมมติว่า  $\varepsilon_t$  เป็นตัวแปรสุ่มที่เป็นอิสระกัน ค่าเฉลี่ยเท่ากับ 0 และ ค่าความแปรปรวนคงที่

ส่วนของค่าสังเกตในอดีตที่อยู่ในตัวแบบเรียกเป็นเทอมการถดถอยในตนเอง (Autoregressive หรือ AR) และส่วนของความคลาดเคลื่อนสุ่มในอดีตที่อยู่ในตัวแบบเรียกเป็นเทอมค่าเฉลี่ยเคลื่อนที่ (Moving Average หรือ MA) และอนุกรมเวลาที่ต้องมีการหาผลต่างก่อนเพื่อแปลงให้เป็นอนุกรมเวลาที่มีลักษณะคงที่เรียกว่าเป็นอนุกรมเวลาที่ปรับให้มีลักษณะคงที่แล้ว (Integrated version of a stationary series หรือ I) ตัวแบบ ARIMA นิยมเขียนโดยบอกอันดับ  $p$  ของ AR อันดับ  $d$  ของการหาผลต่างของอนุกรมเพื่อให้เป็นอนุกรมเวลาลักษณะคงที่และอันดับ  $q$  ของ MA โดยเขียนแทนด้วย  $ARIMA(p, d, q)$  ถ้า  $q=0$  จะเป็นตัวแบบ  $AR(p)$  แต่ถ้า  $p=0$  จะเป็นตัวแบบ  $MA(q)$  ในการวิเคราะห์อนุกรมเวลานั้น Box-Jenkins ได้แบ่งออกเป็น 3 ขั้นตอนหลักๆ คือ

1. การกำหนดตัวแบบ
2. การประมาณค่าพารามิเตอร์ของตัวแบบ ARIMA



### 3. การตรวจสอบความเหมาะสมของตัวแบบ

#### 2.2.1 การกำหนดตัวแบบ ARIMA

##### 2.2.1.1 ฟังก์ชันสหสัมพันธ์ในตัวเอง (Autocorrelation Function: ACF)

ฟังก์ชันสหสัมพันธ์ในตัวเองเป็นมาตรวัดความสัมพันธ์ในค่าของข้อมูลที่เกิดขึ้น ณ เวลา ต่างๆ สัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร  $Y_t$  และ  $Y_{t+k}$  ในอนุกรมเวลาที่ อยู่ห่างกัน  $k$  ช่วงเวลา เรียกว่าสหสัมพันธ์ในตัวเองที่อยู่ห่างกัน  $k$  ช่วงเวลาแทนด้วย ช่วงเวลา (Sample Autocorrelation of Lag  $k$ ) แทนด้วย

$$r_k = \frac{\sum_{t=1}^{n-k} (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2} ; k = 1, 2, \dots, n$$

โดยที่

$$\bar{Y} = \sum_{t=1}^n \frac{Y_t}{n} ; t = 1, 2, \dots, n$$

ในการทดสอบความมีนัยสำคัญของสหสัมพันธ์ในตัวเองจากตัวอย่างมีขั้นตอนในการทดสอบดังนี้

1. สมมติฐานในการทดสอบ:  $H_0 : \rho_k = 0$

$$H_1 : \rho_k \neq 0$$

2. ตัวสถิติที่ใช้ทดสอบ:  $r_k$

3. กำหนดระดับนัยสำคัญ:  $\alpha = 0.05$

4. เขตวิกฤต: ปฏิเสธ  $H_0$  ถ้า  $|r_k| = \frac{t_{\frac{\alpha}{2}, df}}{\sqrt{n}} = \frac{1.96}{\sqrt{n}} \approx \frac{2}{\sqrt{n}}$

5. สรุปผล: ถ้าปฏิเสธ  $H_0$  แสดงว่าสหสัมพันธ์ในตัวเองที่ อยู่ห่างกัน  $k$  ช่วงเวลาแตกต่างจากศูนย์

### 2.2.1.2 ฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วน (Partial Autocorrelation Function: PACF)

สหสัมพันธ์ในตัวเองบางส่วนคือสหสัมพันธ์ในตัวเอง (autocorrelation) ระหว่างตัวแปร  $Y_t$  และ  $Y_{t+1}$  ในอนุกรมเวลาที่อยู่ห่างกัน  $k$  ช่วงเวลาที่ซัดอิทธิพลของตัวแปรที่อยู่ระหว่างตัวแปรทั้งสองได้แก่  $Y_{t+1}, Y_{t+2}, \dots, Y_{t+k-1}$  ออกไป แทนด้วย  $\sigma_{kk}$  และสามารถประมาณได้ด้วยสหสัมพันธ์ในตัวเองบางส่วนจากตัวอย่างที่อยู่ห่างกัน  $k$  ช่วงเวลา (Sample Partial Autocorrelation of Lag  $k$ ) แทนด้วย

$$r_{kk} = \begin{cases} r_1 & \text{if } k = 1 \\ \frac{r_k - \sum_{j=1}^{k-1} r_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1,j} r_j} & \text{if } k = 2, 3, \dots \end{cases}$$

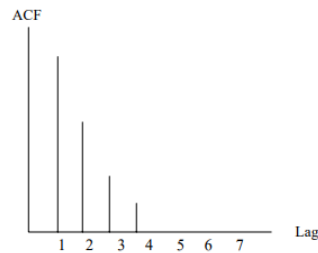
โดยที่  $r_{kj} = r_{k-1,j} - r_{kk} r_{k-1,k-j} ; j = 1, 2, \dots, k-1$

ในการทดสอบความมีนัยสำคัญของสหสัมพันธ์ในตัวเองบางส่วนจากตัวอย่างมีขั้นตอนในการทดสอบดังนี้

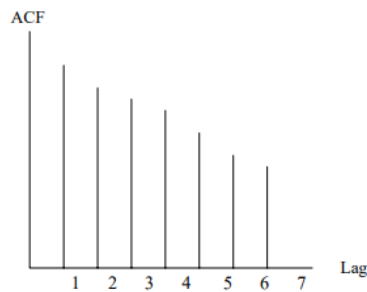
1. สมมติฐานในการทดสอบ:  $H_0 : \rho_{kk} = 0$   
 $H_1 : \rho_{kk} \neq 0$
2. ตัวสถิติที่ใช้ทดสอบ:  $r_{kk}$
3. กำหนดระดับนัยสำคัญ:  $\alpha = 0.05$
4. เขตวิกฤติ: ปฏิเสธ  $H_0$  ถ้า  $|r_{kk}| = \frac{t_{\frac{\alpha}{2}, df}}{\sqrt{n}} = \frac{1.96}{\sqrt{n}} \approx \frac{2}{\sqrt{n}}$
5. การสรุปผล: ถ้าปฏิเสธ  $H_0$  แสดงว่าสหสัมพันธ์ในตัวเองบางส่วนที่อยู่ห่างกัน  $k$  ช่วงเวลาแตกต่างจากศูนย์

ข้อมูลอนุกรมเวลาที่จะนำมาสร้างตัวแบบต้องมีสมบัติคงที่ (Stationary) คือลักษณะทาง สถิติ เช่น ค่าเฉลี่ยความแปรปรวน และ ลักษณะของสหสัมพันธ์ในตัวเองของข้อมูลต้องคงที่ตลอดช่วงเวลา การพิจารณาว่าของข้อมูลที่ stationary หรือไม่อาจดูได้จากฟังก์ชันสหสัมพันธ์ในตนเองในเชิงทฤษฎี

(Theoretical Autocorrelation Function : ACF) กล่าวคือ ACF ของอนุกรมเวลาที่คงที่ จะมีลักษณะลดลง (die down) เข้าสู่ศูนย์อย่างรวดเร็วเมื่อข้อมูลอนุกรมเวลามีระยะห่างกันเพิ่มขึ้นหรือถูกตัดออก (cut off) ที่ระยะห่างช่วงใดช่วงหนึ่ง แต่ถ้า ACF มีลักษณะลดลงเข้าสู่ศูนย์ช้าๆ แสดงว่าอนุกรมเวลาข้อมูลไม่คงที่ (Non-stationary) ถ้าข้อมูลอนุกรมเวลามีแนวโน้ม หรือความแปรปรวนไม่คงที่ จะต้องปรับให้มีลักษณะคงที่ จะต้องปรับให้มีลักษณะคงที่ ก่อนโดยการหาผลต่าง หรือการแปลงรูปกำลัง (Power transformation) (สุชัยศรี โลออนส์, 2540)



รูปที่ 2.6 ACF (Stationary)



รูปที่ 2.7 ACF (Non-Stationary)

### 2.2.1.3 การกำหนดตัวแบบและอันดับของ $p$ และ $q$

การกำหนดตัวแบบและอันดับของตัวแบบทำโดยการเปรียบเทียบลักษณะของฟังก์ชันสหสัมพันธ์ในตัวเอง (ACF) และฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วน (Partial Auto-Correlation Function : PACF) ของค่าสังเกตกับลักษณะของ ACF และ PACF ตามทฤษฎีของตัวแบบ ARIMA อันดับต่าง ๆ ซึ่งเสนอโดย Box-Jenkins (1976) โดยการกำหนดอันดับของ  $p$  และ  $q$  พิจารณา จาก ACF และ PACF ประกอบกัน ถ้าเป็น ตัวแบบ AR อันดับ  $p$  ( $AR(p)$ ) สหสัมพันธ์ในตัวเอง (ACF) จะมีลักษณะลดลงเข้าสู่ศูนย์อย่างรวดเร็ว และ สหสัมพันธ์ในตัวเองบางส่วน (PACF) ที่ห่างกันเกิน  $p$  ช่วงเวลาจะมีลักษณะลดลงเข้าสู่ศูนย์อย่างรวดเร็ว

ถ้าเป็นตัวแบบ MA อันดับ  $q$  ( $MA(q)$ ) สหสัมพันธ์ในตัวเองที่ห่างกันเกิน  $q$  ช่วงเวลา จะมีลักษณะลดลงเข้าสู่ศูนย์อย่างรวดเร็วและสหสัมพันธ์ในตัวเองบางส่วนจะมีลักษณะลดลงเข้าสู่ศูนย์อย่างรวดเร็ว

ถ้าเป็นแบบผสมระหว่าง AR และ MA (Mixed Autoregressive Moving Average :  $ARMA(p, q)$ ) สหสัมพันธ์ในตัวเองที่ห่างกันเกิน  $q$  ช่วงเวลา และสหสัมพันธ์ในตัวเองบางส่วนที่ห่างกันเกิน  $p$  ช่วงเวลา จะมีลักษณะลดลงเข้าสู่ศูนย์อย่างรวดเร็ว

ถ้าทั้งสหสัมพันธ์ในตัวเองและสหสัมพันธ์ในตัวเองบางส่วนมีค่าเท่ากับศูนย์ทุกช่วงห่างจะได้แบบจำลองที่เรียกว่า White Noise นิยามคือ

$$Y_t = \theta_0 + \varepsilon_t$$

โดยที่

$$\varepsilon_t \sim IN(0, \sigma_A^2)$$

ถ้าทั้งสหสัมพันธ์ในตัวเองและสหสัมพันธ์ในตัวเองบางส่วนจากข้อมูลอนุกรมเวลาที่ได้จากการแปลงให้เป็นอนุกรมเวลาคงที่โดยการหาผลต่าง มีค่าเท่ากับศูนย์ จะได้แบบที่เรียกว่า Random Walk นิยามคือ

$$Y_t = \theta_{t-1} + \theta_0 + \varepsilon_t$$

โดยที่

$$\varepsilon_t \sim IN(0, \sigma_A^2)$$

## 2.2.2 การประมาณค่าพารามิเตอร์ (Estimation)

กำลังสอง ( $\sum_{i=1}^n \varepsilon_i^2$ ) มีค่าน้อยที่สุด โดยที่  $\hat{\varepsilon}_i$  คือค่าประมาณความคลาดเคลื่อน  $\varepsilon_i$  จากนั้นนำแบบจำลองที่ได้ไปตรวจสอบเหมาะสม

## 2.2.3 การตรวจสอบความเหมาะสมของตัวแบบ (Diagnostic Checking)

ในขั้นตอนการตรวจสอบความเหมาะสมของแบบจำลอง เป็นการตรวจสอบว่าแบบจำลองที่เลือกมีความเหมาะสมกับข้อมูลอนุกรมเวลาที่ทำarviเคราะห์หรือไม่โดย

### 2.2.3.1 ทดสอบค่าประมาณพารามิเตอร์ในแบบจำลองตามขั้นตอนดังนี้

1. สมมติฐานในการทดสอบ:  $H_0$  : พารามิเตอร์ = 0

$H_1$  : พารามิเตอร์  $\neq 0$

2. ตัวสถิติที่ใช้ทดสอบ:  $t = \frac{\text{ตัวประมาณพารามิเตอร์}}{\text{ส่วนเบี่ยงเบนมาตรฐาน}}$

3. กำหนดระดับนัยสำคัญ:  $\alpha = 0.05$

4. เขตวิกฤต: ปฏิเสธ  $H_0$  ถ้า  $|t| > t_{\frac{\alpha}{2}, df}$

5. สรุปผล: ถ้าผลการทดสอบนำไปสู่การปฏิเสธสมมติฐานว่าง แสดงว่าพารามิเตอร์ตัวที่ทดสอบนั้นแตกต่างจากศูนย์ ควรรวมพารามิเตอร์ตัวนั้นอยู่ในแบบจำลอง

### 2.2.3.2 พิจารณาค่าประมาณความคลาดเคลื่อนที่คำนวณได้จากการสร้างสมการพยากรณ์แบบจำลอง

1. สร้างกราฟค่าประมาณความคลาดเคลื่อนเทียบกับเวลา ถ้ากราฟที่ได้แสดงให้เห็นว่าค่าประมาณความคลาดเคลื่อนมีการกระจายไม่คงที่เปลี่ยนแปลงตามเวลา แสดงว่าค่าความคลาดเคลื่อน ถ้ามีแบบจำลองเป็น White Noise แสดงว่าแบบจำลองที่ได้เป็นแบบจำลองที่เหมาะสม

2. ดูจาก Theoretical Autocorrelation Function (ACF) และ Theoretical Partial Autocorrelation Function (PACF) ของค่าประมาณความคลาดเคลื่อน ถ้ามีแบบจำลองเป็น White Noise แสดงว่าแบบจำลองที่ได้เป็นแบบจำลองที่เหมาะสม

3. การทดสอบสหสัมพันธ์ในตนเองของค่าประมาณความคลาดเคลื่อน Box และ Pierce ได้เสนอวิธีทดสอบ โดยสร้างตัวสถิติที่มีการแจกแจงแบบไคสแควร์ที่มีค่าองศาแห่งความอิสระ (degree of freedom) เท่ากับจำนวนสหสัมพันธ์ในตัวเองจากตัวอย่าง (sample autocorrelations) ของค่าประมาณความคลาดเคลื่อน ( $k$ ) ตัวที่ใช้ในการทดสอบลบจำนวนพารามิเตอร์ ( $m$ ) ที่ประมาณค่าในแบบจำลอง ตัวสถิตินี้เรียกว่า Box-Pierce Chi-Square Statistic

$$Q = n \sum_{k=1}^K r_k^2(\hat{\epsilon})$$

โดยที่  $n$  แทนจำนวนข้อมูลของค่าความคลาดเคลื่อนของค่าสังเกตอนุกรมเวลา  
 หลังจากปรับให้เป็นอนุกรมเวลาคงที่แล้ว  
 $K$  แทนจำนวนสหสัมพันธ์ในตัวเองของค่าประมาณความคลาดเคลื่อนที่ใช้  
 ในการคำนวณค่า  $Q$   
 $r_K^2(\xi)$  แทนฟังก์ชันสหสัมพันธ์ในตัวเองของค่าประมาณความคลาดเคลื่อน

ตัวสถิติ Box-Pierce คำนวณจากฟังก์ชันสหสัมพันธ์ในตัวเองของค่าประมาณความ  
 คลาดเคลื่อนในอนุกรมเวลาที่ห่างกันในหน่วยเวลาต่าง ๆ ซึ่งจะทดสอบว่าความคลาดเคลื่อนนี้เป็นอิสระต่อกัน  
 หรือไม่

1. สมมติฐานในการทดสอบ  $H_0 : \rho_K = 0$  ทุกๆค่าที่  $k \leq K$   
 $H_1 : \rho_K \neq 0$  อย่างน้อย 1 ค่าที่  $k \leq K$
2. ตัวสถิติที่ใช้ทดสอบ:  $t = n \sum_{K=1}^K r_K^2(\xi)$
3. กำหนดระดับนัยสำคัญ:  $\alpha = 0.05$
4. เขตวิกฤต: ปฏิเสธ  $H_0$  ถ้า  $Q > \chi_{\alpha, K-m}$
5. สรุปผล: ถ้าผลการทดสอบนำไปสู่การปฏิเสธสมมติฐานว่าง แสดงว่าความ  
 คลาดเคลื่อนยังมีสหสัมพันธ์กันอยู่ จึงต้องปรับปรุงแบบจำลองใหม่แต่ถ้าผลการ  
 ทดสอบยอมรับสมมติฐาน แสดงว่าแบบจำลองที่ได้เป็นแบบจำลองที่เหมาะสม  
 ถ้าพบว่าแบบจำลองที่ได้ไม่เหมาะสมจะต้องพิจารณาเลือกแบบจำลองใหม่คือกลับไป  
 ที่ขั้นตอนกำหนดแบบจำลองใหม่จนกว่าจะได้แบบจำลองที่เหมาะสม

### 2.3 ตัวแบบ Autoregressive Integrated Moving Average with Exogenous Variable (ARIMAX)

(Wiwik Anggraeni ,2017) เป็นตัวแบบอธิบายร่วมกันของแบบจำลองอาร์ีมา กับตัวแปรภายนอกหรือ  
 ปัจจัยอื่น ( $x$ ) ที่น่าจะมีอิทธิพลต่อ  $y$ , โดย ตัวแบบARIMAX ถูกมองเชิงตัวแบบการถดถอยพหุคูณ ตัวแบบมี  
 Autoregressive (AR) อย่างน้อยหนึ่งเทอมและ Moving Average (MA) อย่างน้อยหนึ่งเทอม โดยทั่วไปตัวแบบ  
 ARIMAX มีรูปแบบดังนี้

Autoregressive model with exogenous variables (ARX):

$$y_t = \phi(L)y_t + \beta X_t + \varepsilon_t$$

Moving average model with exogenous variables (MAX):

$$y_t = \beta X_t + \theta(L)\varepsilon_t$$

Autoregressive Integrated Moving Average with exogenous variables model (ARIMAX):

$$\phi(L)\Delta^d y_t = \beta X_t + \theta(L)\varepsilon_t$$

โดย  $d$  แทนจำนวนครั้งของการหาผลต่างเพื่อให้อนุกรมเวลามีคุณสมบัติคงที่ (Stationary)

$p$  แทนอันดับของ Autoregressive

$q$  แทนอันดับของ Moving Average

$y_t$  แทนค่าสังเกตของอนุกรมเวลาที่เวลา  $t$

$\Delta^d$  แทนผลต่างระดับที่  $d$  ( $d$ -th differences operator)

$\phi(L)$  แทน Autoregressive Polynomial ซึ่ง  $\phi(L) = (1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)$

$\theta(L)$  แทน Moving Average Polynomial ซึ่ง

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

$\varepsilon_t$  แทนค่าส่วนเหลือเวลา  $t$  กำหนดให้มีคุณสมบัติ White Noise Process

ซึ่งมีข้อสมมติว่า  $\varepsilon_t$  เป็นตัวแปรสุ่มที่เป็นอิสระกันค่าเฉลี่ยเท่ากับ 0 และ ค่าความแปรปรวนคงที่

$\beta$  แทนค่าสัมประสิทธิ์ของตัวแปรภายนอกที่เวลา  $t$

$X_t$  แทนตัวแปรภายนอกที่เวลา  $t$

## 2.4 โครงข่ายประสาทเทียม (Artificial Neural Network: ANN)

โครงข่ายประสาทเทียม (Artificial Neural Network: ANN) เป็นวิธีที่ถูกพัฒนามาจากการทำงานของเซลล์ประสาททางชีววิทยา (Biological Neuron) ซึ่งมีความสามารถในการปรับตัวเมื่อข้อมูลไม่เป็นเชิงเส้น มีความสามารถในการเรียนรู้จากตัวอย่างและการทำให้เป็นทั่วไป (Generalize) โครงข่ายจะถูกฝึกการเรียนรู้โดยการ

แสดงรูปแบบ (Pattern) ต่างๆ ที่ต้องการให้โครงข่ายเรียนรู้ด้วยกฎการเรียนรู้ (Learning Rule) การที่โครงข่ายถูกทำให้เป็นกรณีทั่วไปจะทำให้ตัวโครงข่ายสามารถที่จะจำแนกรูปแบบของข้อมูลนำเข้าแบบใหม่ๆ ที่ตัวโครงข่ายไม่รู้จักมาก่อนในระดับที่เป็นที่ยอมรับได้ ตัวโครงข่ายจะทำการเก็บข้อมูลความรู้ (Knowledge) ในระหว่างขั้นตอนของการเรียนรู้ โดยเก็บไว้ที่น้ำหนักประสาท (Synaptic Weights)

ในการประยุกต์ใช้โครงข่ายประสาทเทียมในการทำนาย โดยการกำหนดเซตของค่าเป้าหมาย (Target)  $N$  ค่า คือ  $\{y(t_1), y(t_2), y(t_3), \dots, y(t_N)\}$  ณ เวลา  $t_1, t_2, \dots, t_N$  ตามลำดับ ในการพยากรณ์ผลล่วงหน้า คือ  $y(t_{N+1})$  ซึ่งเป็นค่า ณ เวลา  $t_{N+1}$  ในอนาคต การใช้ฟังก์ชันประมาณค่าจึงจำเป็นต้องทำการกำหนด ชนิดของฟังก์ชันถ่ายโอน (Activation Function) โดยปกติแล้วโครงข่ายจะประกอบไปด้วยนิวรอนหลายๆ ตัว และเชื่อมต่อแบบขนานกันหลายๆ ชั้นเรียกว่า Layer แต่ละชั้นมีน้ำหนักประสาท (Weight) ค่าเอนเอียง (Biased) ของชั้นนั้นๆ ชั้นแรกเรียกว่าชั้นนำเข้ามี  $R$  อินพุต ชั้นที่หนึ่งมี  $s^1$  นิวรอนไปเรื่อยๆ จนถึงชั้นสุดท้ายคือชั้นผลลัพธ์ซึ่งมี  $s^N$  นิวรอน โครงข่ายที่ไม่ใช่ชั้นนำเข้าและชั้นผลลัพธ์ เรียกว่าชั้นซ่อน (Hidden Layer) ซึ่งในชั้นซ่อนนั้น ช่วยแปลงข้อมูลจากไม่เชิงเส้นให้กลายเป็นเชิงเส้น (Srikaew, 2009) ค่าผลลัพธ์ในรูปแบบทั่วไปในชั้นซ่อนของโครงข่าย แสดงดังสมการ

$$a_j = f \sum_{i=0}^d w_{ji} x_i - b_j$$

เมื่อพิจารณาจากผลลัพธ์ (Output) ของโครงข่ายประสาทเทียมในรูปแบบทั่วไป

$$o_i = f \sum_{j=1}^m w_{ki} g \left( \sum_{i=0}^d w_{ji} x_i - b_i \right) - b_j$$

โดยกำหนดให้  $i$  แทนจำนวนนิวรอนในชั้นนำเข้าที่  $i = 1, 2, \dots, d$

$j$  แทนจำนวนนิวรอนในชั้นซ่อนที่  $j = 1, 2, \dots, m$

$w_{ji}$  แทนค่าน้ำหนัก ของเส้นเชื่อมจากชั้นนำเข้าของนิวรอนที่  $i$

$w_{kj}$  แทนค่าน้ำหนัก ของเส้นเชื่อมจากชั้นซ่อนของนิวรอนที่  $j$

$b_i$  แทนค่าเอนเอียงของเส้นเชื่อมจากชั้นนำเข้าของนิวรอนที่  $i$

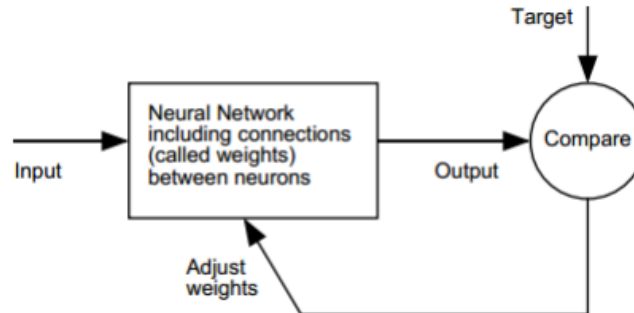
$b_j$  แทนค่าเอนเอียงของเส้นเชื่อมจากชั้นซ่อนของนิวรอนที่  $j$

$g$  แทนฟังก์ชันถ่ายโอนในชั้นซ่อน

$f$  แทนฟังก์ชันถ่ายโอนในชั้นผลลัพธ์



$a_j$  แทนผลลัพธ์ในชั้นซ่อนที่  $j$   
 $o_i$  แทนผลลัพธ์ที่  $i$



รูปที่ 2.8 หลักการทำงานของตัวแบบ ANN (Beale et al., 2013)

#### 2.4.1 การประยุกต์ใช้ Neural network ในงานต่างๆ ได้แก่

ปัจจุบันมีการนำตัวแบบ ANN นำมาใช้งานต่างๆ หลายประเภท เช่น สถาบันการเงินหลายแห่งใช้ ANN ทำนายราคาหุ้นในอนาคต ซึ่งก็สามารถทำได้เนื่องจากข้อมูลราคาหุ้นในอดีตนั้นมีแนวโน้ม (Trend) หรือมีปัจจัยที่พอจะบ่งบอกลักษณะ (Pattern) ของข้อมูล สิ่งเหล่านี้จะช่วยให้การ สร้างตัวแบบมีความแม่นยำมากขึ้น

ดังนั้นสิ่งที่สำคัญและจำเป็นต้องเข้าใจ คือต้องทราบว่าตัวแปรอิสระที่จะนำไปพยากรณ์ตัวแปรตามนั้นต้องมีความสัมพันธ์กันเกี่ยวเนื่องกัน หรือต้องสงสัยว่าเกี่ยวข้องกันทางใดทางหนึ่ง (สัมพันธ์ไปในทางเดียวกัน หรือตรงกันข้าม) ซึ่งความสัมพันธ์ที่ว่ามานี้คงจะไม่ใช้ 100% เป็นลักษณะที่กำกวมบ้าง ไม่ชัดเจนไปเสียทีเดียว (Noisy) ก็ถือเป็นขอบเขตที่ ANN ที่จะทำงานต่อไปโดยการปล่อยให้ตัวแบบเรียนรู้ความสัมพันธ์ระหว่างข้อมูลนำเข้า (Input) และข้อมูลที่ออกมา (Output) หรือเรียกว่า Training สามารถแบ่งออกได้เป็น 2 ประเภท คือ

1. **Supervised training** คือตัวแบบที่มีการเทรน โดยตัวแบบจะ fit กับข้อมูลชุดเรียนรู้ (Training set) โดยจะนำเข้าข้อมูลเป็นแถว ๆ (Record) ที่มีข้อมูลทั้ง Input และ Output ครบถ้วนมักเป็นข้อมูลเก่า เช่น ข้อมูลราคาหุ้นในอดีต ข้อมูลส่วนตัวของลูกค้าบัตรเครดิต ข้อมูลปฏิกิริยาที่หุ่นยนต์ตอบสนองเมื่อเจอกับสิ่งเร้าที่มากกระทบ เป็นต้น จากนั้นทำการเทรนแล้วนำตัวแบบที่ได้ ไปทดสอบประสิทธิภาพกับข้อมูลชุดอื่นที่ตัวแบบไม่เคยเจอมาก่อนเรียกว่า ข้อมูลชุดทดสอบ (Test set) ตัวแบบ ANN ส่วนมากจัดอยู่ในหมวดนี้ ได้แก่ Feedforward network , Multilayer perceptron (MLP), Nonlinear autoregressive (NAR) network ,

Nonlinear autoregressive with exogenous input (NARX) network, Time delay network, Recurrent network, Radial Basis network, Probabilistic network , Generalized Regression network

**2. Unsupervised training** คือตัวแบบที่ไม่มี Output มีเฉพาะ Input หลักการทำงานของตัวแบบประเภทนี้ คือการจัดกลุ่ม (Clustering) ข้อมูลในแต่ละ record ที่มีความคล้ายคลึงกัน อยู่เป็นกลุ่มเดียวกัน และพยายามให้แต่ละกลุ่มมีความแตกต่างกันมากที่สุด ตัวแบบ neural network ที่จัดอยู่ในหมวดนี้ ได้แก่ Self-Organizing Feature Map (SOFM, หรือ Kohonen) networks งานวิจัยชิ้นนี้สนใจเฉพาะตัวแบบประเภท Supervised training โดยเฉพาะตัวแบบ NAR เนื่องจากเป็นตัวแบบที่ใช้ในงานนี้ และจะกล่าวลงในรายละเอียดต่อไป ส่วนตัวแบบอื่น ๆ ทั้ง Supervised และ Unsupervised สามารถค้นคว้าศึกษาเพิ่มเติมได้ในตำราวารสารวิชาการอื่นๆ กล่าวโดยสรุปการใช้งาน ANN แบ่งออกเป็น 4 กลุ่มคือ

### 1. Fitting a Function

เป็นการประมาณค่าฟังก์ชันลักษณะการทำงานคล้ายกับเทคนิค Regression ในการวิเคราะห์ทางสถิติ

### 2. Recognizing Patterns

เป็นการพยากรณ์ข้อมูลที่เป็นกลุ่ม จัดกลุ่มข้อมูลที่มีการเรียนรู้ (Supervised learning) หรือเรียกว่า Pattern recognition ลักษณะการทำงานคล้ายกับเทคนิค Logistic regression, Discriminant analysis ในการวิเคราะห์ทางสถิติ

### 3. Clustering Data

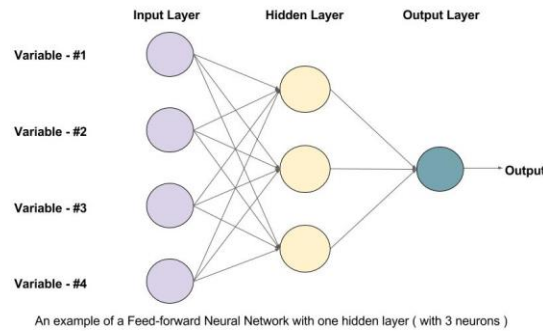
จัดกลุ่มข้อมูลที่มีความคล้ายกัน โดยใช้คำนวณระยะห่างระหว่างจุดเป็นตัวแบ่ง เป็นการจัดกลุ่มข้อมูลที่ไม่มีการเรียนรู้ (Unsupervised learning) เช่น การแบ่งกลุ่มลูกค้าที่เข้ามาซื้อสินค้าหรือเข้ามาใช้บริการ (Market segmentation) การทำเหมืองข้อมูล (Data mining) การจัดกลุ่มยีน Bioinformatic analysis ลักษณะการทำงานคล้ายกับเทคนิค Clustering ในการวิเคราะห์ทางสถิติ

### 4. Time series forecasting

ตัวแบบ ANN สามารถประยุกต์นำมาใช้พยากรณ์ข้อมูลอนุกรมเวลาได้ คล้ายกับ Function fitting แต่ต้องปรับตัวแปรอิสระ โดยทำให้ข้อมูลในอดีตกลายเป็นตัวแปรอิสระ ตามช่วงเวลาต่างๆ หรือเรียกว่า Lag (กล่าวถึงรายละเอียดในหัว 2.4.5) ลักษณะการทำงานคล้ายกับเทคนิค ARIMA ในการวิเคราะห์ทางสถิติ

### 2.4.2 เครือข่าย Feedforward (Feedforward neural network)

นิยมเรียกอีกอย่างว่า Multilayer perceptron (MLP) เป็นโครงสร้างการทำงานของ ANN เป็นลักษณะเป็นไปตามรูปข้างล่าง



รูปที่ 2.9 โครงสร้างของ Feedforward network (Vikas Gupta,2017)

ค่า Inputs จากแต่ละคนตัวแปร จะถูกส่งผ่าน Input neurons ไปยัง Hidden neurons แบบทั่วถึงกันหมด (Fully connected) จากนั้นก็เข้าสู่ Output neurons ในขั้นสุดท้ายตัวแบบ Multilayer Perceptron (MLP) เป็นตัวแบบพยากรณ์ (Predictive model) ที่นิยมใช้พยากรณ์ข้อมูล แบ่งโครงสร้างออกเป็น 3 ชั้น คือ

1. **Input layer** คือชั้นที่รับข้อมูลเพื่อสร้างตัวแบบ จะมี 1-layer ประกอบด้วย Input neurons จำนวน Input neurons ขึ้นอยู่กับจำนวนตัวแปรอิสระที่เลือกใช้ในการสร้างตัวแบบ อาจมี 1 ตัวหรือหลายตัวก็ได้
2. **Hidden layer** คือชั้นถัดจาก Input layer ปกติจะมีชั้นเดียวเนื่องจากมีความสามารถเพียงพอต่อการพยากรณ์แทบทุกปัญหา (Zhang, 2003, Khashei and Bijari, 2011) ภายใน Hidden layer จะประกอบด้วย Hidden neurons เป็นรูปวงกลมดังภาพ จะมีจำนวนที่ไม่แน่นอน ต้องทำการทดลองหาค่า จำนวน Hidden neurons ที่เหมาะสม จัดเป็นหนึ่งในพารามิเตอร์ที่ผู้วิจัยต้องกำหนดขอบเขตแทนค่าเอง
3. **Output layer** คือขั้นสุดท้ายของ Network เป็นชั้นที่ได้ค่าพยากรณ์ที่นำไปใช้จริง จะมีเพียงชั้นเดียว ประกอบด้วย Output neurons

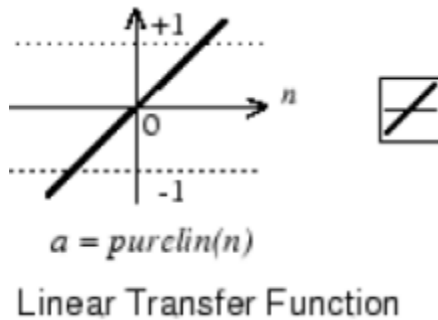
### 2.4.3 ฟังก์ชัน Activation (Activation function)

Activation function หรือ Transfer function, Threshold function จะอยู่ 2 ที่ คือ

1. อยู่ใน Hidden neurons ทำการแปลงค่าข้อมูลที่นำเข้ามาจาก Input neurons แปลงให้เป็นค่าส่งออก ไปยัง Output neuron
2. อยู่ใน Output neuron ทำการแปลงค่าข้อมูลที่นำเข้ามาจาก Hidden neurons แปลงให้เป็นค่าส่งออกที่สามารถนำไปใช้ ซึ่งก็คือค่าพยากรณ์นั่นเอง โดยทั่วไป Activation นั้นมีหลายฟังก์ชันที่ใช้กัน แต่ที่นิยมใช้ที่สุด ได้แก่

#### 2.4.3.1 Linear function หรือ Identity function

$f(x) = x$  เป็นฟังก์ชันที่อยู่ใน Output neuron จะอยู่ในชั้นสุดท้ายของเครือข่าย ใช้ในการพยากรณ์ที่เป็นตัวเลข การประมาณค่า การพยากรณ์อนุกรมเวลา และนำมาใช้ในงานวิจัยชิ้นนี้ด้วย



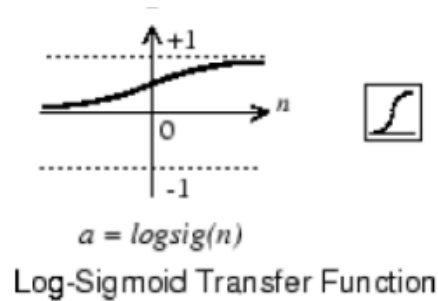
รูปที่ 2.10 Linear Transfer Function (Beale et al., 2013)

#### 2.4.3.2 Sigmoid logistics function

$$f(x) = \frac{1}{1 + e^{-x}}$$

ทำการแปลงสถานะของข้อมูลนำเข้าให้อยู่ในรูปแบบการจัดแบ่งเชิงกลุ่ม

เช่น 0 หรือ 1, ใช่ หรือ ไม่ใช่, ON หรือ OFF เป็นต้น ใช้ในทุกๆ Neurons ใน Hidden layer ยกเว้น Output neuron ค่าที่ได้หลังจากผ่านฟังก์ชันนี้จะอยู่ระหว่าง 0 ถึง 1



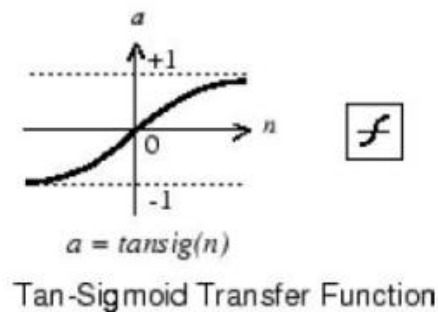
รูปที่ 2.11 Log-Sigmoid Transfer Function (Beale et al., 2013)

#### 2.4.3.3 Tan-sigmoid function หรือ Hyperbolic tangent function

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

ทำคล้าย Sigmoid function ใช้ที่เดียวกัน ใช้แทนกันได้ แต่ค่าที่ได้

หลังจากผ่านฟังก์ชันนี้ คือ -1 ถึง 1



รูปที่ 2.12 Tan-Sigmoid Transfer Function (Beale et al., 2013)

ทั้ง Sigmoid และ Tan-sigmoid สามารถใช้ใน Output neurons ในปัญหาการพยากรณ์เชิงกลุ่ม เนื่องจากค่าที่ได้จะอยู่ช่วง ที่มีขอบเขตจำกัด และค่านี้เองจะแปลงกลับเป็นค่าพยากรณ์ที่ต้องการ เช่น ค่าพยากรณ์=0.345 ก็จะแปลงกลับเป็นกลุ่มตามที่ทำการใส่รหัสไว้ในข้างต้น แต่จะไม่เหมาะที่จะนำมาใช้พยากรณ์เชิงตัวเลข สำหรับปัญหา Regression และ Time series เนื่องจากใช้ Linear function จะดีกว่าเพราะว่าไม่ต้องแปลงกลับไปมางานวิจัยชิ้นนี้เลือกใช้ ฟังก์ชัน Hyperbolic tangent function เนื่องจากเป็นที่นิยมใช้ทั่วไป และให้ค่าพยากรณ์ที่มีความแม่นยำ (BekirKarlik and A. Vehbi Olgac, 2010)

#### 2.4.4 การแปลงข้อมูลก่อนและหลังกระบวนการเรียนรู้ (Data preprocessing and postprocessing)

การแปลงข้อมูลก่อนนำข้อมูลสร้างตัวแบบ สำหรับ ANN เป็นสิ่งที่ควรทำเนื่องจากจะทำให้การสร้างตัวแบบใช้เวลาอันน้อยลงหรือมันจะทำให้ได้คำตอบที่มีประสิทธิภาพมากยิ่งขึ้นการแปลงข้อมูล (Data transformation) ให้อยู่ในช่วงที่เหมาะสม ได้แก่

#### 2.4.4.1 การใส่ Logarithm ข้อมูลดิบ

ส่วนใหญ่นิยม ใส่ Logarithm ฐาน 10 หรือ ฐานธรรมชาติ ลักษณะข้อมูลที่น่าจะเหมาะกับการใส่ Logarithm คือข้อมูลที่มีการเพิ่มขึ้นแบบก้าวกระโดด เป็นแบบ Exponential จัดได้ว่าข้อมูลมีความสัมพันธ์แบบผลคูณ (Multiplicative relations) การใส่ Logarithm เข้าไปเปรียบเสมือนการปรับความสัมพันธ์เป็นแบบผลบวก (Additive relations)

#### 2.4.4.2 การแปลงข้อมูลเชิงเส้นตรง (Linear transformation)

การแปลงข้อมูลเชิงเส้นหรือเรียกว่า Normalized คือการนำข้อมูลดิบ มาจำกัดช่วง เช่น อยู่ในช่วง 0 ถึง 1 และ -1 ถึง 1

$$\text{ช่วง } 0 \text{ ถึง } 1 : Y_t = \frac{Y_t - Y_{\min}}{Y_{\max} - Y_{\min}}$$

$$\text{ช่วง } -1 \text{ ถึง } 1 : Y_t = \frac{2(Y_t - Y_{\min})}{Y_{\max} - Y_{\min}} - 1$$

โดยที่

$Y_t$  แทนค่าหลังจากการแปลง

$Y_t$  แทนข้อมูลดิบ

$Y_{\max}$  แทนข้อมูลดิบที่มีค่ามากที่สุด

$Y_{\min}$  แทนข้อมูลดิบที่มีค่าน้อยที่สุด

#### 2.4.4.3 การแปลงข้อมูลแบบ Standardization

แปลงข้อมูลโดยใช้คุณสมบัติทางสถิติ ที่มีค่าเฉลี่ยและค่า Standard deviation ของข้อมูลดิบมาเกี่ยวข้องด้วย

$$Y_t = \frac{Y_t - \bar{Y}}{\sigma_Y}$$

โดยที่

$Y_t$  คือ ค่าหลังจากการแปลง

$\bar{Y}$  คือ ข้อมูลดิบ

$\sigma_Y$  คือ ค่า Standard deviation ของข้อมูลดิบ

#### 2.4.4.4 การแปลงข้อมูลแบบ Standardization

แปลงข้อมูลโดยใช้คุณสมบัติทางสถิติ ที่มีค่าเฉลี่ยและค่า Standard deviation ของข้อมูลดิบมาเกี่ยวข้องด้วย

#### 2.4.4.5 การ Differencing

การนำข้อมูลดิบตัวติดกันมาลบกัน ดังเช่นการ Differencing ในการ fit ตัวแบบ ARIMA เป็นการทำให้ข้อมูลให้เรียบขึ้น กำจัดแนวโน้ม (Trends) และ ลักษณะฤดูกาล (Seasonality) ออกไป การแปลงข้อมูลให้อยู่ในช่วงที่เหมาะสมนั้น ถือเป็นเรื่องจำเป็นเนื่องจากตัวแบบ ANN ไม่สามารถทำการพยากรณ์ได้เมื่อข้อมูลมี Scale ที่หลากหลาย มีค่าขอบเขตของข้อมูลในตัวแปรที่นำเข้าต่างกันมาก งานวิจัยชิ้นนี้ เลือกทำการแปลงข้อมูลให้อยู่ในช่วง -1 ถึง 1 เพื่อให้ค่าความห่างของข้อมูล สอดคล้องกับ Activation ฟังก์ชันที่เลือกใช้ คือ Hyperbolic tangent sigmoid function ภายหลังจากเทรน เรียบร้อยแล้ว Output ที่ได้จะต้องทำการแปลงกลับให้อยู่ในรูปแบบเดิม ก่อนทำ Preprocessing เรียกกระบวนการนี้ว่า Postprocessing ค่าที่ได้นี้จะเป็ค่าพยากรณ์ที่สามารถนำไปใช้จริง

#### 2.4.5 ตัวแบบ ANN สำหรับการพยากรณ์อนุกรมเวลา (ANN for time series forecasting)

ตัวแบบ ANN สามารถประยุกต์ใช้การพยากรณ์อนุกรมเวลาได้ ส่วนใหญ่นิยมใช้พยากรณ์ข้อมูล

อนุกรมเวลาชุดเดียว ตัวแปรเดียว ๆ เช่นพยากรณ์ยอดขายสินค้า พยากรณ์ความต้องการใช้พลังงาน ทั้งประเทศ พยากรณ์ราคาหุ้น ราคาทองคำ เป็นต้น โดยที่ใช้ตัวแปรที่ต้องการพยากรณ์นั้นเพียงตัวแปรเดียว ไม่ใช่ตัวแปรอิสระอื่น ๆ เพื่อมาพยากรณ์ ซึ่งเรียกว่า Univariate time series การพยากรณ์ข้อมูลอนุกรมเวลา มีเป้าหมาย คือการพยากรณ์ข้อมูลล่วงหน้า โดยใช้ข้อมูลเก่าในอดีตของข้อมูลชุดนั้นเพียงอย่างเดียว มาพยากรณ์ค่าในอนาคต โดยไม่ใช้ตัวแปรอื่นๆ มาวิเคราะห์ โดยส่วนมากการพยากรณ์อนุกรมเวลา จะเกี่ยวข้องกับตัวเลข ปัญหาเรื่องนี้จึงคล้ายคลึงกับปัญหา Regression การปรับเปลี่ยนปัญหาการพยากรณ์แบบ Regression ให้เป็นการพยากรณ์อนุกรมเวลาที่สำคัญ คือการเปลี่ยนตัวแปรอิสระ ดังเช่นตัวอย่างต้องการพยากรณ์ยอดขายสินค้าชนิดหนึ่ง จากเดิมที่มีตัวแปรอิสระหลาย ๆ ตัว แล้วนำมาเทรนตัวแบบ แต่สำหรับอนุกรมเวลาจะใช้ข้อมูลตัวแปรยอดขายสินค้า เพียงตัวแปรเดียว โดยนำตัวแปรยอดขายมาแบ่งเป็นตัวแปรย่อย ณ เวลาย้อนหลัง ใช้เป็นตัวแปรอิสระในการพยากรณ์ รูปแบบการแปลงตัวแปรที่จะพยากรณ์เป็นตัวแปรอิสระนั้น เป็นไป ตามรูปที่ 2.13

Input1	Input2	Input3	(Output) Predicted value	Real observation
Y1	Y2	Y3	$\hat{Y}_4$	Y4
Y2	Y3	Y4	$\hat{Y}_5$	Y5
Y3	Y4	Y5	$\hat{Y}_6$	Y6
Y4	Y5	Y6	$\hat{Y}_7$	Y7
....	Training set	....	....	....
....	....	....	....	....
....	....	....	....	....
Y101	Y102	Y103	$\hat{Y}_{104}$	Y104
Y105	Y106	Y107	$\hat{Y}_{107}$	Y107
....	Testing set	....	....	....
....	....	....	....	....
Y130	Y131	Y132	Y133	Y134

รูปที่ 2.13 แผนผังตัวแปรอิสระและตัวแปรตามสำหรับตัวแบบ ANN

ในการพยากรณ์ข้อมูลอนุกรมเวลา (ภัทร วรภู,2556)

เช่นต้องการพยากรณ์ราคาหุ้นที่เวลา  $Y_{t+1}$  โดยรู้ข้อมูลยอดขายตั้งแต่  $Y_1, Y_2, \dots, Y_t$  จากนั้นกำหนดจำนวนจุดเวลาย้อนหลังมาทำเป็นตัวแปร (Lag variable) เช่นกำหนดจำนวน Lag เป็น 1 นั่นคือทำการพยากรณ์  $Y_{t+1}$  ด้วย  $Y_t$  หรือกำหนดจำนวน Lag เป็น 2 นั่นคือทำการพยากรณ์  $Y_{t+1}$  ด้วย  $Y_t$  และ  $Y_{t-1}$  หรือกำหนดจำนวน Lag เป็น 3 นั่นคือทำการพยากรณ์  $Y_{t+1}$  ด้วย  $Y_t$  และ  $Y_{t-1}$  และ  $Y_{t-2}$  เมื่อจำนวน Lag เพิ่มขึ้น ก็ทำเช่นเดียวกันนี้ไปเรื่อย ๆ ตารางข้างบน แสดงการพยากรณ์ข้อมูล เมื่อกำหนดจำนวน Lag เป็น 3 จะมี input 3 ตัว นำเข้าเพื่อทำการเทรน และจะได้ค่าพยากรณ์ 1 ตัว ทุกๆ ชุดที่ใส่ Input เข้าไป



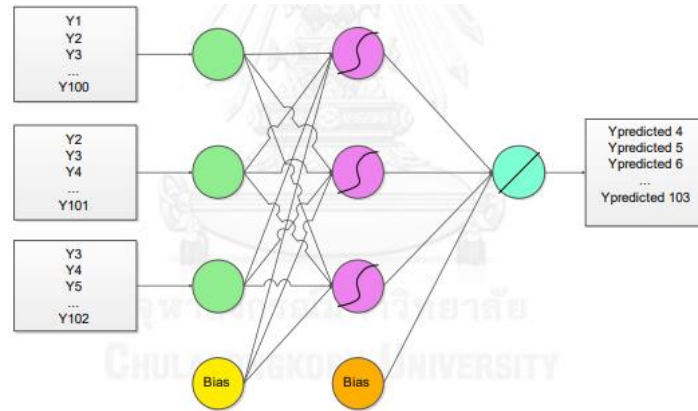
จำนวน Lag ที่กำหนดนั้น ปัจจุบันยังไม่มีกฎเกณฑ์ตายตัวว่าจะต้องเป็นเท่าไร จึงจะได้คำตอบที่ดีที่สุด ขึ้นอยู่กับ ความซับซ้อนของข้อมูล ความแปรผันของข้อมูล ลักษณะเฉพาะของข้อมูล เช่น มีแนวโน้ม มีปัจจัยเชิงฤดูกาล มีการส่ายไปมาไร้รูปแบบ(Noise) จำเป็นต้องทำการทดลอง เริ่มจากกำหนดจำนวน Lag เป็น 1 เพิ่มไปเรื่อยๆ ทีละ 1 จนกระทั่ง Lag มีค่าเป็น 2,3,4,5,... แต่ส่วนใหญ่มักมีค่าไม่มาก อยู่ใน 1-10 เนื่องจากยิ่งกำหนดจำนวน Lag ยิ่งมาก ความสัมพันธ์ระหว่างข้อมูล ณ จุดเวลา Lag (Autocorrelation) นั้นจะน้อย และจะมีความสัมพันธ์กันเองระหว่าง Lag อื่นๆ มากยิ่ง ทำให้ใช้เวลาในการเทรนนาน และทำให้ได้ผลลัพธ์ที่แย่ สำหรับข้อมูลอนุกรมอนุกรมเวลา ที่มีปัจจัยเชิงฤดูกาล เช่น ข้อมูลปริมาณการใช้ไฟฟ้ารายเดือน จะมีรอบของฤดูกาลประมาณ 11,12,13 คือ ปริมาณการใช้ไฟฟ้ามากที่สุดในเดือน เมษายน ปี 2554 ในปี 2555 มากที่สุดในเดือน เมษายน และ ในปี 2556 มากที่สุดในเดือน พฤษภาคม การกำหนดจำนวน Lag ให้คลุมรอบของปัจจัยฤดูกาลนี้ จะมีโอกาสเพิ่มความแม่นยำในการพยากรณ์ให้สูงขึ้น จากตารางข้างบน จำนวนค่าพยากรณ์ มีจำนวนน้อยกว่าข้อมูลจริง ดังความสัมพันธ์

$$\text{จำนวนค่าพยากรณ์} = \text{จำนวนข้อมูลจริง} - \text{จำนวน Lag ที่เลือกใช้}$$

ดังนั้น ถ้าข้อมูลที่ใช้มีจำนวนน้อยการกำหนดจำนวน Lag ยิ่งมาก ยิ่งทำให้จำนวนข้อมูลลดลง ถ้าเรามีข้อมูลอนุกรมเวลา 130 ตัว กำหนดจำนวน Lag 20 ตัว นั้นจะทำให้จำนวนข้อมูลลดลงเหลือ 110 ตัว

#### 2.4.5.1 โครงสร้างตัวแบบ ANN สำหรับการพยากรณ์อนุกรมเวลา (ANN for time series architecture)

รูปแบบเครือข่ายของตัวแบบ ANN ที่ใช้ในการพยากรณ์อนุกรมเวลาจะคล้ายกับตัวแบบ MLP แต่จะมีการปรับเปลี่ยนตัวแปรนำเข้า รูปที่ 2.14



รูปที่ 2.14 แผนผังของตัวแบบ Feedforward network ในการพยากรณ์ข้อมูลอนุกรมเวลา  
(ภัทร วรภู, 2556)

สอดคล้องกับตารางด้านบน เส้นที่แสดงเป็นตัว S ใน Hidden neurons หมายถึง Hyperbolic tangent transfer function ส่วนเส้นตรงแนวทแยงใน Output neuron หมายถึง Linear transfer function ซึ่งทั้งสองเป็นฟังก์ชัน Activation ที่เลือกใช้ในงานวิจัยชิ้นนี้ สามารถเขียนนี้เป็นความสัมพันธ์ในรูปแบบสมการดังนี้

$$\hat{y}_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-n}) + \varepsilon_t$$

โดยที่

$\hat{y}_t$  แทนค่าพยากรณ์เวลา  $t$

$y_{t-1}$  แทนค่าจริงที่เวลา  $t-1$

$y_{t-2}$  แทนค่าจริงที่เวลา  $t-2$

$y_{t-n}$  แทนค่าจริงที่เวลา  $t-n$  , โดยที่  $n$  คือจำนวน Lag ช่วงห่างเวลาที่

นำเข้าสู่ตัวแบบ

$\varepsilon_t$  แทนค่าเรซิดิวลที่เวลา  $t$

สามารถเรียกรูปแบบ Neural network ในลักษณะนี้ว่า Nonlinear Autoregressive หรือ NAR เนื่องจากนำค่าในอดีตมาพยากรณ์ค่าในอนาคต คล้ายกับ ARIMA เพียงแต่เป็นรูปแบบไม่เป็นเชิงเส้นตรงส่วนประกอบอื่น ๆ ของโครงข่าย ทั้งจำนวน Hidden layer, Output neuron, Activation function, Training algorithm จะเหมือนกับตัวแบบ Neural networks สำหรับการพยากรณ์ Regression จำนวนค่าพยากรณ์ล่วงหน้า สามารถแบ่งได้เป็น 2 แบบ ใหญ่ คือ

1.การพยากรณ์ 1 ค่าล่วงหน้า (One step ahead forecasting) ทำการพยากรณ์ล่วงหน้า 1 ค่าล่วงหน้า ถ้าต้องการพยากรณ์ข้อมูลในวันพรุ่งนี้ ต้องมีข้อมูลในวันนี้

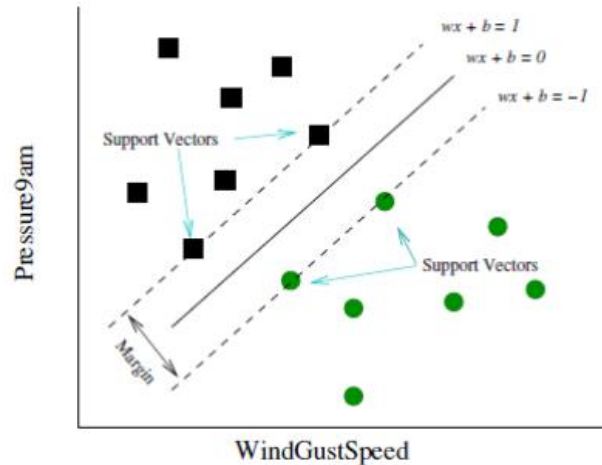
2.การพยากรณ์ล่วงหน้า k ค่าล่วงหน้า (K steps ahead forecasting) ทำการพยากรณ์ล่วงหน้า k ค่าล่วงหน้าเช่นพยากรณ์ 5 หรือ 10 หรือ 50 ค่าล่วงหน้า

การพยากรณ์ล่วงหน้า 1 ค่าจะได้เปรียบ มีความแม่นยำในการพยากรณ์สูงกว่าการพยากรณ์ หลายค่า ล่วงหน้านานวิจัยขึ้นนี้จะพุ่งเป้าเน้นไปที่พยากรณ์ 1 ค่าล่วงหน้าเท่านั้นเนื่องจากต้องการเปรียบเทียบประสิทธิภาพ ตัวแบบพยากรณ์อนุกรมเวลา 6 ตัวแบบ จึงต้องการเน้นไปที่ความแม่นยำที่สูงสุดของแต่ละตัวแบบเป็นหลัก และการพยากรณ์อนุกรมเวลาล่วงหน้ามากจะทำให้ความผิดพลาดในการพยากรณ์ (Cumulative errors) มีค่าสูงมากยิ่งขึ้น (Pai and Lin, 2005)

## 2.5 ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM)

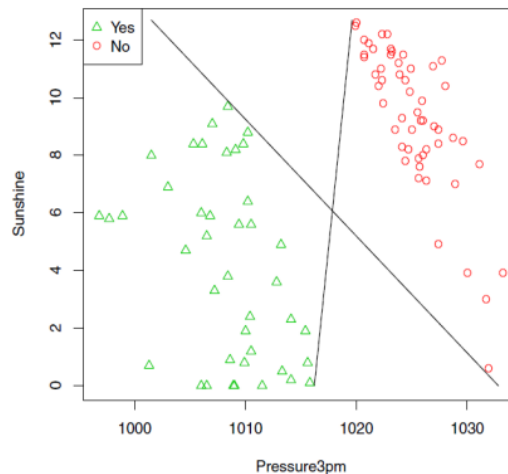
Support vector machine เป็นตัวแบบพยากรณ์ จัดอยู่ในกลุ่ม Machine learning ที่เริ่มเกิดขึ้นเมื่อปี 1995 โดย Vapnik (Vapnik V, 1995) เป็นตัวแบบที่ได้รับความสนใจจากผู้วิจัย เนื่องจากให้ผลการพยากรณ์ที่ดี มีความแม่นยำสูงตัวแบบ SVM มีลักษณะคล้ายกับตัวแบบ ANN ตรงที่สามารถนำไปประยุกต์ใช้พยากรณ์ข้อมูลแบบเดียวกัน ทั้งการพยากรณ์เชิงกลุ่ม (Pattern recognition หรือ Classification) แต่แตกต่างกันตรงที่หลักการ Minimization กล่าวคือ Support vector machine จะทำให้ความเสี่ยงเชิงโครงสร้างต่ำสุด (Structural Risk Minimization : SRM) ขณะที่Neural networks ใช้หลักการลดความเสี่ยงทดลองให้ต่ำสุด (Empirical Risk Minimization: ERM) ERM จะพุ่งเน้นพยายามทำให้กระบวนการเทรน ในข้อมูลชุด Training เกิด Error ต่ำที่สุด นั่นคือ โครงสร้างของตัวแบบ ANN เช่น มีหลายจำนวน Hidden layer หรือ มีจำนวน Hidden neurons หลายตัว อาจทำให้ตัวแบบเรียนรู้ดีเกินไปในข้อมูลชุด Training นั้นอาจทำให้ตัวแบบพยากรณ์ไม่ แม่นยำในข้อมูลชุด Test ได้ เรียกปัญหานี้ว่า Overfitting แต่ SRM จะมีเทอมหรือฟังก์ชันที่เรียกว่า Regularization penalty กำหนดขอบเขตมากอยควบคุม เสมือนเป็นตัวเบรก ไม่ให้ตัวแบบเรียนรู้ดี เกินไป ซึ่งก็เป็นการป้องกันการเกิด ปัญหา Overfitting

หลักการของ SVM มาจากการหาระนาบหรือเส้นแบ่งในกรณีหลายๆ มิติ เรียกว่า Hyperplane ที่แบ่งกลุ่มข้อมูลหรือเรียกว่าแบ่ง Class โดยมีระยะห่างระหว่าง 2 กลุ่มมากที่สุด ระยะห่างนั้น เรียกว่า Margin ดังรูปที่ 2.15



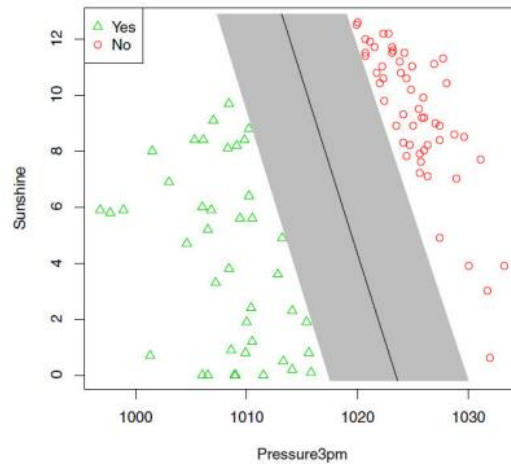
รูปที่ 2.15 เส้นตรงแบ่งกลุ่มข้อมูลที่มี 2 ตัวแปร สำหรับตัวแบบ SVM (Williams, 2011)

ตัวอย่างการหา Hyperplane สำหรับตัวแบบ SVM ที่มี 2 ตัวแปร หรือ 2 มิติ คือมีตัวแปรอิสระและตัวแปรตามอย่างละตัว ตัวแปรตามมีสองกลุ่ม คือ ที่เป็นรูปสี่เหลี่ยมและรูปวงกลม รูปที่ 2.16



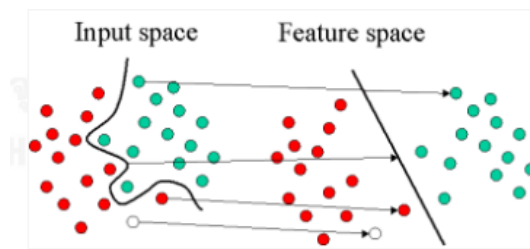
รูปที่ 2.16 เส้นตรงที่ใช้แบ่งกลุ่มข้อมูลที่เป็นไปได้ สำหรับตัวแบบ SVM (Williams, 2011)

จากภาพ 2.16 จะเห็นได้ว่ามีเส้นตรงหลายเส้นที่สามารถแบ่งข้อมูลเป็นกลุ่มข้อมูลเป็น 2 class แต่ SVM จะค้นหาเส้นตรงที่แบ่งกลุ่มข้อมูลโดยให้มีระยะ Margin มากที่สุด ซึ่งนั่นคือมีพื้นที่ (Space) ระหว่างกลุ่มมากที่สุด ดังภาพด้านล่างคือพื้นที่สี่เทา



รูปที่ 2.17 เส้นตรงที่ใช้แบ่งกลุ่มที่ดีที่สุด (Williams, 2011)

เมื่อข้อมูลที่ซับซ้อน ยากที่จะหาเส้นตรงมาแบ่งกลุ่มได้ หรือเมื่อข้อมูลใหญ่ขึ้นมีตัวแปรอิสระหลายตัวหรือข้อมูลมีหลายมิติก็ใช้หลักการเดียวกันในการหา Hyperplane ที่ดีที่สุด แต่ยิ่งแทบเป็นไปไม่ได้เลยที่จะหา Hyperplane แนวตรง กรณีเช่นนี้ ต้องใช้การแปลงข้อมูล โดยใช้ Kernel function หรือเรียกว่า Kernel trick จะแปลงจุดพิกัดให้มีการจัดเรียงใหม่ (Mapping) เพื่อให้สามารถค้นหา Hyperplane แนวเส้นตรงได้ดังภาพด้านล่าง



รูปที่ 2.18 รูปแบบการแปลงจุด เพื่อให้สามารถหาเส้นตรงมาแบ่ง (Williams, 2011)

## 2.6 ซัพพอร์ตเวกเตอร์เรเกรสชัน (Support Vector Regression: SVR)

เป็นเทคนิคที่ใช้วิธีการของซัพพอร์ตเวกเตอร์แมชชีน มาวิเคราะห์ความถดถอยระหว่างอินพุตเวกเตอร์ (Input vector) และตัวแปรเอาต์พุต (Output variables) ซึ่งนำมาใช้กับการพยากรณ์อนุกรมเวลาได้ โดยเปลี่ยนการจำแนกคลาสด้วย SVM เป็นการทำนายค่าด้วย SVR โดยมีเป้าหมายคือต้องการค้นหาความสัมพันธ์เชิงเส้นระหว่างอินพุตเวกเตอร์ในมิติ  $n(X \in R^n)$  และตัวแปรเอาต์พุต  $(Y \in R)$  และเพราะ SVR ดัดแปลงจาก SVM ดังนั้น สมการถดถอย SVR จึงคล้ายคลึงกับสมการไฮเปอร์เพลน (Hyperplane) ของ SVM มีฟังก์ชันเป็นดังนี้

$$y = w\phi(x) + b \quad (2.1)$$

โดยที่  $\phi(x)$  เรียกว่า dimensional feature space ซึ่งเป็น nonlinear mapped จาก input space  $x$  ค่า  $w$  และ  $b$  คือ สัมประสิทธิ์ หาได้จากการ Minimize

$$R(C) = C \frac{1}{N} \sum_{i=1}^N L_\varepsilon(d_i, y_i) + \frac{1}{2} \|w\|^2 \quad (2.2)$$

$$L_\varepsilon(d, y) = \begin{cases} |d - y| - \varepsilon & ; |d - y| \geq \varepsilon \\ 0 & ; \text{others} \end{cases} \quad (2.3)$$

สมการที่ (2.2) เรียกว่า Empirical error(risk) หา  $\varepsilon$  - insensitive loss function จากสมการที่ (2.3) เทอม  $\frac{1}{2} \|w\|^2$  ในสมการ (2.2) คือ Regularization term ซึ่งจะหาค่า Flatness ของฟังก์ชัน ขณะที่  $C > 0$  ทำหน้าที่เป็นตัว Trade-off ระหว่าง Empirical risk และ Flatness ของตัวแบบการ Trade-off ตรงนี้ทำให้ตัวแบบมีความคงทนต่อ Outliers นั้นทำให้ SVM แตกต่างจากเทคนิค Traditional error minimization

การเพิ่มขึ้นของ  $C$  จะทำให้ Empirical risk เพิ่มขึ้นเมื่อเทียบกับ Regularization term การที่จะได้ค่า  $w$  และ  $b$  ในสมการ (2.2) ต้องทำการแปลงเป็น Primal function ในสมการ ที่โดยการเพิ่มตัวแปร Slack ที่เป็นบวกคือ  $\xi_i$  และ  $\xi_i^*$  ซึ่งแสดงระยะห่างระหว่างค่าจริงและค่าขอบเขตภายใต้  $\varepsilon$ -tube จะได้ชุดของสมการที่เป็นลักษณะ Quadratic programming ดังนี้

*Minimize :*

$$R(w, \xi, \xi^*) = \frac{1}{2} ww^T + c^* \left[ \sum_{i=1}^N (\xi + \xi^*) \right] \quad (2.4)$$

$$w\phi(x_i) + b_i - d_i \leq \varepsilon + \xi_i^*, \quad (2.5)$$

$$d_i - w\phi(x_i) - b_i \leq \varepsilon + \xi_i^*, \quad (2.6)$$

$$\xi, \xi^* \geq 0 \quad (2.7)$$

$$i = 1, 2, \dots, N \quad (2.8)$$

ขั้นตอนสุดท้ายจะนำตัวคูณ Lagrange และทำการ Maximizing dual function ในสมการที่ (2.4) ซึ่งจะเปลี่ยนแปลงไปเป็นดังสมการที่ (2.9)

$$R(\alpha_i - a_i^*) = \sum_{i=1}^N d_i (\alpha_i - a_i^*) - \varepsilon \sum_{i=1}^N (\alpha_i - a_i^*) - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - a_i^*) (\alpha_j - a_j^*) K(x_i, x_j) \quad (2.9)$$

$$\sum_{i=1}^N (\alpha_i - a_i^*) = 0 \quad (2.10)$$

$$0 \leq \alpha_i \leq C \quad (2.11)$$

$$0 \leq \alpha_i^* \leq C \quad (2.12)$$

$$i = 1, 2, \dots, N \quad (2.13)$$

ในสมการที่ (2.9) ตัว  $\alpha_i$  และ  $\alpha_i^*$  เรียกว่าตัวคูณ Lagrange โดยที่

$$a_i \cdot a_j^* = 0 \quad (2.14)$$

$$f(x, \alpha, \alpha^*) = \sum_{i=1}^l (\alpha - \alpha_i^*) K(x, x_i) + b \quad (2.15)$$

สมการที่ (2.9) เป็นจุดที่สำคัญในการแก้ปัญหา Nonlinear SVM เนื่องจากมีเทอม  $K(x_i, x_j)$

เรียกว่า Kernel function หรือเรียกว่า “kernel trick” (Vapnik V, 1995) สมการรูปแบบของ Kernel function นี้คือ dot product  $K(x_i, x_j) = \phi(x_i) * \phi(x_j)$  หน้าที่ของ Kernel function คือ การแปลงข้อมูลจาก Input space ไปสู่ High dimensional feature space ใช้ทั้งการพยากรณ์ทั้ง Classification และ Regression เช่นเดียวกัน

## 2.6.2 Kernel function

Kernel function เป็นฟังก์ชันที่มีอยู่ในรูปแบบ Inner product ระหว่างจุดสองจุด ใน Feature space มีหน้าที่หลักคือทำให้ SVM สามารถแก้ปัญหาที่ไม่เป็นเชิงเส้นตรง (Nonlinear) ได้ ซึ่ง Kernel function สามารถมีได้หลายรูปแบบ ดังเช่น

1. Linear kernel มีสูตรดังนี้  $K(x_i, x_j) = \langle x_i, x_j \rangle$

2. Gaussian radial basis kernel มีสูตรดังนี้  $K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$

3. Polynomial kernel มีสูตรดังนี้  $K(x_i, x_j) = (s\langle x_i, x_j \rangle + c)^d$

4. Sigmoid l kernel มีสูตรดังนี้  $K(x_i, x_j) = \tanh(s\langle x_i, x_j \rangle + c)$

5. Convex combination of kernels มีสูตรดังนี้  $K(x_i, x_j) = \lambda_1 k_1(x_i, x_j) + \lambda_2 k_2(x_i, x_j)$

6. Normalization kernels มีสูตรดังนี้  $K(x_i, x_j) = \frac{k'(x_i, x_j)}{k'(x_i, x_i)k'(x_j, x_j)}$

โดยที่  $s, c, d$  และ  $\lambda_i$  เป็น Parameter สำหรับ Kernel ในแต่ละฟังก์ชัน

จนกระทั่งถึงปัจจุบันนี้ ยังไม่มีกฎเกณฑ์ตายตัวในการเลือกชนิดของ Kernel function ที่ดีที่สุดสำหรับทุกชุดข้อมูล (Hong et al., 2010)

งานวิจัยชิ้นนี้เลือก Gaussian kernel function เนื่องจากง่ายต่อการนำไปใช้เนื่องจากมี



พารามิเตอร์ที่ต้องใส่ค่าตัวเดียว คือ  $\sigma$  และมีประสิทธิภาพที่ดีในการแปลง Nonlinear mapping จาก Input space ไปสู่ High dimensional feature space (Chen and Wang, 2007)

ดังนั้นการพยากรณ์ SVR โดยเลือกใช้ Gaussian kernel function จึงมี 3 พารามิเตอร์ที่ต้องใส่ค่า ได้แก่  $\varepsilon, c$  และ  $\sigma$

## 2.7 ขั้นตอนการสร้างตัวแบบผสมที่มีกระบวนการทำต่อเนื่องกัน (Traditional hybrid)

(Zhang, P.G., 2003) ได้เสนอตัวแบบอนุกรมเวลาแบบผสม ซึ่งมีข้อสมมติว่าข้อมูลอนุกรมเวลาที่นำมาวิเคราะห์นั้นมี 2 องค์ประกอบ คือ องค์ประกอบที่เป็นเส้นตรง (linear component) และองค์ประกอบที่ไม่เป็นเส้นตรง (non-linear component) ทั้งนี้ Zhang ได้นำตัวแบบ ARIMA ซึ่งเป็นฟังก์ชันเชิงเส้นตรงในเทอมของค่าพารามิเตอร์มารวมกับตัวแบบไม่เป็นเชิงเส้น ซึ่งใช้สำหรับพยากรณ์ข้อมูลส่วนที่ไม่เป็นฟังก์ชันเชิงเส้นตรงในเทอมของค่าพารามิเตอร์ โดยตัวแบบผสมดังกล่าวมีขั้นตอนการสร้างดังนี้

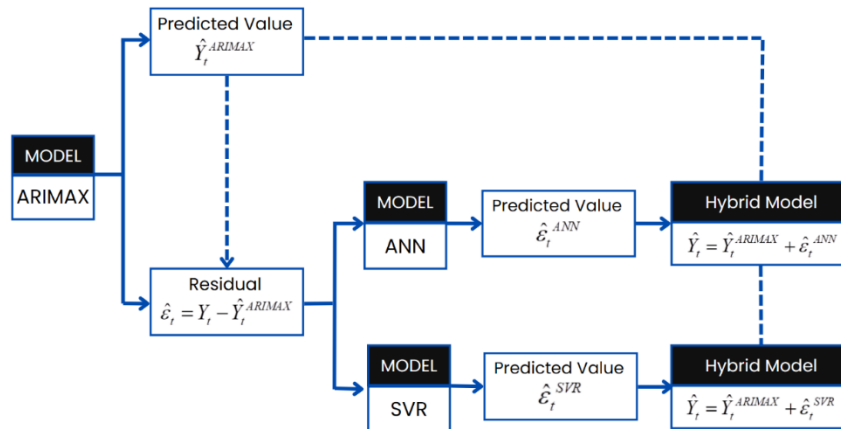
**ขั้นตอนที่1:** นำข้อมูลชุด Train เข้าสู่ตัวแบบ ARIMAX

**ขั้นตอนที่2:** ทำการพยากรณ์ข้อมูลชุด test ตัวแบบ ARIMAX จะได้ค่าพยากรณ์ คือ  $\hat{Y}_t^{ARIMAX}$

**ขั้นตอนที่3 :** คำนวณค่าส่วนเหลือ (residual) จากตัวแบบARIMAX โดยที่  $\varepsilon_t = Y_t - \hat{Y}_t^{ARIMAX}$

**ขั้นตอนที่4 :** นำค่าส่วนเหลือจากตัวแบบARIMAX โดยที่  $\varepsilon_t = Y_t - \hat{Y}_t^{ARIMAX}$  เข้าสู่ตัวแบบ ANN และ SVR เพื่อพยากรณ์ค่าส่วนเหลือจากตัวแบบ ANN ( $\varepsilon_t^{ANN}$ ) และ SVR ( $\varepsilon_t^{SVR}$ ) ตามลำดับ

**ขั้นตอนที่5:** นำค่าส่วนเหลือที่ได้จากจากขั้นตอนที่ 4 และ ค่าพยากรณ์ตัวแบบ ARIMAX ( $\hat{Y}_t^{ARIMAX}$ ) จากขั้นตอนที่2 ไปคำนวณในสมการพยากรณ์ของตัวแบบตัวแบบผสมที่มีกระบวนการทำต่อเนื่องกัน (Traditional hybrid) ได้แก่ ตัวแบบ ARIMAX-ANN และ ARIMAX-SVR สมการพยากรณ์ตัวแบบ ARIMAX-ANN และ ARIMAX-SVR คือ  $\hat{Y}_t = \hat{Y}_t^{ARIMAX} + \varepsilon_t^{ANN}$  และ  $\hat{Y}_t = \hat{Y}_t^{ARIMAX} + \varepsilon_t^{SVR}$  ตามลำดับ



รูปที่ 2.19 ขั้นตอนการสร้างตัวแบบ Traditional hybrid

## 2.8 ตัวแบบที่ทำการพยากรณ์แยกส่วนกัน (Hybrid combined)

(Khairalla, M., Xu-Ning and AL-Jallad, N.T.,2017) หลักการของตัวแบบผสมชนิดนี้ คือ คื้อมองว่า ข้อมูลก่อนนำเข้าสู่ตัวแบบ สามารถแบ่งได้เป็น 2 องค์ประกอบ คือ องค์ประกอบที่เป็นเชิงเส้นตรง (Linear component) และองค์ประกอบที่ไม่เป็นเชิงเส้นตรง (Nonlinear component) เหมือนกับตัวแบบ Traditional hybrid แต่สร้างตัวแบบพยากรณ์แต่ละตัวแยกกันไป แล้วนำพยากรณ์มารวมกัน โดยจัดให้อยู่ในรูปแบบผลรวมเชิงเส้นตรง (Linear combination) แล้วจึงหาค่าน้ำหนัก (Weights) ที่ทำให้ค่าพยากรณ์มีความผิดพลาด (Errors) น้อยที่สุด โดยตัวรวมผสมมีขั้นตอนการสร้างดังนี้

**ขั้นตอนที่1:** นำข้อมูลชุด Train เข้าสู่ตัวแบบ ARIMAX, ANN และ SVR

**ขั้นตอนที่2:** ทำการพยากรณ์ข้อมูลชุด test สำหรับแต่ละตัวแบบ ARIMAX, ANN และ SVR จะได้ค่าพยากรณ์ ตามลำดับดังนี้  $\hat{Y}_{ARIMAX}, \hat{Y}_{ANN}, \hat{Y}_{SVR}$

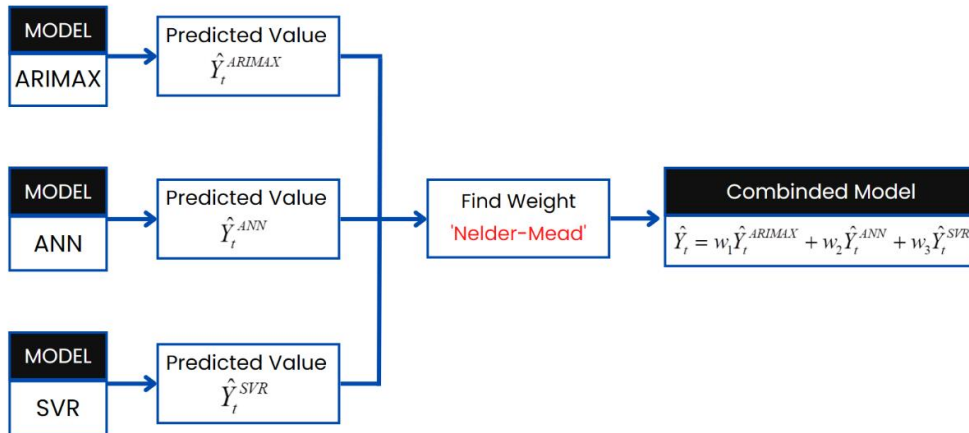
**ขั้นตอนที่3:** ค้นหาค่า Weights ที่ดีที่สุดในข้อมูลชุดทดสอบภายใต้เงื่อนไขค่า MSE มีค่าต่ำสุด ค่า Weights ที่ งานวิจัยชิ้นนี้นำเสนอตัวแบบ Combined (ARIMAX-ANN-SVM) ดังนั้นจะมีค่าน้ำหนัก (weight) จำนวน 3 ค่าคือ ค่า Weigh ของ ARIMAX, ค่า Weight ของ ANN และค่า Weight ของ SVR ทำการค้นหาชุดของค่า Weights ทั้งสามที่ดีที่สุดที่ทำให้ค่า MSE ต่ำที่สุด โดยใช้วิธีการ Nelder-Mead หรือ "Nelder-Mead simplex method" ซึ่งเป็นระเบียบวิธีเชิงตัวเลขที่สามารถค้นหาค่า weights ที่ให้ค่า MSE ที่ต่ำที่สุดได้โดยไม่ต้องคอยไล่ทุกค่า weights ที่เป็นไปได้ ทำให้สามารถหาค่า weights ได้อย่างรวดเร็วและแม่นยำ ซึ่งค่า Weights ที่ได้ในแต่ละตัวแบบต้องรวมกันได้เท่ากับ 1

**ขั้นตอนที่4:** นำค่าพยากรณ์ทั้ง 3 ตัวแบบ มาทำการใส่ค่าweight ซึ่งมีค่าตั้งแต่ 0

ถึง 1 โดย

$$\hat{Y}_t = w_1 \hat{Y}_t^{ARIMAX} + w_2 \hat{Y}_t^{ANN} + w_3 \hat{Y}_t^{SVR}$$

โดยที่  $w_1 + w_2 + w_3 = 1$



รูปที่ 2.20 ขั้นตอนการสร้างตัวแบบ Hybrid Combined

## 2.9 แถบโบลิงเจอร์ (Bollinger Bands: BB)

Bollinger Bands (BB) (Shah & Manubhai, 2015) เป็น Indicator ที่เคลื่อนไหวตามเทรนด์และโมเมนตัมที่เกิดขึ้นภายในตลาด คิดค้นขึ้นในช่วงปี 1980 โดย John Bollinger นักวิเคราะห์ผู้มากประสบการณ์ โดยมีพื้นฐานมาจาก Moving Average Indicator ที่ทำงานร่วมกับการคำนวณโดยใช้ค่าเบี่ยงเบนมาตรฐาน (Standard Deviation: SD) เพื่อใช้วัดค่าความผันผวนของราคา ซึ่งค่าความผันผวนสามารถดูได้จากการเคลื่อนที่ของกรอบ Indicator ว่ากว้างขึ้นหรือแคบลง Bollinger Bands ประกอบด้วยเส้นสามเส้น ได้แก่

1. เส้นขอบบน (Upper Band) คือเส้นค่าเฉลี่ยเคลื่อนที่  $= \bar{X} + m \cdot \sigma^2$
2. เส้นกลาง (Middle Band) คือเส้นค่าเฉลี่ยเคลื่อนที่ (Simple Moving Average)  $= \bar{X}$
3. เส้นขอบล่าง (Lower Band) คือเส้นค่าเฉลี่ยเคลื่อนที่  $= \bar{X} - m \cdot \sigma^2$

โดยที่

$\bar{X}$  คือ เส้นค่าเฉลี่ยเคลื่อนที่ มีสูตรดังนี้

$m$  คือ ค่าสัมประสิทธิ์เบี่ยงเบนมาตรฐานของเส้นค่าเฉลี่ยเคลื่อนที่ มีโดยปกติมีค่าเท่ากับ 2

$\sigma$  คือ ส่วนเบี่ยงเบนมาตรฐานของเส้นค่าเฉลี่ยเคลื่อนที่ มีสูตรดังนี้

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{N}}$$

$\bar{X}$  คือ เส้นค่าเฉลี่ยเคลื่อนที่ มีสูตรคือ

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{N} ; x_i \text{ แทน ราคาวันที่ } i \text{ ในจากทั้งหมด } N \text{ วัน (ปกติใช้ 20 วัน)}$$

### 2.9.1 ความกว้างของ Bandwidth (Bollinger Bands Width: BBW)

(สนธิ อังสนากุล, 2547) ความกว้างของ Bandwidth เป็นการวัดความกว้างของ แถบโบลิ้งเจอร์ แล้วนำไปใช้เป็นตัวบ่งชี้ ใช้อุณหภูมิหรือทิศทางของการเคลื่อนที่ของราคาว่ามั่นคงหรือไม่ เป็นการนำความกว้างของ แถบจากเครื่องมือแถบโบลิ้งเจอร์มาทำเป็นดัชนีบ่งชี้ ถ้ามีค่าน้อยแสดงว่าแถบแคบบ่งชี้ว่าจะมีการเปลี่ยนแปลง ราคาอย่างรุนแรงในไม่ช้า เมื่อมีการเปลี่ยนแปลงราคาที่รุนแรง Bandwidth จะเพิ่มขึ้นหรือชี้ขึ้น แสดงถึงความมั่นคงของแนวโน้มหรือทิศทางของราคาหลักทรัพย์เคลื่อนที่ไป ถ้าราคาตกลงอย่างมั่นคง ถ้าราคาขึ้นก็ขึ้นอย่างมั่นคง จนกระทั่งแถบ Bandwidth แสดงยอดคลื่น อันแสดงถึงแนวโน้มหรือทิศทางราคาได้อ่อนตัวลง

$$Bandwidth = \frac{U_B - L_B}{\mu}$$

โดยที่

$U_B$  แทน แถบด้านบน (Upper band)

$L_B$  แทน แถบด้านล่าง (Lower band)

$\mu$  แทน ค่าเฉลี่ยเคลื่อนที่ (Moving average)

## 2.10 การวิเคราะห์ปัจจัย (Factor Analysis)

การวิเคราะห์ปัจจัย (นงลักษณ์ วิรัชชัย, 2538 ) เป็นวิธีการทางสถิติที่จะช่วยค้นหาลักษณะของตัวแปรหลายๆ ตัวที่สัมพันธ์ซึ่งกันและกันเป็นการลดจำนวนตัวแปรให้น้อยลงเพื่อให้ง่ายต่อการเข้าใจ ทำให้สามารถมองเห็นโครงสร้าง และแบบแผนของตัวแปรในลักษณะของความสัมพันธ์ระหว่างตัวแปร ซึ่งจะช่วยอธิบายความหมายและลักษณะของตัวแปรที่ทำให้สามารถให้จำกัดความของตัวแปรให้ชัดเจนยิ่งขึ้น ช่วยตัดสินใจว่าควรจะศึกษาตัวแปรด้วย ใดบ้างและตัวแปรใดเกี่ยวข้องกับตัวแปรใดการวิเคราะห์องค์ประกอบจะเป็นศูนย์รวมความหลากหลายของการวิเคราะห์ความสัมพันธ์ซึ่งถูกออกแบบมาเพื่อใช้ตรวจสอบความสัมพันธ์ภายในระหว่างตัวแปรต่างๆ ที่สังเกตหรือวัดได้การวิเคราะห์องค์ประกอบมักทำใน 2 ลักษณะคือ

1. เพื่อสำรวจหรือค้นหาตัวแปรแฝงที่ซ่อนอยู่ภายใต้ตัวแปรที่สังเกตหรือวัดได้ เรียกว่า การวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory Factor Analysis)
2. เพื่อพิสูจน์ตรวจสอบหรือยืนยันทฤษฎีที่ผู้อื่นค้นพบ เรียกว่า การวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis)

จุดมุ่งหมายของการวิเคราะห์องค์ประกอบในการวิเคราะห์องค์ประกอบมีหลักใหญ่ๆ ที่ต้องการคำตอบอยู่ 2 ลักษณะ (อุทุมพร ทองอุไทย, 2523 ) คือ “มีความง่ายเชิงสถิติและมีความหมายในเนื้อหา” จุดมุ่งหมายในการวิเคราะห์องค์ประกอบอาจจำแนกได้ ดังนี้

1. ช่วยบรรยายเกี่ยวกับขอบเขตที่ต้องการศึกษา
2. ช่วยตรวจสอบทฤษฎีที่เกี่ยวกับความสัมพันธ์ระหว่างตัวแปร
3. ช่วยสร้างความสัมพันธ์เชิงหน้าที่ระหว่างตัวแปร
4. วิเคราะห์บุคคลหรือวัตถุและจัดให้เป็นประเภทต่าง ๆ
5. วิเคราะห์โครงสร้างองค์ประกอบของตัวแปรที่เป็นเกณฑ์ และระบุตัวแปรที่เป็นประโยชน์ในสมการถดถอย

ขั้นตอนในกระบวนการวิเคราะห์องค์ประกอบและแปลความหมาย แบ่งได้ 4 ขั้นตอน (วิเชียร ไชยบัง, 2544 : 1 - 12) คือ

1. เตรียมข้อมูลและตรวจสอบความสัมพันธ์ระหว่างตัวแปร
2. การสกัดองค์ประกอบ
3. การหมุนแกนองค์ประกอบ
4. การแปลความหมายและกำหนดชื่อองค์ประกอบ

### 2.10.1 เตรียมข้อมูลและตรวจสอบความสัมพันธ์ระหว่างตัวแปร

เตรียมข้อมูลและตรวจสอบความสัมพันธ์ระหว่างตัวแปรข้อมูลที่ใช้ในการวิเคราะห์ องค์ประกอบนั้น จะต้องมีลักษณะเป็นตัวแปรต่อเนื่องหลาย ๆ ตัวแปรที่เก็บจากกลุ่มตัวอย่างหลังจากนั้นทำการตรวจสอบว่าตัวแปรต่างๆ มีความสัมพันธ์กันหรือไม่ ถ้าตัวแปรมีความสัมพันธ์กันมากหรือมีความสัมพันธ์กันอย่างมีนัยสำคัญจะสามารถใช้เทคนิค Factor Analysis ได้ถ้าตัวแปรไม่มีความสัมพันธ์กัน หรือมีความสัมพันธ์กันน้อย ไม่ควรใช้เทคนิค Factor Analysis โดยการตรวจสอบทำได้หลายวิธี ดังนี้

**วิธีที่1 การตรวจสอบโดยใช้ค่าสัมประสิทธิ์สหสัมพันธ์** โดยการสร้างเมทริกซ์แสดงสมประสิทธิ์สหสัมพันธ์ของตัวแปรทุกคู่

- ถ้าค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรคู่ใดมีค่าเข้าใกล้ +1 หรือ -1 แสดงว่าตัวแปรคู่นั้นมีความสัมพันธ์กันมากควรอยู่ใน Factor เดียวกัน
- ถ้าค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรคู่ใดมีค่าใกล้ศูนย์ แสดงว่าตัวแปรคู่นั้นไม่มีความสัมพันธ์กัน หรือสัมพันธ์กันน้อยควรอยู่คนละ Factor
- ถ้าตัวแปรที่ไม่มีความสัมพันธ์กับตัวแปรอื่น ๆ หรือมีความสัมพันธ์กับตัวแปรอื่นๆ ที่เหลือน้อยมาก ควรตัดตัวแปรนั้นออกจากการวิเคราะห์

**วิธีที่2 ใช้สถิติ KMO (Kaiser-Meyer-Olkin Measure of Sampling Adequacy)** ซึ่งเป็นค่าที่ใช้วัดความเหมาะสมของข้อมูลตัวอย่างที่จะนำมาวิเคราะห์โดยเทคนิค Factor Analysis โดยที่

เมื่อ  $r$  คือค่าสัมประสิทธิ์สหสัมพันธ์ซึ่งทำให้ค่า  $0 \leq KMO \leq 1$

- ถ้าค่า KMO มีค่าน้อย (เข้าสู่ศูนย์) แสดงว่าเทคนิค Factor analysis ไม่เหมาะสมกับข้อมูลที่มีอยู่
- ถ้าค่า KMO มีค่ามาก (เข้าสู่หนึ่ง) แสดงว่าเทคนิค Factor analysis เหมาะสมกับข้อมูลที่มีอยู่
- โดยทั่วไปถ้าค่า KMO < 0.5 จะถือว่า ข้อมูลที่มีอยู่ไม่เหมาะสมที่จะใช้เทคนิค Factor Analysis

**วิธีที่3 Bartlett' s Test of Sphericity** เป็นการทดสอบค่าไค-สแควร์ (Chi-square)

ของดีเทอร์มิแนนท์ (Determinant) ของเมทริกซ์สหสัมพันธ์ มีการทดสอบสมมติฐานว่าเมทริกซ์สหสัมพันธ์นั้น เป็นเมทริกซ์เอกลักษณ์ (Identity matrix) หรือไม่จากสมมติฐาน

$H_0$  : ตัวแปรต่างๆ ไม่มีความสัมพันธ์กัน

$H_1$  : ตัวแปรต่างๆ มีความสัมพันธ์กัน

ดังนั้นถ้ายอมรับ  $H_0$  แสดงว่าตัวแปรไม่มีความสัมพันธ์กัน จึงไม่ควรใช้เทคนิค Factor analysis แต่ถ้าปฏิเสธ  $H_0$  (หรือยอมรับ  $H_1$ ) นั่นคือตัวแปรมีความสัมพันธ์กันจึงสามารถใช้เทคนิค Factor analysis ได้

### 2.10.2 การสกัดองค์ประกอบ (Factor extraction)

มีจุดมุ่งหมาย คือการหาจำนวน Factor ที่สามารถใช้แทนตัวแปรทั้งหมดทุกตัวได้หรือเป็นการดึงรายละเอียดจากตัวแปรมาไว้ใน Factor วิธีการสกัดองค์ประกอบมีหลายวิธี (นงลักษณ์ วิรัชชัย, 2538) เช่น

2.10.2.1 Principal Component Analysis หรือ PCA เป็นเทคนิคที่มีวัตถุประสงค์ที่จะนำรายละเอียดของตัวแปรที่มีจำนวนตัวแปรหลายๆ ไว้ในองค์ประกอบที่มีเพียงไม่กี่ตัว โดยจะพิจารณาจากรายละเอียดทั้งหมดจากแต่ละตัวแปรในการ ในการวิเคราะห์ PCA จะสร้างการเชื่อมรวมกันเชิงเส้น (Linear combination)

2.10.2.2 Principal axis factors เป็นการวิเคราะห์องค์ประกอบร่วมแบบหนึ่งที่ใช้หลักการแบบเดียวกับ PCA ข้อแตกต่างคือ Principal Axis Factors มิได้ใช้ค่าการร่วม (Communality) ของตัวแปรเป็น 1.0 เหมือนใน PCA

2.10.2.3 Least Squares Method เป็นการสกัดองค์ประกอบสำหรับการวิเคราะห์องค์ประกอบร่วมแบบหนึ่ง ประกอบด้วยวิธีการแตกต่างกัน 3 แบบ คือ วิธีการถ่วงน้ำหนักน้อยที่สุดไม่ถ่วงน้ำหนัก (Unweighted Least Squares Method) วิธีการถ่วงน้ำหนักน้อยที่สุดทั่วไป (Generalized Least Squares Method) และวิธีพิเศษเหลือ้น้อยที่สุด (Minimum Residuals Method)

### 2.10.3 การหมุนแกนองค์ประกอบ (Factor Rotation)

กรณีที่ค่า Factor Loading มีค่ากลาง ๆ ทำให้ไม่สามารถจัดตัวแปรว่าควรอยู่ใน Factor ใด นั้นจะต้องทำการหมุนแกน ดังนั้นวัตถุประสงค์ของการหมุนแกนปัจจัยคือเพื่อให้ค่า Factor Loading ของตัวแปรค่ามากขึ้น หรือลดลงจนกระทั่งทำให้ทราบว่าตัวแปรนั้นควรอยู่ใน Factor ใด หรือไม่ควรอยู่ใน Factor ใด วิธีการหมุนแกนมี 2 วิธีคือ Orthogonal และ Oblique (สายพิน ศิริสุวรรณรัตน์ , 2540 : 30)

2.9.3.1 Orthogonal Rotation เป็นการหมุนแกนไปแล้วยังคงทำให้ Factor ตั้งฉากกันหรือเป็นอิสระกัน แต่ทำให้ค่า Factor Loading เพิ่มขึ้น หรือลดลง

2.9.3.2 Oblique Rotation เป็นการหมุนแกนไปในลักษณะที่ Factor ไม่ตั้งฉากกัน หรือ Factor ไม่เป็นอิสระกัน แต่ทำให้ค่า Factor loading เพิ่มขึ้นหรือลดลง

#### 2.10.4 การแปลความหมายและกำหนดชื่อองค์ประกอบ

การพิจารณาจำนวนองค์ประกอบจะพิจารณาจากองค์ประกอบที่มีค่า Eigenvalue มากกว่า 1 การพิจารณาวาดัแปรโตบรรจุอยู่ในองค์ประกอบใดให้พิจารณาที่ค่า Loading โดยพิจารณาเลือกค่าน้ำหนักองค์ประกอบ โดย Hair (1995 : 385) ได้เสนอความสัมพันธ์ระหว่างค่า loading ที่มีนัยสำคัญทางสถิติที่ระดับ 0.05 ต่อจำนวนตัวอย่างดังนี้

**ตารางที่ 2.1** ความสัมพันธ์ระหว่างค่า loading ที่มีนัยสำคัญทางสถิติที่ระดับ 0.05 ต่อจำนวนตัวอย่างดังนี้

Factor loading	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75
จำนวนตัวอย่าง	350	250	200	150	120	100	85	70	60	50

ทำการตั้งชื่อองค์ประกอบให้สอดคล้องกับตัวแปร การตั้งชื่อให้แต่ละองค์ประกอบ มีกฎในการตั้งชื่อ ชื่อขององค์ประกอบควรจะสั้น อาจตั้งชื่อเพียง 1- 2 คำ มีความหมายสอดคล้องกันระหว่างตัวแปรที่อยู่ในองค์ประกอบ

#### 2.11 วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

งานวิจัย (Zhang, 2003) ได้ทำการพยากรณ์ข้อมูลอนุกรมเวลาโดยใช้ตัวแบบ ARIMAANN โดยสมมติเงื่อนไขความสัมพันธ์ของการHybrid เป็นผลบวก เปรียบเทียบการพยากรณ์กับตัวแบบเดี่ยวทั้ง ARIMA และANN ทำการพยากรณ์ล่วงหน้า 1 ค่า (One step ahead forecasting) สำหรับ ตัวแบบ ANN ได้ใช้ตัวแบบในลักษณะ Feedforward network มี 1 Hidden layer 1 Output neuron ใช้ Logistic function เป็น Activation function ใช้เทคนิคการเรียนรู้ generalized reduced gradient (GRG2) สำหรับข้อมูลที่ใช้มีจำนวน 3 ชุด คือ ข้อมูลจำนวนจุดดับบนดวงอาทิตย์รายปี(Sunspot),ข้อมูลจำนวนแมวน้ำที่ถูกจับได้ในประเทศแคนาดา (Canadian lynx) และ ข้อมูลอัตราแลกเปลี่ยนเงินสกุลปอนด์อังกฤษต่อเงินสกุลดอลลาร์สหรัฐ ผลการทดลองพบว่าตัวแบบ ARIMAANN มีประสิทธิภาพดีกว่าตัวแบบ ARIMA และANN

งานวิจัย (Ömer Faruk, 2010)งานวิจัยชิ้นนี้ได้ศึกษาประสิทธิภาพของตัวแบบ ARIMAANN โดยสมมติเงื่อนไขความสัมพันธ์ของการHybrid เป็นผลบวก เทียบกับตัวแบบ ARIMA และANN โดยทำการพยากรณ์คุณภาพ



ของน้ำที่แม่น้ำ Buyuk Menderes ภายใต้อัตรา 3 ตัววัด คือระดับอุณหภูมิน้ำ, ปริมาณโบรอนในน้ำ และปริมาณออกซิเจนที่ละลายในน้ำ ใช้ข้อมูลตัวอย่างรายเดือนจำนวนทั้งสิ้น 108 ข้อมูล แบ่งข้อมูลออกเป็น 2 ชุด คือข้อมูลชุดสร้างตัวแบบจำนวน 72 ข้อมูล และข้อมูลชุดทดสอบจำนวน 36 ข้อมูล สำหรับตัวแบบ ANN ใช้เทคนิคการเรียนรู้ Scale conjugate gradient (SCG) ผลการทดลองพบว่า ตัวแบบ ARIMAANN มีความแม่นยำมากกว่าตัวแบบเดียวทั้งสองค่า Correlation ระหว่าง ค่าพยากรณ์ของตัวแบบผสมกับค่าจริงของข้อมูลชุดทดสอบของปริมาณ โบรอนในน้ำ, ปริมาณออกซิเจนที่ละลายในน้ำและระดับอุณหภูมิน้ำ มีค่า 0.902, 0.893 และ 0.909 ตามลำดับ ซึ่งถือเป็นที่น่าพอใจในงานการพยากรณ์คุณภาพน้ำ

งานวิจัย (Tseng et al., 2002) นำเสนอตัวแบบผสม Hybrid ทำการผสมตัวแบบ Seasonal ARIMA (SARIMA) กับตัวแบบ Neural networks แบบ Backpropagation เรียกว่า SARIMABP ทำการเปรียบเทียบกับตัวแบบเดียว ๆ ที่มีการDifference และ Deseasonalized และตัวแบบ SARIMA รวมเป็น 3 ตัวแบบนำมาพยากรณ์ข้อมูลอนุกรมเวลาที่มีปัจจัยเชิงฤดูกาล 2 ชุดข้อมูล ได้แก่ การผลิตเครื่องจักรรวมและเครื่องดื่มของประเทศไต้หวันรวม ทำการพยากรณ์แบบ 1 ค่าล่วงหน้า ใช้ข้อมูลทดลอง 2 ชุด แต่ละชุดแบ่งข้อมูลเป็น 2 กลุ่มคือชุดเรียนรู้และชุดทดสอบ เปรียบเทียบตัวแบบทั้ง 3 ด้วย MSE, MAE และMAPE ผลการทดลองพบว่าตัวแบบ SARIMABP ให้ค่าทั้งสามต่ำที่สุด ดีกว่าตัว แบบทั้งสาม ทั้ง 2 ชุดข้อมูล

งานวิจัย (Pai and Lin, 2005)งานวิจัยชิ้นนี้ทำการพยากรณ์ราคาหุ้นจำนวน 10 ชุดข้อมูล โดยใช้ตัวแบบ Hybrid ARIMASVM เทียบความแม่นยำกับตัวแบบ ARIMA และSVM ทำการ พยากรณ์ 1 ค่าล่วงหน้า (One-step ahead forecasting) เปรียบเทียบความแม่นยำของแต่ละตัวแบบด้วย RMSE,MSE,MAE และMAPE ผลการทดลองพบว่า ตัวแบบ Hybrid ARIMASVM ให้ผลการพยากรณ์มีความแม่นยำกว่าทั้งตัวแบบ ARIMA และตัวแบบ SVM ผู้วิจัยนำเสนอว่าตัวแบบผสมนี้มีประสิทธิภาพที่ดีเนื่องจากนำส่วนที่ดีของแต่ละตัวแบบ คือ ARIMA และ SVM มาใช้ นอกจากนี้ยังเสนอแนะ ว่าควรมีการนำตัวแบบ ARIMA และSVM มาผสมกันโดยใช้เงื่อนไขอื่นนอกจากการบวกกัน เช่น การคูณ เป็นต้น และแนะนำการใช้เทคนิคในการค้นหาพารามิเตอร์ของตัวแบบ SVM เพื่อให้ได้ตัวแบบที่มีความแม่นยำและใช้เวลาค้นหาสั้นลง

งานวิจัย (Shi et al., 2012)งานวิจัยชิ้นนี้ทำการพยากรณ์ เพื่อจัดการในโรงงานการผลิตกระแสไฟฟ้าด้วยความเร็วลม โดยทำการพยากรณ์ข้อมูลสองชุด คือความเร็วลม และกำลังไฟฟ้าจากกังหันลม ทดสอบโดยใช้ตัวแบบผสม 2 ชนิด คือ Hybrid ARIMAANN และARIMASVM เทียบความแม่นยำกับตัวแบบเดียว คือ ARIMA, ANN และSVM โดยใช้ตัวประเมินตัวแบบ คือ MAE และRMSE ผลการทดลองสรุปได้ว่าตัวแบบผสม ไม่ได้ดีไป

กว่าตัวแบบเดียว อาจเนื่องมาจากเงื่อนไขระหว่างตัวแบบเชิงเส้นตรงคือ ARIMA และตัวแบบไม่เป็นเชิงเส้นตรงคือ ANN และ SVM ไม่ได้เป็นผลบวก อาจมีความสัมพันธ์เป็นอย่างอื่นเช่น ผลคูณ ซึ่งนั่นทำให้ค่าเศษเหลือที่นำไปสู่สร้างตัวแบบทั้ง ANN และ SVM ไม่ได้เป็นข้อมูลที่ไม่เป็นเชิงเส้นตรงเพียงอย่างเดียวแต่มีความสัมพันธ์เชิงเส้นตรงรวมอยู่ด้วย

งานวิจัย (Wang et al., 2012) ทำการพยากรณ์ราคาหุ้น โดยใช้ตัวแบบ Exponential smoothing (ESM), ARIMA และ Back propagation neural network (BPNN) นำมาพยากรณ์ และนำเสนอตัวแบบผสม Hybrid combined โดยนำตัวแบบทั้งสามมารวมกันด้วยค่าน้ำหนัก โดยใช้ Genetic algorithm ในการค้นหาค่า Weights เรียกตัวแบบที่นำเสนอว่า (Proposed hybrid model :PHM) วัดผลความขนาดความแม่นยำด้วย MAE, RMSE, MAPE, ME และวัดผลทิศทางความแม่นยำด้วยตัววัด DA (Direction accuracy) เทียบกับตัวแบบทั้งสาม และเทียบกับตัวแบบ EWH คือตัวแบบ Hybrid ที่กำหนดให้น้ำหนักมีค่าเท่ากัน และ RWM (Random walk model) ซึ่งเป็นตัวแบบที่นิยมใช้ในการพยากรณ์ราคาหุ้น ผลการทดลองพบว่าตัวแบบ PHM มีความแม่นยำทั้งขนาดและทิศทางมากกว่าทุก ๆ ตัวแบบ เหมาะที่จะนำไปใช้กับการพยากรณ์ราคาหุ้น

งานวิจัย (Wang et al., 2010) ทำการพยากรณ์กำลังไฟฟ้าเพื่อที่จะผลิตกระแสไฟฟ้าให้เพียงพอและประหยัดงานวิจัยชิ้นนี้ นำเสนอตัวแบบผสม Hybrid combined โดยใช้ตัวแบบ SARIMA, Exponential smoothing แบบมีปัจจัยฤดูกาล และ Support vector machine แบบถ่วงน้ำหนักโดยนำ Adaptive particle swarm optimization (APSO) มาช่วยในการค้นหาค่า Weights ที่เหมาะสม ตัวแบบผสมในที่นี้ทำการรวมทั้งสามตัวแบบให้อยู่ในรูปแบบผลรวมเชิงเส้นตรง โดยที่ใช้ APSO ในการค้นหา Weights ทั้งสามตัวแบบ ใช้ Regression significance test ทดสอบสมการผลรวมเชิงเส้นตรงที่ได้ว่าดีและผ่านเงื่อนไขหรือไม่ ผลการทดลองพบว่าตัวแบบผสมที่นำเสนอนี้ดีกว่าตัวแบบเดียว ทั้งสามตัวแบบ เช่น ความแม่นยำของตัวแบบที่นำเสนอเพิ่มขึ้นกว่าตัวแบบ S-ARIMA, SESM, W-SVM คิดเป็นเปอร์เซ็นต์ คือ 30.746%, 45.358%, 45.494% ตามลำดับ การพยากรณ์ ด้วยตัวแบบเดียวนั้นไม่สามารถเก็บรายละเอียดของข้อมูลได้ดีเพียงพอการผสมตัวแบบจะเป็นข้อดี ของแต่ละตัวแบบ ทำให้เก็บรายละเอียดของข้อมูลได้ดีมากยิ่งขึ้น ตัวแบบผสมที่นำเสนอนี้ สามารถใช้ นำไปใช้พยากรณ์กำลังไฟฟ้าในการจัดการการผลิต เพื่อจะลดปริมาณกำลังไฟฟ้าที่ปล่อยออกมา เกินความจำเป็นช่วยบริษัทผลิตกระแสไฟฟ้าในการตัดสินใจ

งานวิจัย (Chen and Wang, 2007) นำเสนอตัวแบบผสมโดยนำประสิทธิภาพของตัวแบบ SARIMA และตัวแบบ Support Vector Machine (SVM) ทำการพยากรณ์ข้อมูลอนุกรมเวลาที่มี ปัจจัยเชิงฤดูกาล โดยใช้ข้อมูลตัวอย่างของมูลค่าผลผลิตเครื่องจักรกลในประเทศไต้หวัน เทียบ ประสิทธิภาพตัวแบบผสมกับตัวแบบเดียว

คือ SARIMA และ SVM สำหรับข้อมูลที่นำมาทดสอบ แบ่งเป็นสองชุด คือชุดเรียนรู้เพื่อสร้างตัวแบบและชุดทดสอบตัวแบบ ข้อมูลชุดเรียนรู้ตัวแบบมี จำนวน 72 ตัวอย่าง ส่วนข้อมูลชุดทดสอบมีจำนวน 10 ตัวอย่างสำหรับตัวแบบ SVM มีการนำเทคนิค Genetic algorithm มาใช้เพื่อช่วยค้นหาพารามิเตอร์ของ SVM ตัวแบบผสมใช้หลักการ Traditional hybrid สมมติความสัมพันธ์ระหว่าง SARIMA และ SVM เป็นผลบวก (Additive) โดยแบ่งเป็น 3 ตัวแบบย่อยที่ใช้ทดลอง คือ SARIMASVM1, SARIMASVM2, SARIMASVM3 ที่มีการนำ Lag นำเข้าตัวแบบ SVM ที่ต่างกัน สำหรับตัววัดประสิทธิภาพของตัวแบบใช้ Normalized mean square error (NMSE), mean absolute percentage error (MAPE) ตัวสถิติวัดความแม่นยำด้าน ทิศทาง และขนาดสัมประสิทธิ์ความสัมพันธ์ ( $R^2$ ) ผลการทดลองพบว่าตัวแบบ SARIMASVM2 ซึ่งมี Lag นำเข้า คือ  $(Y_{t-1}, Y_{t-2}, \varepsilon_t)$  ให้ผลความแม่นยำดีที่สุดทั้งในแง่ขนาดความผิดพลาดที่น้อย ที่สุด มีความแม่นยำดีที่สุดและสามารถพยากรณ์ทิศทางได้แม่นยำอีกด้วย

### บทที่ 3

#### วิธีการดำเนินงานวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อพยากรณ์ตัวส่งสัญญาณการซื้อ-ขายล่วงหน้าของตัวบ่งชี้ Bollinger Band ร่วมกับการใช้ตัวแบบเดี่ยว ได้แก่ ARIMAX, ANN, SVM และตัวแบบผสม ได้แก่ ARIMAX-ANN, ARIMAX-SVR, ARIMAX-ANN-SVR เพื่อใช้สำหรับการหาตัวแบบพยากรณ์ดัชนีราคา SET50 ที่ดีที่สุด โดยใช้โปรแกรม SPSS ในการทดสอบทางสถิติ และ Python ในการพัฒนาตัวแบบ ซึ่งมีขั้นตอนการวิจัยดังนี้

1. กลุ่มตัวอย่างและการเก็บรวบรวมข้อมูล
2. การสร้างตัวแบบและการวิเคราะห์ข้อมูล
3. การคำนวณหาความคลาดเคลื่อน
4. การพยากรณ์สัญญาณการซื้อ-ขายล่วงหน้าและการวัดประสิทธิภาพ
5. แผนผังการดำเนินงานวิจัย

#### 3.1 กลุ่มตัวอย่างและการเก็บรวบรวมข้อมูล

งานวิจัยฉบับนี้ทำการเก็บข้อมูลทุติยภูมิ (Secondary Data) รายวันในช่วง 1 มกราคม พ.ศ. 2563 – 18 เมษายน พ.ศ. 2566 ซึ่งรวมเป็นจำนวน 628 วัน เพื่อมุ่งเน้นในการนำชุดข้อมูลดังกล่าวไปสร้าง ตัวแบบการพยากรณ์ราคาดัชนี SET50 Index ซึ่งจะใช้ข้อมูลราคาปิดของดัชนี SET50 Index ผ่านเครื่องมือพยากรณ์ทางทั้ง 6 วิธีประกอบด้วย ตัวแบบเดี่ยว 3 ตัว ได้แก่ ARIMAX model, ANN model, SVR model ตัวแบบผสม 2 ตัว ได้แก่ ARIMAX-ANN , ARIMAX-SVR และตัวแบบรวม คือ ARIMAX-ANN-SVR รวมทั้งหมด 6 ตัวแบบ

เนื่องจากงานวิจัยนี้มีการใช้การพยากรณ์ผ่านเครื่องมือ ARIMAX Model ซึ่งเป็น เครื่องมือที่พัฒนาจากตัวแบบพยากรณ์ ARIMA โดยมีการนำปัจจัยที่มีผลต่อดัชนี SET50 Index ในรูปแบบของความสัมพันธ์ระหว่างตัวแปรตาม (Dependent Factor) และตัวแปรอิสระ (Independent Factor) มารวมคำนวณในตัวแบบพยากรณ์ด้วย โดยตัวแปรอิสระประกอบด้วย 1.ดัชนีราคาหุ้นตลาดหลักทรัพย์แห่งประเทศไทยย้อนหลัง1วัน ( $SET50_{t-1}$ ) 2. ดัชนี Dow Jones Industrial Average ประเทศสหรัฐอเมริกา (DOW) 3.ราคาน้ำมันดิบเวสต์เท็กซัส (WTI) 4. อัตราแลกเปลี่ยน (USD) 5.ราคาทองคำแท่ง (GOLD) 6.ดัชนี Hang Seng 7.ดัชนี Nikkie 225 ประเทศญี่ปุ่น 8. ดัชนี Composite ประเทศอินโดนีเซีย (JKSE) 9.ดัชนี PSEI ประเทศฟิลิปปินส์ 10.ดัชนี Merval Buenos Aires ประเทศอาร์เจนตินา 11.ดัชนีตลาดหลักทรัพย์ ประเทศสิงคโปร์ (STI) และ12.ดัชนี FTSE 100 ประเทศอังกฤษ (FTSE) รวมทั้งสิ้น 12 ตัวแปร โดยตัวแปรตาม คือ ดัชนีหลักทรัพย์แห่งประเทศไทย SET50 Index

โดยข้อมูลทั้งตัวแปรตามและตัวแปรอิสระมีการแบ่งส่วนข้อมูลเพื่อใช้ตามวัตถุประสงค์ต่างๆเป็น 3 ส่วน ดังนี้ ส่วนที่ 1 สำหรับสร้างตัวแบบ ส่วนที่ 2 สำหรับทดสอบตัวแบบ และ ส่วนที่ 3 สำหรับเปรียบเทียบผลลัพธ์จากการพยากรณ์สัญญาณซื้อ-ขายล่วงหน้าของตัวแบบกับสัญญาณจริง โดยส่วนที่ 1 ใช้เป็น 60% แรกของข้อมูลทั้งหมดคือวันที่ 1 มกราคม 2563 ถึงวันที่ 28 ธันวาคม 2564 ส่วนที่ 2 ใช้เป็น 20% ต่อมาจากส่วนแรกคือวันที่ 29 ธันวาคม 2564 ถึงวันที่ 19 สิงหาคม 2565 และส่วนที่ 3 ใช้เป็น 20% ต่อมาจากส่วนที่สองคือวันที่ 22 สิงหาคม 2565 ถึงวันที่ 18 เมษายน 2566 ซึ่งตัวแปรตามและตัวแปรอิสระที่ใช้ในการวิเคราะห์ มีแหล่งที่มาของข้อมูลที่รวบรวมดังนี้

**ตารางที่ 3.1** ตารางชี้แจงแหล่งที่มาของข้อมูลในการนำมาใช้พยากรณ์

ตัวแปรอิสระ (Independent Factor)	แหล่งที่มา
1.ดัชนีราคาหุ้นตลาดหลักทรัพย์แห่งประเทศไทยย้อนหลัง 1 วัน (SET50 <sub>t-1</sub> )	setsmart.com
2.ดัชนี Dow Jones Industrial Average ประเทศสหรัฐอเมริกา (DOW)	finance.yahoo.com
3.ราคาน้ำมันดิบเวสต์เท็กซัส (WTI)	investing.com
4.อัตราแลกเปลี่ยน (USD)	oanda.com
5.ราคาทองคำแท่ง (GOLD)	Swivel.com
6.ดัชนี Hang Seng (Hang)	investing.com
7.ดัชนี Nikkie 225 ประเทศญี่ปุ่น (Nik)	investing.com
8.ดัชนี Composite ประเทศอินโดนีเซีย (JKSE)	investing.com
9.ดัชนี PSEI ประเทศฟิลิปปินส์ (PSEI)	investing.com
10.ดัชนี Merval Buenos Aires ประเทศอาร์เจนตินา (Merval)	investing.com
11.ดัชนีตลาดหลักทรัพย์ ประเทศสิงคโปร์ (STI)	investing.com
12.ดัชนี FTSE 100 ประเทศอังกฤษ (FTSE)	investing.com
ตัวแปรตาม (Dependent Factor)	แหล่งที่มา
ดัชนีราคาหุ้นตลาดหลักทรัพย์แห่งประเทศไทย (SET50)	setsmart.com

### 3.2 การสร้างตัวแบบและวิเคราะห์ข้อมูล

#### 3.2.1 การวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis : PCA)

เนื่องจากในงานวิจัยนี้จะไม่นำปัจจัยภายนอกทั้ง 12 ปัจจัยมาใช้ร่วมกับตัวแบบพยากรณ์ทันทีแต่จะนำปัจจัยภายนอกทั้งหมดมาผ่านการวิเคราะห์องค์ประกอบหลัก (PCA) ก่อน แล้วค่อยนำมาใช้ร่วมกับตัวแบบพยากรณ์ เพื่อป้องกันปัญหาการเกิด Multicollinearity ที่จะมีต่อปัจจัยภายนอกของมันเอง ซึ่งมีวิธีการดังนี้

##### 3.2.1.1 เตรียมข้อมูลและตรวจสอบความสัมพันธ์ระหว่างตัวแปร

เตรียมข้อมูลและตรวจสอบความสัมพันธ์ระหว่างตัวแปร ข้อมูลที่ใช้ในการวิเคราะห์องค์ประกอบนั้น จะต้องมีลักษณะเป็นตัวแปรต่อเนื่องหลาย ๆ ในทุกตัวแปร หลังจากนั้นทำการตรวจสอบว่าตัวแปรต่าง ๆ มีความสัมพันธ์กันหรือไม่ถ้าตัวแปรมีความสัมพันธ์กันมาก หรือมีความสัมพันธ์กันน้อยอย่างมีนัยสำคัญจะสามารถใช้เทคนิค PCA ได้ถ้าตัวแปรไม่มีความสัมพันธ์กัน หรือมีความสัมพันธ์กันน้อย ไม่ควรใช้เทคนิค PCA โดยการตรวจสอบทำได้หลายวิธี (กลยาวันิชย์บัญชา, 2548 : 204 -205) ดังนี้

**วิธีที่ 1** การตรวจสอบโดยใช้ค่าสัมประสิทธิ์สหสัมพันธ์ โดยการสร้างเมทริกซ์แสดงสัมประสิทธิ์สหสัมพันธ์ของตัวแปรทุกคู่

- ถ้าค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรคู่ใดมีค่าเข้าใกล้ +1 หรือ -1 แสดงว่าตัวแปรคู่นั้นมีความสัมพันธ์กันมากควรอยู่ใน Factor เดียวกัน
- ถ้าค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรคู่ใดมีค่าใกล้ศูนย์ แสดงว่าตัวแปรคู่นั้นไม่มีความสัมพันธ์กันหรือสัมพันธ์กันน้อยควรอยู่คนละ Factor
- ถ้ามีตัวแปรที่ไม่มีความสัมพันธ์กับตัวแปรอื่นๆ หรือมีความสัมพันธ์กับตัวแปรอื่นๆ ที่เหลือน้อยมาก ควรตัดตัวแปรนั้นออกจากการวิเคราะห์

**วิธีที่ 2** ใช้สถิติ KMO (Kaiser-Meyer-Olkin Measure of Sampling Adequacy) ซึ่งเป็นค่าที่ใช้วัดความเหมาะสมของข้อมูลตัวอย่างที่จะนำมาวิเคราะห์โดยเทคนิค Factor Analysis โดยที่เมื่อ  $r$  คือค่าสัมประสิทธิ์สหสัมพันธ์ซึ่งทำให้ค่า  $0 \leq KMO \leq 1$

- ถ้าค่า KMO มีค่าน้อย (เข้าสู่ศูนย์) แสดงว่าเทคนิค PCA ไม่เหมาะสมกับข้อมูลที่มีอยู่
- ถ้าค่า KMO มีค่ามาก (เข้าสู่หนึ่ง) แสดงว่าเทคนิค PCA เหมาะสมกับข้อมูลที่มีอยู่
- โดยทั่วไปถ้าค่า  $KMO < 0.5$  จะถือว่า ข้อมูลที่มีอยู่ไม่เหมาะสมที่จะใช้เทคนิค PCA

**วิธีที่ 3** Bartlett's Test of Sphericity เป็นการทดสอบค่า ไค-สแควร์ (Chi-square) ของดีเทอร์มิแนนต์ (Determinant) ของเมทริกซ์สหสัมพันธ์ มีการทดสอบสมมติฐานว่าเมทริกซ์สหสัมพันธ์นั้นเป็นเมทริกซ์เอกลักษณ์ (Identity matrix) หรือไม่ จากสมมติฐาน

$H_0$  : ตัวแปรต่างๆ ไม่มีความสัมพันธ์กัน หรือ

$H_1$  : ตัวแปรต่างๆ มีความสัมพันธ์กัน

ดังนั้นถ้ายอมรับ  $H_0$  แสดงว่าตัวแปรไม่มีความสัมพันธ์กัน จึงไม่ควรใช้เทคนิค PCA แต่ถ้าปฏิเสธ  $H_0$  (หรือยอมรับ  $H_1$ ) นั่นคือตัวแปรมีความสัมพันธ์กันจึงสามารถใช้เทคนิค PCA ได้

### 3.2.1.2 การสกัดองค์ประกอบ (Factor extraction)

มีจุดมุ่งหมาย คือ การหาจำนวน Factor ที่สามารถใช้แทนตัวแปรทั้งหมดทุกตัวได้ หรือเป็นการดึงรายละเอียดจากตัวแปรมาไว้ใน Factor เดียว โดยงานวิจัยนี้ใช้เทคนิค Principle Component Analysis หรือเรียกย่อๆ ว่า PCA ตามที่กล่าวมาข้างต้น เป็นเทคนิคที่มีวัตถุประสงค์ที่จะนำรายละเอียดของตัวแปรที่มีจำนวนตัวแปรหลายๆ มาไว้ในองค์ประกอบที่มีเพียงไม่กี่ตัว โดยจะพิจารณาจากรายละเอียดทั้งหมดจากแต่ละตัวแปร ในการวิเคราะห์ PCA จะสร้างการเชื่อมรวมกันเชิงเส้น (Linear combination) ของตัวแปร โดยที่

- Factor ที่ 1 จะเป็น Linear Combination แรกและมีรายละเอียดจากตัวแปรทั้งหมดมากที่สุด หรือกล่าวได้ว่ามีค่าแปรปรวนสูงสุด

- Factor ที่ 2 ก็เป็น Linear Combination ของตัวแปร เช่นกัน และสามารถนำรายละเอียดที่เหลือจาก Factor ที่ 1 โดยพยายามนำรายละเอียดจากที่เหลือมาใส่ใน Factor ที่ 2 ให้มากที่สุด โดยที่ Factor ที่ 2 จะต้องตั้งฉาก (Orthogonal) กับ Factor แรก หรือกล่าวว่า Factor ที่ 2 ไม่มีความสัมพันธ์กับ Factor ที่ 1

- Factor ที่ 3 เป็น Linear Combination ของตัวแปรเช่นกัน ไม่มีความสัมพันธ์กับ Factor ที่ 1 และ 2 และสามารถนำ Information ที่เหลือจากตัวแปรให้มากที่สุด

- ในทำนองเดียวกัน การสร้าง Factor ที่ 4, 5, 6, ... , 12 ก็ใช้หลักเกณฑ์ดังที่กล่าวข้างต้น

### 3.2.2 การสร้างตัวแบบ ARIMAX

แบบจำลอง ARIMAX เป็นการผสมผสานแบบจำลอง ARIMA กับปัจจัยภายนอกที่มีแนวโน้มที่มีอิทธิพลต่อดัชนี SET50 Index โดยมีด้วยกันดังนี้

$$CP = f(Factor_1, Factor_2, \dots, Factor_{12}, AR(p), MA(q))$$

$Factor_i$  คือ ปัจจัยที่  $i$  ที่ผ่านการวิเคราะห์องค์ประกอบหลักจาก 12 ปัจจัยภายนอก

$AR(p)$  คือ Auto Regressive Process ที่แสดงว่าข้อมูลอนุกรมเวลาขึ้นอยู่กับค่าของตัวเองในอดีต โดย  $p$  คือจำนวนระยะห่าง (lag) ของข้อมูลในอดีตจากข้อมูลปัจจุบัน

$MA(q)$  คือ Moving Average Process แสดงว่าข้อมูลอนุกรมเวลาขึ้นอยู่กับความคลาดเคลื่อนที่อยู่ก่อนหน้า โดย  $q$  คือจำนวนระยะห่าง (lag) ของความคลาดเคลื่อนในอดีตจากปัจจุบัน

เนื่องด้วยแบบจำลอง ARIMAX เป็นวิธีการพัฒนาแนวคิดของ Box and Jenkins โดยข้อมูลที่นำมาใช้จะต้องมีลักษณะนิ่ง (Stationary) ดังนั้นจึงต้องเริ่มจากการตรวจสอบคุณสมบัติของข้อมูลก่อนที่จะหารูปแบบจำลองที่เหมาะสมกับการพยากรณ์ได้โดยเรียงตามวิธีดังต่อไปนี้

### 3.2.2.1 การทดสอบคุณสมบัติหนึ่งของข้อมูลผ่านวิธี Unit Root Test

ทำการทดสอบคุณสมบัติความคงที่ (Stationary) ของข้อมูลด้วยวิธีการทดสอบ Unit Root Test ด้วยวิธีของ Augmented Dickey – Fuller Test (ADF) (Dickey & Fuller, 1979) โดยจะทำการทดสอบกรณีมีทั้งค่าคงที่และแนวโน้มเวลาโดยในการพิจารณาว่าข้อมูลมีลักษณะคงที่หรือไม่คงที่นั้นมีสมมติฐานดังนี้

สมมติฐานหลัก

$H_0 : p\text{-value} \geq 0.05$  ข้อมูลเป็นข้อมูลแบบ non-stationary

สมมติฐานรอง

$H_1 : p\text{-value} < 0.05$  ข้อมูลเป็นข้อมูลแบบ stationary

ถ้าข้อมูลที่ทำการทดสอบยอมรับสมมติฐานหลัก  $H_0$  แสดงว่าข้อมูลนั้นมีลักษณะไม่คงที่ (Non-Stationary) แต่ถ้าข้อมูลที่ทำการทดสอบปฏิเสธสมมติฐานหลัก  $H_0$  แสดงว่าข้อมูลนั้นมีลักษณะคงที่ (Stationary)

### 3.2.2.2 การกำหนดรูปแบบจำลอง ARIMAX( $p, d, q$ )

การกำหนดรูปแบบจำลอง ARIMAX โดยพิจารณาจาก Correlogram ซึ่งดูจากค่า Autocorrelation Function (AFC) และ Partial Autocorrelation Function (PACF) เพื่อที่จะระบุว่าเป็นแบบจำลองนั้นควรจะมี Autoregressive AR( $p$ ) และ Moving Average MA( $q$ ) เท่าใด และจะทำการสร้างแบบจำลองไว้หลายๆ รูปแบบโดยใช้สถิติเพื่อประกอบการพิจารณาการเลือกแบบจำลองที่ดีที่สุดจากค่า Akaike' Information Criterion (AIC) ที่มีค่าต่ำที่สุด เพื่อหาแบบจำลองที่เหมาะสมที่สุดที่จะใช้ในการพยากรณ์

### 3.2.2.3 การวิเคราะห์ความถูกต้อง (Diagnostic Checking)

การตรวจสอบรูปแบบจำลองว่ามีความเหมาะสมหรือไม่ด้วยวิธี (Correlogram) ของ  $r_K$  ของค่าความคลาดเคลื่อน และการทดสอบความเหมาะสมของแบบจำลองโดยการทดสอบ Box และ Ljung ซึ่งพิจารณาค่า Q-statistic ดังสมการนี้



$$Q_m = n(n+2) \sum_{k=1}^m \left\{ \frac{r_k^2(\varepsilon_t)}{(n-k)} \right\}$$

โดยที่  $\varepsilon_t$  แทนค่าความคลาดเคลื่อนของการพยากรณ์ ณ เวลา  $t$

$n$  แทนจำนวนอนุกรมของ  $\varepsilon_t$  ทั้งหมด

$m$  แทนช่วงเวลาที่ห่างกันมากที่สุดของ  $\varepsilon_t$  ทั้งหมด

โดยค่า  $Q$  ที่ได้มีการแจกแจงแบบ Chi-Square และมีองศาความเป็นอิสระ (Degree of Freedom) เท่ากับ  $m-n$  โดยมีสมมติฐานว่าเป็นพจน์ของความคลาดเคลื่อนที่ได้จากการประมาณที่มีลักษณะเป็น White Noise หมายถึง แบบจำลองไม่มีอัตโนมัติสัมพันธ์ และสามารถนำแบบจำลองนี้ไปพยากรณ์ต่อได้

### 3.2.3 การสร้างตัวแบบ ANN

การสร้างตัวแบบพยากรณ์ด้วยเทคนิคการเรียนรู้ของเครื่องถือเป็นการนำเทคโนโลยีเข้ามาประยุกต์ใช้ในการจัดการและทำการวิเคราะห์ข้อมูลอนุกรมเวลาทางการเงิน โดยจะทำการเขียน โค้ดในภาษาไพธอน (Python) ผ่านโปรแกรม Google Collaboratory runs on Python 3.9 เพื่อสร้างตัวแบบพยากรณ์จากอัลกอริทึมต่าง ๆ ทั้งนี้ จะมีการใช้ประโยชน์จากแหล่งข้อมูลเปิด (Open Source Library) ที่สำคัญในการเขียนโค้ดไม่ว่าจะเป็นแหล่งข้อมูล “ไซคิตเลิร์น (Scikit-learn)” ซึ่งถือเป็น แหล่งข้อมูลหลักที่ใช้ในการจัดการและวิเคราะห์ข้อมูล รวมถึงอัลกอริทึมเพื่อใช้งานในด้านการจัดประเภทข้อมูลและการวิเคราะห์ถดถอย และในส่วนของ การสร้างตัวแบบการพยากรณ์จากเทคนิคการเรียนรู้เชิงลึก จะใช้แหล่งข้อมูลเปิด “เทนเซอร์โฟลว์ (Tensorflow)” โดยจะทำงานร่วมกับเอพีไอ (Application Programming Interface) “คีราส (Keras)” ในการสร้างตัวแบบ ANN โดยมีการกำหนดส่วนประกอบของ ANN ดังนี้

- โครงสร้างที่ใช้เป็นแบบ Feed Forward Back Propagation
- จำนวนข้อมูลที่นำเข้า (batch) เท่ากับ 5, 10, 15 และ 20 ค่า
- จำนวน Input neuron = จำนวน feature ที่นำเข้า
- จำนวน Hidden layer = 2, จำนวน Hidden neurons คือ 10, 20, 50, 80 และ 100
- จำนวน Output neuron = 1
- Activation function เลือกใช้ ReLu function

- Output function เลือกใช้ Linear function เป็นฟังก์ชันที่ใช้โดยทั่วไปในการพยากรณ์ข้อมูลที่เป็นตัวเลข (Regression and time series)
- ทำการทดลองซ้ำ (epoch) จำนวน 5, 10, 50 และ 100 รอบภายใต้ Initial weights ที่ต่างกัน
- ใช้ MSE เป็นตัวประเมินความแม่นยำของตัวแบบในข้อมูลชุดทดสอบ สำหรับแต่ละชุดของพารามิเตอร์ที่ใส่ลงไป

สรุปมีตัวแปร 3 ตัวที่ต้องค้นหาค่าเอง ได้แก่ จำนวน batch, จำนวน Hidden neurons และจำนวนการรันซ้ำ (epoch) ปัจจุบันยังไม่มีกฎหรือทฤษฎีการเลือกพารามิเตอร์ที่เหมาะสม งานวิจัยชิ้นนี้จึงทำการค้นหาพารามิเตอร์ในปริมาณที่มากเพื่อโอกาสที่เจอพารามิเตอร์ที่ทำให้ค่าพยากรณ์มีความแม่นยำสูงยิ่งขึ้น รวมจำนวนครั้งการรันทั้งสิ้น 3300 ครั้ง

### 3.2.4 การสร้างตัวแบบ SVR

ในการสร้างตัวแบบพยากรณ์ด้วยเทคนิคการเรียนรู้ของเครื่อง โดยจะทำการเขียน โค้ดในภาษาไพธอน (Python) ผ่านโปรแกรม Google Collaboratory runs on Python 3.9 สำหรับตัวแบบ SVR ได้ใช้อัลกอริทึมต่างๆ จากแหล่งข้อมูลเปิด (Open Source Library) ซึ่งก็คือ “ไซคิดเลิร์น (Scikit-learn)” โดยมีการกำหนดส่วนประกอบของ SVR ดังนี้

- Kernel function ที่ใช้คือ Radial basis function (rbf) และ Linear function
- พารามิเตอร์ gamma ที่ใช้คือ ‘scale’ และ ‘auto’
- พารามิเตอร์ C ค่าที่ใส่ได้แก่ 1, 5, 10, 25, 30, 35, 50, 70, 100, 400, 800, 1500, 2500 และ 3500 รวมทั้งสิ้น 14 ค่า
- พารามิเตอร์ epsilon ค่าที่ใส่ได้แก่ 0.00001, 0.0001, 0.001, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 และ 0.99 รวมทั้งสิ้น 14 ค่า
- ใช้ MSE เป็นตัวประเมินความแม่นยำของตัวแบบในข้อมูลชุดทดสอบสำหรับแต่ละชุดของพารามิเตอร์ที่ใส่ลงไป

สรุปมีตัวแปร 4 ตัวที่ต้องค้นหาค่าเอง ได้แก่ ฟังก์ชันของ gamma, ค่าพารามิเตอร์ C, ค่าพารามิเตอร์ epsilon และ ชนิดของ Kernel function ปัจจุบันยังไม่มีกฎหรือทฤษฎีการเลือกพารามิเตอร์ที่เหมาะสม จึงทำการสร้างตัวแบบที่เป็นไปได้ทั้งหมด รวมจำนวนทั้งสิ้น 784 ตัวแบบ และหาพารามิเตอร์ที่เหมาะสมที่สุดโดยการประเมินความแม่นยำจากค่า MSE ในแต่ละตัวแบบที่มีค่าต่ำที่สุด

### 3.2.5 การสร้างตัวแบบ Traditional hybrid ARIMA-ANN และ ARIMA-SVR

นำค่าเศษเหลือ (Residual) ที่ได้จากการหาผลต่างระหว่างค่าจริงและค่าพยากรณ์จากตัวแบบ ARIMAX มาพยากรณ์ด้วย ANN หรือ SVR โดยยึดขอบเขตและพารามิเตอร์ที่ใช้สร้างตัวแบบเดียวกันทั้ง ANN และ SVM ขั้นตอนการสร้างตัวแบบก็จะเป็นดังขั้นตอนการสร้างตัวแบบเดียว ANN หรือ SVR เมื่อได้ค่าพยากรณ์ ANN หรือ SVR ก็นำค่าพยากรณ์นั้นรวมกับค่าพยากรณ์ ARIMAX ก็จะได้ค่าพยากรณ์รวมทั้งหมด

### 3.2.6 การสร้างตัวแบบ Hybrid combined (ARIMA-ANN-SVM)

ทำการพยากรณ์ด้วย ARIMAX, ANN และ SVR แยกส่วนกัน แล้วนำค่าทั้งสามค่ารวมกันโดยใส่ค่า weights ในแต่ละตัวแบบ ค้นหาค่า Weights ที่ดีที่สุดในข้อมูลชุดทดสอบภายใต้เงื่อนไขค่า MSE มีค่าต่ำสุด ค่า Weights ที่ใช้ในตัวแบบ Combined จะขึ้นอยู่กับจำนวนตัวแบบที่นำมาใช้ สำหรับ งานวิจัยชิ้นนี้นำเสนอตัวแบบ Combined (ARIMAX-ANN-SVM) ดังนั้นจะมีค่า Weights จำนวน 3 ค่าคือ ค่า Weight ของ ARIMAX, ค่า Weight ของ ANN และค่า Weight ของ SVM ทำการค้นหาชุด ของค่า Weights ทั้งสามที่ดีที่สุดที่ทำให้ค่า MSE ต่ำที่สุด โดยใช้วิธีการ Nelder-Mead หรือ "Nelder-Mead simplex method" ที่สามารถค้นหาค่า weights ที่ให้ค่า MSE ที่ต่ำที่สุดได้โดยไม่ต้องคอยไล่ทุกค่า weights ที่เป็นไปได้ ทำให้สามารถหาค่า weights ได้อย่างรวดเร็วและแม่นยำ ซึ่งค่า Weights ที่ได้ในแต่ละตัวแบบต้องรวมกันได้เท่ากับ 1

## 3.3 การคำนวณหาความคลาดเคลื่อน

การประเมินประสิทธิภาพของตัวแบบเพื่อวัดความผิดพลาดเชิงปริมาณ เพื่ออธิบายความสามารถในการให้ค่าพยากรณ์ที่ถูกต้องและเหมาะสมในการอธิบายข้อมูลอนุกรมเวลา โดยใช้เกณฑ์ที่ใช้สำหรับการวัดประสิทธิภาพเพื่อคัดเลือกตัวแบบที่ดีที่สุดของตัวแบบในการพยากรณ์ในงานวิจัยนี้มีด้วยกันทั้งหมด 4 เกณฑ์ ดังนี้

### 3.3.1 ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย

$$MAE = \frac{\sum_{t=1}^n |Y_t - \hat{Y}_t|^2}{n}$$

### 3.3.2 รากของความคลาดเคลื่อนกำลังสองเฉลี่ย

$$RMSE = \frac{\sqrt{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}}{n}$$

### 3.3.3 ความคลาดเคลื่อนกำลังสองเฉลี่ย

$$MSE = \frac{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}{n}$$

### 3.3.4 ร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ย

$$MAPE = 100 \times \frac{\sum_{t=1}^n \left| 1 - \frac{\hat{Y}_t}{Y_t} \right|}{n}$$

## 3.4 การพยากรณ์สัญญาณการซื้อขายล่วงหน้าและการวัดประสิทธิภาพ

การพยากรณ์สัญญาณการซื้อขายโดยใช้ตัวแบบพยากรณ์ที่ดีที่สุดจากทั้งหมด 6 ตัวแบบ ร่วมกับเครื่องมือ Bollinger Band ซึ่งจะทำให้การพยากรณ์สัญญาณการซื้อขายใน 2 ส่วนหลัง ซึ่งคิดเป็น 40% (20%+20%) หลังของข้อมูลดัชนี SET50 หรือตั้งแต่วันที่ 29 ธันวาคม 2564 ถึงวันที่ 18 เมษายน 2566 โดยเปรียบเทียบกับสัญญาณการซื้อขายที่ได้จากเครื่องมือ Bollinger Band กับข้อมูลดัชนี SET50 จริงในส่วนการเปรียบเทียบประสิทธิภาพด้วยอัตรากำไร และเปรียบเทียบกับสัญญาณการซื้อขายที่ได้จากเครื่องมือ Moving Average กับข้อมูลดัชนี SET50 จริงด้วยอัตราซื้อขายสำเร็จ

### 3.4.1 เปรียบเทียบประสิทธิภาพด้วยอัตรากำไร

เนื่องด้วยตัวแบบพยากรณ์ที่ได้จากการทดลองจะเป็นตัวแบบที่มีความแม่นยำในการพยากรณ์มากที่สุดทำให้ค่าพยากรณ์ที่ได้อาจมีความใกล้เคียงกับข้อมูลจริงทำให้สัญญาณการซื้อขายที่ได้จากเครื่องมือ Bollinger Band จะไม่ค่อยมีความแตกต่างกันมากนัก ในเรื่องของจุดเวลาการซื้อขาย ดังนั้นการเปรียบเทียบประสิทธิภาพจึงจำเป็นต้องเปรียบเทียบกับอัตรากำไร ซึ่งมีสมการดังนี้

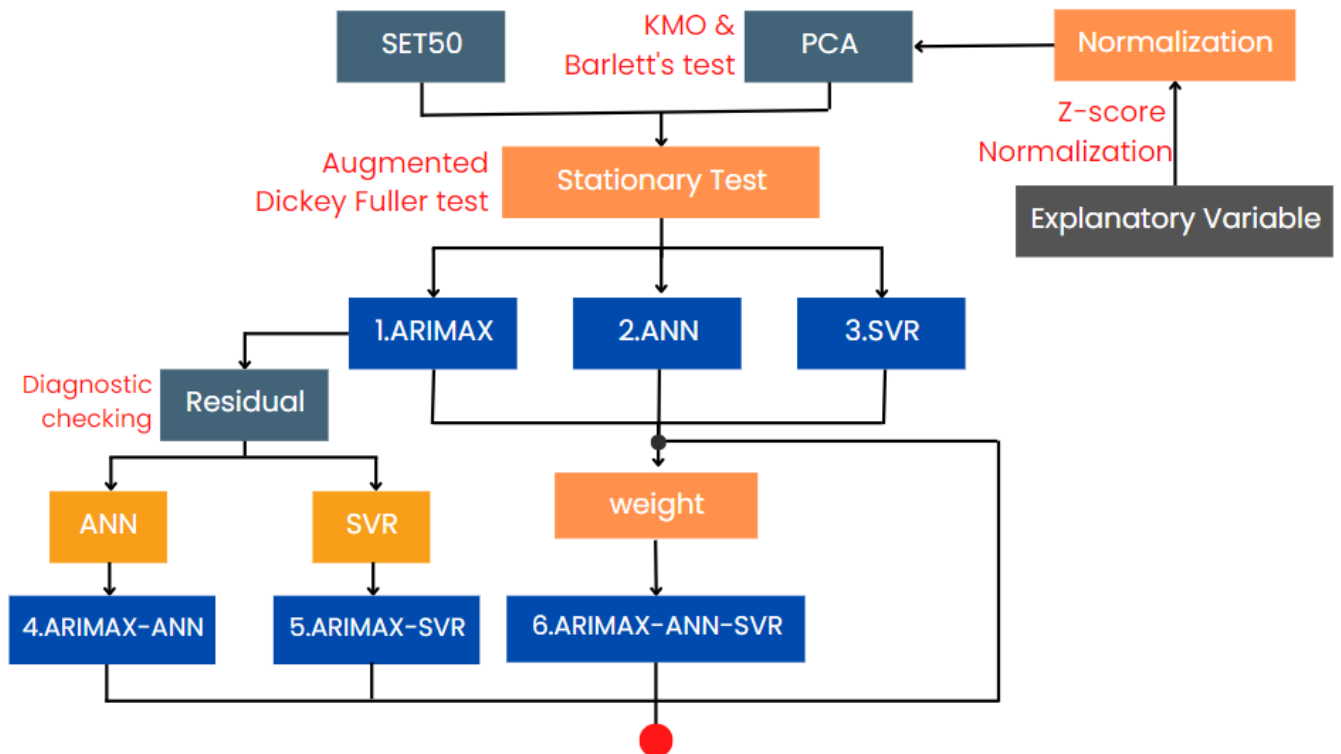
$$\text{profit rate (\%)} = 100 \times \sum_{t=1}^n \frac{\text{close}_{\text{buy}_t} - \text{close}_{\text{sell}_t}}{\text{close}_{\text{buy}_t}}$$

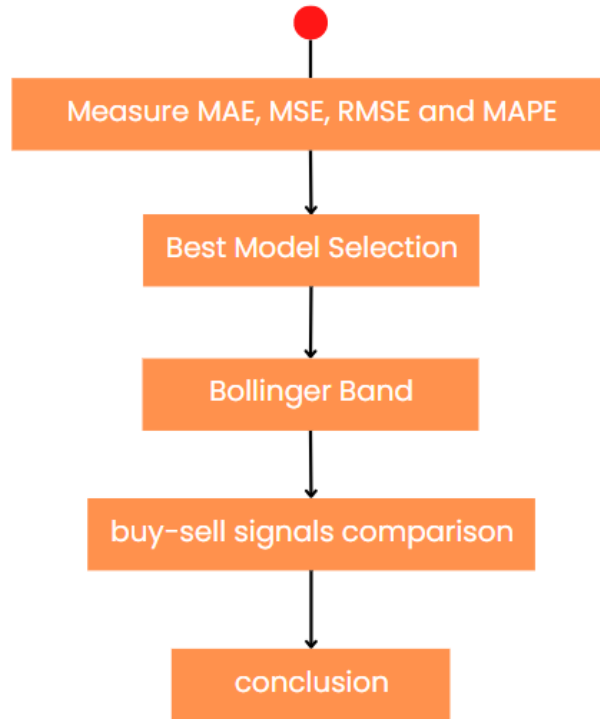
### 3.4.2 เปรียบเทียบประสิทธิภาพด้วยอัตราซื้อขายสำเร็จ

การเปรียบเทียบประสิทธิภาพด้วยอัตราซื้อขายสำเร็จจะเปรียบเทียบระหว่างสัญญาณที่ได้จากเครื่องมือ Moving Average กับข้อมูลดัชนี SET50 จริง กับสัญญาณที่ได้จากเครื่องมือ Bollinger Band กับข้อมูลดัชนี SET50 ที่ได้จากการพยากรณ์จากตัวแบบที่ดีที่สุด เนื่องจากการหาประสิทธิภาพว่าสัญญาณซื้อ-ขายที่ได้จากตัวแบบพยากรณ์ดีกว่าเครื่องมืออย่าง Moving Average ทั่วๆไปหรือไม่ โดยมีสมการดังนี้

$$\text{success rate (\%)} = 100 \times \frac{\sum_{t=1}^n \{1 | \text{close}_{\text{buy}_t} < \text{close}_{\text{sell}_t}, 0 | \text{close}_{\text{buy}_t} \geq \text{close}_{\text{sell}_t}\}}{n}$$

### 3.5 แผนผังการดำเนินงานวิจัย



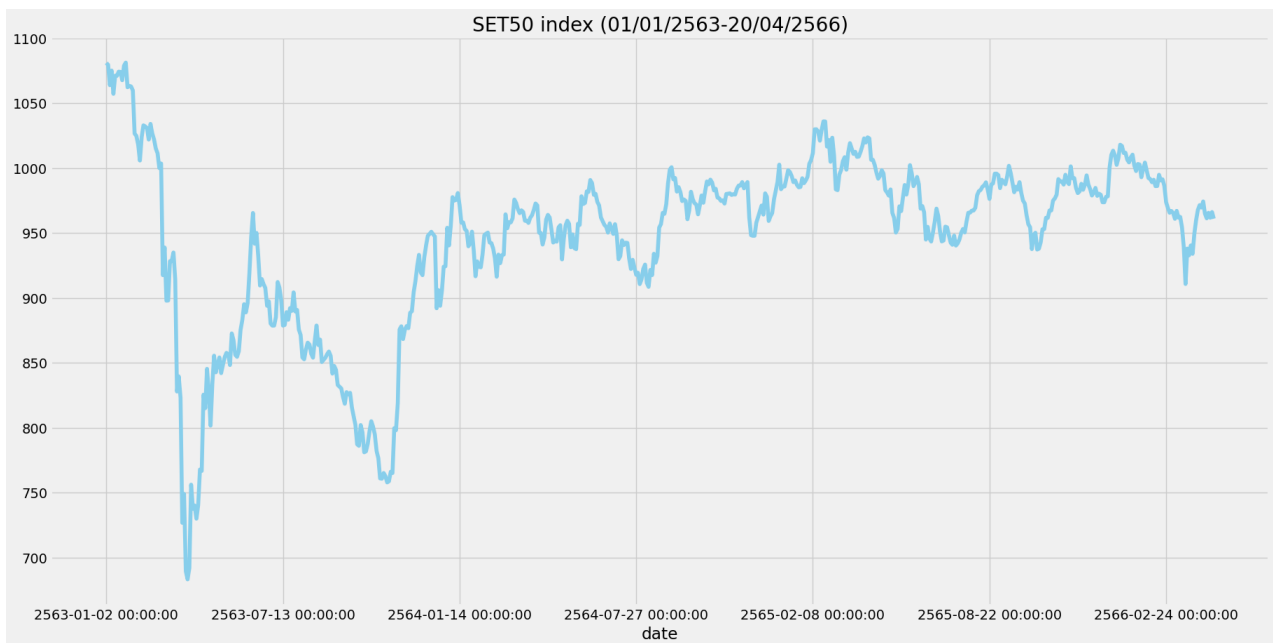


## บทที่ 4

### ผลการวิจัยและอภิปรายผล

#### 4.1 กลุ่มตัวอย่าง และการเก็บรวบรวมข้อมูล

การศึกษาในครั้งนี้เป็นการใช้ข้อมูลกลุ่มตัวอย่างของดัชนี SET50 ซึ่งเป็นข้อมูลราคาดัชนี SET50 ซึ่งเป็นข้อมูลในรูปแบบอนุกรมเวลาแบบรายวันตั้งแต่วันที่ 1 มกราคม พ.ศ. 2563 ถึงวันที่ 18 เมษายน พ.ศ. 2566 โดยไม่นับรวมวันหยุดของตลาดหลักทรัพย์แห่งประเทศไทยโดยมีจำนวนข้อมูล รวมทั้งสิ้น 628 ข้อมูล



รูปที่ 4.1 ข้อมูลอนุกรมเวลา SET50 Index ตั้งแต่ช่วง 1 มกราคม พ.ศ. 2563 ถึง 18 เมษายน พ.ศ.2566

และในการศึกษานี้มีการใช้ข้อมูลตัวแปรอิสระเพิ่มเติมที่นำมาเป็นส่วนประกอบในการพัฒนาเครื่องมือ ARIMAX Model และ Machine learning Model (SVR และ ANN) โดยจะทำการเก็บข้อมูลดังกล่าวในวันเดียวกันกับวันทำการของตลาดหลักทรัพย์แห่งประเทศไทยซึ่งเป็นกลุ่มข้อมูลหลักในการทำงานวิจัย โดยข้อมูลตัวแปรประกอบด้วยข้อมูล 12 ตัวแปร คือ 1.ดัชนีราคาหุ้นตลาดหลักทรัพย์แห่งประเทศไทยย้อนหลัง1วัน ( $SET50_{t-1}$ ) 2.ดัชนี Dow Jones Industrial Average ประเทศสหรัฐอเมริกา (DOW) 3.ราคาน้ำมันดิบเวสต์เทกซัส (WTI) 4. อัตราแลกเปลี่ยน (USD) 5.ราคาทองคำแท่ง (GOLD) 6.ดัชนี Hang Seng 7.ดัชนี Nikkie 225 ประเทศญี่ปุ่น 8. ดัชนี Composite ประเทศอินโดนีเซีย (JKSE) 9.ดัชนี PSEI ประเทศฟิลิปปินส์ 10.ดัชนี Merval Buenos Aires

ประเทศอาร์เจนตินา 11.ดัชนีตลาดหลักทรัพย์ ประเทศสิงคโปร์ (STI) และ 12.ดัชนี FTSE 100 ประเทศอังกฤษ (FTSE) ซึ่งจะนำปัจจัยเหล่านี้ไปทำการวิเคราะห์หองค์ประกอบหลัก เพื่อนำมาใช้ในแบบในขั้นตอนถัดไป (4.1.2)

4.1.1 การทดสอบคุณสมบัติหนึ่งของข้อมูลผ่านวิธี Unit Root Test ผลลัพธ์จากการทำ Unit Root Test ด้วยวิธี Augmented Dickey-Fuller (ADF) ของทั้งก่อน และหลังการแก้ไขข้อมูล

**ตารางที่ 4.1** ผลลัพธ์จากการทดสอบ Augmented Dickey-Fuller test ของข้อมูล SET50 ก่อนการแก้ไขข้อมูล หรือ ณ ผลต่างลำดับที่ 0 (At Level)

<i>Null Hypothesis: SET50 has a unit root</i>			
<i>Exogenous: Constant</i>			
<i>Lag Length: 11 (Automatic – based on AIC, max lag = 32)</i>			
		<i>t-Statistic</i>	<i>Prob*</i>
<i>Augmented Dickey -Fuller test statistics</i>	Test statistics	-3.094297	0.0279
<i>Test critical values;</i>	1% level	-3.447914	
	5% level	-2.869176	
	10% level	-2.570905	

หมายเหตุ \*MacKinnon(1996) one-side p -value



**ตารางที่ 4.2** ผลลัพธ์จากการทดสอบ Augmented Dickey-Fuller test ของข้อมูล SET50 หลังการแก้ไขข้อมูล หรือ ณ ผลต่างลำดับที่ 1 (1st Difference)

<i>Null Hypothesis: D(SET50) has a unit root</i>			
<i>Exogenous: Constant</i>			
<i>Lag Length: 11 (Automatic – based on AIC, max lag =32)</i>			
		<i>t-Statistic</i>	<i>Prob*</i>
<i>Augmented Dickey -Fuller test statistics</i>	<b>Test statistics</b>	-5.175130	0.000
<i>Test critical values;</i>	1% level	-3.447963	
	5% level	-2.869198	
	10% level	-2.570917	

หมายเหตุ \*MacKinnon(1996) one-side p -value

จากการทดสอบข้อมูลผ่านวิธี Unit Root Test ได้ผลลัพธ์ p-value จากผลการทดสอบมีค่าดังนี้ ที่ข้อมูลดั้งเดิมหรือข้อมูลก่อนการแก้ไข ให้ค่า p-value เท่ากับ 0.0279 และในส่วนข้อมูลหลังการแก้ไขหรือข้อมูลที่ทำการหาผลต่างครั้งที่ 1 ให้ค่า p-value ที่ 0.0000 ซึ่งข้อมูลทั้งก่อนและหลังการทำผลต่างครั้งที่ 1 อยู่ในระดับที่ต่ำกว่าระดับนัยสำคัญ (Level of Significant) ที่ 0.05 ทั้งคู่ หรือสามารถกล่าวได้ว่าข้อมูลดัชนี SET50 ทั้งก่อนและหลังการทำผลต่างครั้งที่ 1 เป็นข้อมูลที่มีคุณสมบัตินิ่ง (Stationary) ที่ระดับนัยสำคัญที่ 0.05

#### 4.1.2 การวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis : PCA)

การทำ PCA กับปัจจัยภายนอกที่จะนำมาใช้ในการสร้างตัวแบบเพื่อพยากรณ์ จะช่วยจัดปัญหาเรื่องตัวแปรมีความสัมพันธ์กันเอง (multicollinearity) ได้ โดยปัญหาดังกล่าวจะสังเกตได้จากผลลัพธ์ดังรูปที่ 4.2

		Correlations												
		SET50	SET50_Shiff	Dow	WTI	USD	GOLD	Hang	Nik	JKSE	PSEI	MERVAL	STI	FTSE
SET50	Pearson Correlation	1	.985**	.682**	.674**	.272**	-.104**	-.120**	.607**	.782**	.733**	.395**	.444**	.347**
	Sig. (2-tailed)		.000	.000	.000	.000	.009	.003	.000	.000	.000	.000	.000	.000
	N	628	628	628	628	628	628	628	628	628	628	628	628	628
SET50_Shiff	Pearson Correlation	.985**	1	.673**	.676**	.272**	-.108**	-.124**	.599**	.782**	.729**	.397**	.455**	.342**
	Sig. (2-tailed)	.000		.000	.000	.000	.007	.002	.000	.000	.000	.000	.000	.000
	N	628	628	628	628	628	628	628	628	628	628	628	628	628
Dow	Pearson Correlation	.682**	.673**	1	.717**	.283**	.301**	-.082**	.896**	.753**	.604**	.420**	.293**	.355**
	Sig. (2-tailed)	.000	.000		.000	.000	.000	.040	.000	.000	.000	.000	.000	.000
	N	628	628	628	628	628	628	628	628	628	628	628	628	628
WTI	Pearson Correlation	.674**	.676**	.717**	1	.637**	.176**	-.498**	.629**	.898**	.383**	.495**	.515**	.064
	Sig. (2-tailed)	.000	.000	.000		.000	.000	.000	.000	.000	.000	.000	.000	.109
	N	628	628	628	628	628	628	628	628	628	628	628	628	628
USD	Pearson Correlation	.272**	.272**	.283**	.637**	1	-.121**	-.876**	.232**	.626**	-.206**	.621**	.510**	-.571**
	Sig. (2-tailed)	.000	.000	.000	.000		.002	.000	.000	.000	.000	.000	.000	.000
	N	628	628	628	628	628	628	628	628	628	628	628	628	628
GOLD	Pearson Correlation	-.104**	-.108**	.301**	.176**	-.121**	1	-.077	.227**	.095**	.026	.244**	-.029	.276**
	Sig. (2-tailed)	.009	.007	.000	.000	.002		.055	.000	.018	.514	.000	.465	.000
	N	628	628	628	628	628	628	628	628	628	628	628	628	628
Hang	Pearson Correlation	-.120**	-.124**	-.082**	-.498**	-.876**	-.077	1	.020	-.486**	.277**	-.655**	-.536**	.582**
	Sig. (2-tailed)	.003	.002	.040	.000	.000	.055		.617	.000	.000	.000	.000	.000
	N	628	628	628	628	628	628	628	628	628	628	628	628	628
Nik	Pearson Correlation	.607**	.599**	.896**	.629**	.232**	.227**	.020	1	.717**	.524**	.352**	.209**	.400**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.617		.000	.000	.000	.000	.000
	N	628	628	628	628	628	628	628	628	628	628	628	628	628
JKSE	Pearson Correlation	.782**	.782**	.753**	.898**	.626**	.095**	-.486**	.717**	1	.508**	.635**	.624**	.077
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.018	.000	.000		.000	.000	.000	.053
	N	628	628	628	628	628	628	628	628	628	628	628	628	628
PSEI	Pearson Correlation	.733**	.729**	.604**	.383**	-.206**	.026	.277**	.524**	.508**	1	.076	.151**	.598**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.514	.000	.000	.000		.058	.000	.000
	N	628	628	628	628	628	628	628	628	628	628	628	628	628
MERVAL	Pearson Correlation	.395**	.397**	.420**	.495**	.621**	.244**	-.655**	.352**	.635**	.076	1	.770**	-.361**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.058		.000	.000
	N	628	628	628	628	628	628	628	628	628	628	628	628	628
STI	Pearson Correlation	.444**	.455**	.293**	.515**	.510**	-.029	-.536**	.209**	.624**	.151**	.770**	1	-.266**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.465	.000	.000	.000	.000	.000		.000
	N	628	628	628	628	628	628	628	628	628	628	628	628	628
FTSE	Pearson Correlation	.347**	.342**	.355**	.064	-.571**	.276**	.582**	.400**	.077	.598**	-.361**	-.266**	1
	Sig. (2-tailed)	.000	.000	.000	.109	.000	.000	.000	.000	.053	.000	.000	.000	
	N	628	628	628	628	628	628	628	628	628	628	628	628	628

\*\* . Correlation is significant at the 0.01 level (2-tailed).

\*. Correlation is significant at the 0.05 level (2-tailed).

## รูปที่ 4.2 การวิเคราะห์สหสัมพันธ์ระหว่างดัชนีราคาหุ้นตลาดหลักทรัพย์แห่งประเทศไทย (SET50) และตัวแปรต่างๆ

จากรูปที่ 4.1 จะพบได้ว่า ตัวแปรทุกตัวมีความสัมพันธ์กับดัชนีราคาหุ้นตลาดหลักทรัพย์แห่งประเทศไทย ที่ระดับนัยสำคัญ 0.01 และนอกจากนี้ยังพบว่า ตัวแปรที่เป็นปัจจัยภายในที่จะนำมาใช้ในการสร้างตัวแบบพยากรณ์ในขั้นตอนถัดไป มีความสัมพันธ์ด้วยกันเอง ที่ระดับนัยสำคัญ 0.05 ซึ่งแสดงให้เห็นถึงปัญหา multicollinearity ที่จะเกิดขึ้นในตัวแบบในงานวิจัยนี้ ดังนั้นจึงได้ทำการวิเคราะห์หองค์ประกอบหลัก หรือการทำ PCA กับปัจจัยภายนอกทุกตัว แต่ก่อนการใช้เทคนิค PCA จำเป็นจะต้องตรวจสอบความเหมาะสมของข้อมูลก่อน โดยการใช้สถิติทดสอบ KMO และ Bartlett's โดยผลลัพธ์ของสถิติทดสอบและผลลัพธ์หลังการจากทำ PCA แสดงดังตารางที่ 4.3 และ รูปที่ 4.4 ตามลำดับ

ตารางที่ 4.3 ผลลัพธ์จากการทดสอบ KMO and Bartlett's

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.726
Bartlett's Test of Sphericity	Approx. Chi-Square	11724.941
	df	78
	Sig.	0.000

จากตารางที่ 4.3 ค่า KMO มากกว่า 0.5 ดังนั้นข้อมูลที่มีอยู่เหมาะสมที่จะใช้เทคนิค Factor Analysis วิธีการทดสอบความเป็นอิสระของข้อมูลด้วย Bartlett's Test of Sphericity มีสมมติฐานว่างคือ ตัวแปรต่างๆ ไม่มีความสัมพันธ์กันซึ่งค่า Sig. = 0.000 อยู่ในระดับที่ต่ำกว่าระดับนัยสำคัญ (Level of Significant) ที่ 0.05 ดังนั้นยอมรับสมมติฐานว่าง กล่าวคือ ตัวแปรต่างๆ ไม่มีความสัมพันธ์กันในระดับนัยสำคัญ 0.05

Correlations															
		SET50	PCA1	PCA2	PCA3	PCA4	PCA5	PCA6	PCA7	PCA8	PCA9	PCA10	PCA11	PCA12	
SET50	Pearson Correlation	1	-.785**	-.353**	.362**	.091*	.121**	.085*	-.167**	-.198**	-.018	.040	-.008	-.012	
	Sig. (2-tailed)		.000	.000	.000	.022	.002	.032	.000	.000	.650	.313	.845	.763	
	N	628	628	628	628	628	628	628	628	628	628	628	628	628	
PCA1	Pearson Correlation	-.785**	1	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	
	Sig. (2-tailed)	.000		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
	N	628	628	628	628	628	628	628	628	628	628	628	628	628	
PCA2	Pearson Correlation	-.353**	.000	1	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	
	Sig. (2-tailed)	.000	1.000		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
	N	628	628	628	628	628	628	628	628	628	628	628	628	628	
PCA3	Pearson Correlation	.362**	.000	.000	1	.000	.000	.000	.000	.000	.000	.000	.000	.000	
	Sig. (2-tailed)	.000	1.000	1.000		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
	N	628	628	628	628	628	628	628	628	628	628	628	628	628	
PCA4	Pearson Correlation	.091*	.000	.000	.000	1	.000	.000	.000	.000	.000	.000	.000	.000	
	Sig. (2-tailed)	.022	1.000	1.000	1.000		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
	N	628	628	628	628	628	628	628	628	628	628	628	628	628	
PCA5	Pearson Correlation	.121**	.000	.000	.000	.000	1	.000	.000	.000	.000	.000	.000	.000	
	Sig. (2-tailed)	.002	1.000	1.000	1.000	1.000		1.000	1.000	1.000	1.000	1.000	1.000	1.000	
	N	628	628	628	628	628	628	628	628	628	628	628	628	628	
PCA6	Pearson Correlation	.085*	.000	.000	.000	.000	.000	1	.000	.000	.000	.000	.000	.000	
	Sig. (2-tailed)	.032	1.000	1.000	1.000	1.000	1.000		1.000	1.000	1.000	1.000	1.000	1.000	
	N	628	628	628	628	628	628	628	628	628	628	628	628	628	
PCA7	Pearson Correlation	-.167**	.000	.000	.000	.000	.000	.000	1	.000	.000	.000	.000	.000	
	Sig. (2-tailed)	.000	1.000	1.000	1.000	1.000	1.000	1.000		1.000	1.000	1.000	1.000	1.000	
	N	628	628	628	628	628	628	628	628	628	628	628	628	628	
PCA8	Pearson Correlation	-.198**	.000	.000	.000	.000	.000	.000	.000	1	.000	.000	.000	.000	
	Sig. (2-tailed)	.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		1.000	1.000	1.000	1.000	
	N	628	628	628	628	628	628	628	628	628	628	628	628	628	
PCA9	Pearson Correlation	-.018	.000	.000	.000	.000	.000	.000	.000	.000	1	.000	.000	.000	
	Sig. (2-tailed)	.650	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		1.000	1.000	1.000	
	N	628	628	628	628	628	628	628	628	628	628	628	628	628	
PCA10	Pearson Correlation	.040	.000	.000	.000	.000	.000	.000	.000	.000	.000	1	.000	.000	
	Sig. (2-tailed)	.313	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		1.000	1.000	
	N	628	628	628	628	628	628	628	628	628	628	628	628	628	
PCA11	Pearson Correlation	-.008	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	1	.000	
	Sig. (2-tailed)	.845	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		1.000	
	N	628	628	628	628	628	628	628	628	628	628	628	628	628	
PCA12	Pearson Correlation	-.012	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	1	
	Sig. (2-tailed)	.763	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		
	N	628	628	628	628	628	628	628	628	628	628	628	628	628	

\*\*. Correlation is significant at the 0.01 level (2-tailed).

\*. Correlation is significant at the 0.05 level (2-tailed).

### รูปที่ 4.3 แสดงการวิเคราะห์สหสัมพันธ์ระหว่างดัชนีราคาหุ้นตลาดหลักทรัพย์แห่งประเทศไทย (SET50) และ ตัวแปรที่ผ่านการทำ PCA

จากรูปที่ 4.3 จะพบว่า ตัวแปรที่ผ่านการทำ PCA มีความสัมพันธ์กับดัชนีราคาหุ้นตลาดหลักทรัพย์แห่งประเทศไทย ที่ระดับนัยสำคัญ 0.01 ได้แก่ ตัวแปร PCA1, PCA2, PCA3, PCA5, PCA7 และ PCA8 ส่วนตัวแปร PCA4 และ PCA6 มีความสัมพันธ์กับดัชนีราคาหุ้นตลาดหลักทรัพย์แห่งประเทศไทย ที่ระดับนัยสำคัญ 0.05 โดยเมื่อสังเกตจะพบว่า ตัวแปรที่ผ่านการทำ PCA นั้นจะไม่มีความสัมพันธ์ด้วยกันเอง ที่ระดับนัยสำคัญ 0.05 ซึ่งแสดงให้เห็นว่าตัวแปรที่ผ่านการทำ PCA ทั้งหมดนี้จะไม่ทำให้เกิดปัญหา Multicollinearity ระหว่างการสร้างหรือวิเคราะห์ตัวแบบในขั้นตอนต่อไป

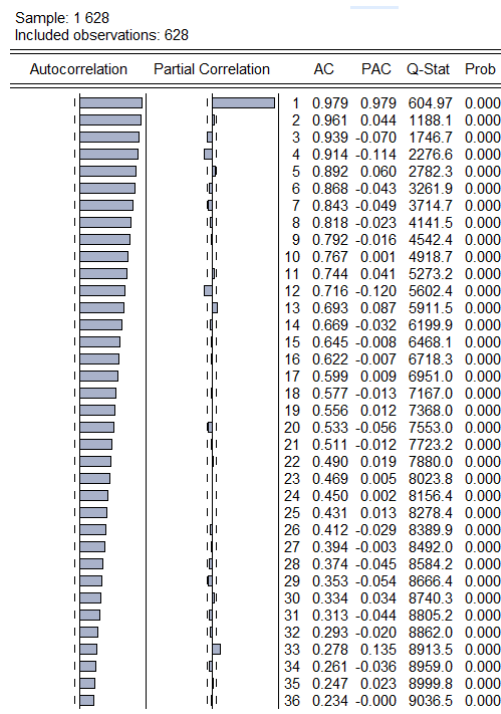
## 4.2 การสร้างตัวแบบ และวิเคราะห์ข้อมูล

### 4.2.1 การสร้างตัวแบบเดี่ยว

#### 4.2.1.1 การสร้างตัวแบบ ARIMAX

##### - การกำหนดรูปแบบ (Identification)

เมื่อพิจารณาผลกราฟฟังก์ชันสหสัมพันธ์ในตัวเอง (ACF) และฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วน (PACF) ของดัชนีราคา SET50 เพื่อกำหนด อันดับ p ของตัวแบบ Autoregressive และอันดับที่ q ของตัวแบบ Moving Average ซึ่งพิจารณา จากค่า Autocorrelation Function (ACF) และค่า Partial Autocorrelation (PACF) ที่แสดงผลที่แสดงค่านอกช่วงความเชื่อมั่นที่ร้อยละ 95 หรือแสดงค่าที่มากที่สุด ซึ่งจะทำให้สามารถกำหนด ตัวแบบที่มีความเป็นไปได้ในการนำมาพยากรณ์เพื่อนำมาพิจารณาหาตัวแบบที่ดีที่สุดต่อไป โดยผล กราฟฟังก์ชันสหสัมพันธ์ในตัวเอง (ACF) และฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วน (PACF) ของดัชนีราคา SET50 มีผลดังรูปที่ 4.4 และรูปที่ 4.5



รูปที่ 4.5 แสดงผล Correlogram ของดัชนี SET50 ณ ผลต่างลำดับที่ 0 (At Level)

Sample: 1 628  
Included observations: 627

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 -0.063	-0.063	2.5027	0.114
		2 0.095	0.092	8.2309	0.016
		3 0.097	0.109	14.147	0.003
		4 -0.068	-0.066	17.097	0.002
		5 0.041	0.013	18.160	0.003
		6 0.035	0.043	18.934	0.004
		7 -0.008	0.005	18.976	0.008
		8 0.022	0.004	19.293	0.013
		9 -0.009	-0.012	19.350	0.022
		10 -0.077	-0.078	23.149	0.010
		11 0.129	0.121	33.730	0.000
		12 -0.112	-0.084	41.829	0.000
		13 0.038	0.016	42.765	0.000
		14 0.006	-0.005	42.791	0.000
		15 -0.049	-0.016	44.353	0.000
		16 0.056	0.036	46.354	0.000
		17 -0.015	-0.004	46.499	0.000
		18 -0.031	-0.031	47.117	0.000
		19 0.068	0.054	50.111	0.000
		20 -0.037	-0.020	50.988	0.000
		21 -0.061	-0.060	53.380	0.000
		22 0.018	-0.026	53.603	0.000
		23 -0.064	-0.013	56.284	0.000
		24 0.001	-0.011	56.285	0.000
		25 -0.000	0.000	56.285	0.000
		26 -0.018	0.004	56.494	0.000
		27 0.047	0.033	57.954	0.000
		28 0.048	0.070	59.488	0.000
		29 -0.068	-0.062	62.509	0.000
		30 0.084	0.040	67.126	0.000
		31 -0.020	0.006	67.380	0.000
		32 -0.059	-0.058	69.673	0.000
		33 0.042	0.002	70.861	0.000
		34 -0.028	0.005	71.397	0.000
		35 -0.026	-0.039	71.858	0.000
		36 -0.040	-0.054	72.921	0.000

รูปที่ 4.6 แสดงผล Correlogram ของดัชนี SET50 ณ ผลต่างลำดับที่ 1 (1<sup>st</sup> Difference)

จากรูปที่ 4.4 และ 4.5 สามารถสรุปได้ว่าตัวแบบสำหรับชุดข้อมูลดัชนีราคา SET50 นั้น คือตัวแบบ  $AR(1)$   $AR(3)$   $AR(2)$   $AR(4)$  และตัวแบบ  $MA(1)$   $MA(2)$   $MA(3)$   $MA(4)$  ณ ผลต่างลำดับที่ 0 (At -Level) และ 1 (1<sup>st</sup> Difference)

ตารางที่ 4.4 เปรียบเทียบตัวแบบ ARIMA ที่เหมาะสมที่สุดผ่านตัวชี้วัด Akaike Info Criterion (AIC)

Model	AIC
$ARIMAX(2,0,2)$	3137.437
$ARIMAX(0,0,0)$	4368.995
$ARIMAX(1,0,0)$	3143.527,
$ARIMAX(0,0,1)$	3940.794
$ARIMAX(0,0,0)$	6270.864,
$ARIMAX(1,0,2)$	3137.306,
$ARIMAX(0,0,2)$	3713.956,

Model	AIC
<i>ARIMAX(1,0,1)</i>	3144.420,
<i>ARIMAX(1,0,3)</i>	3135.699
<i>ARIMAX(0,0,3)</i>	3529.566
<b><i>ARIMAX(2,0,3)</i></b>	<b>3133.381</b>
<i>ARIMAX(3,0,3)</i>	3135.647
<i>ARIMAX(2,0,4)</i>	3135.359
<i>ARIMAX(1,0,4)</i>	3134.535
<i>ARIMAX(3,0,2)</i>	3135.325
<i>ARIMAX(3,0,4)</i>	3137.033
<i>ARIMAX(2,0,3)</i>	3139.866

ผลจากตารางที่ 4.4 พบว่าตัวแบบที่ให้ค่า AIC ต่ำที่สุดคือตัวแบบ ***ARIMA(2,0,3)*** ที่ให้ค่า AIC ที่ต่ำที่สุดที่ค่า 3133.381 โดยจะนำตัวแบบที่ให้ค่า AIC ต่ำที่สุดไปสร้างตัวแบบ ARIMAX ต่อ ซึ่งมีปัจจัยภายนอก (Exogenous Variable) ที่ผ่านการทำ PCA ทั้งหมด 12 ปัจจัย ได้แก่

- 1.ดัชนีราคาหุ้นตลาดหลักทรัพย์แห่งประเทศไทยย้อนหลัง1วัน ( $SET50_{t-1}$ )
- 2.ดัชนี Dow Jones Industrial Average ประเทศสหรัฐอเมริกา (DOW)
- 3.ราคาน้ำมันดิบเวสต์เทกซัส (WTI)
- 4.อัตราแลกเปลี่ยน (USD)
- 5.ราคาทองคำแท่ง (GOLD)
- 6.ดัชนี Hang Seng ประเทศฮ่องกง (Hang)
- 7.ดัชนี Nikkie 225 ประเทศญี่ปุ่น (NIK)
- 8.ดัชนี Composite ประเทศอินโดนีเซีย (JKSE)
- 9.ดัชนี PSEI ประเทศฟิลิปปินส์ (PSEI)
- 10.ดัชนี Merval Buenos Aires ประเทศอาร์เจนตินา (Merval)
- 11.ดัชนีตลาดหลักทรัพย์ ประเทศสิงคโปร์ (STI)
- 12.ดัชนี FTSE 100 ประเทศอังกฤษ (FTSE)

โดยจะเลือกใช้ตัวแปรที่มีความสัมพันธ์ ในระดับนัยสำคัญที่ 0.01 ซึ่งได้แก่ PCA1, PCA2, PCA3, PCA5, PCA7 และ PCA8 รวมทั้งหมด 6 ตัวแปรที่ใช้ในการสร้างตัวแบบ ARIMAX แต่เนื่องจากหลังการสร้างตัวแบบ ARIMAX พบว่ามีค่าพารามิเตอร์ในตัวแบบบางตัวที่มีค่าไม่แตกต่างจาก 0 ที่ระดับนัยสำคัญ 0.05

ตารางที่ 4.5 แสดงผลลัพธ์ของตัวแบบ  $ARIMAX(2,0,3)$

SARIMAX Results						
Dep.Variable: SET50			No. Observation: 380			
Model: ARIMA(2,0,3)			Log Likelihood: -1450.090			
Date: Tue, 25 Apr 2023			AIC: 2926.180			
Time: 09.46.51			BIC: 2977.402			
Sample: 0-380			HQIC: 2946.505			
Covariance Type: opg						
	coef	Std err	Z	P> Z	[0.025	0.975]
const	945.5704	34.940	27.063	0.000	877.090	1014.051
PC1	-12.8721	4.396	-2.928	0.003	-21.487	-4.257
PC2	-31.2620	3.444	-9.077	0.000	-38.012	-24.511
PC3	8.2498	2.719	3.034	0.002	2.920	13.580
PC5	-14.7328	3.481	-4.232	0.000	-21.5356	-7.909
PC7	3.6188	3.646	0.993	0.321	-3.527	10.765
PC8	2.9125	4.167	0.699	0.485	-5.254	11.079
ar.L1	0.1247	0.054	2.296	0.022	0.018	0.231
ar.L2	0.8490	0.053	15.933	0.000	0.745	0.953
ma.L1	0.6142	0.061	9.992	0.000	0.494	0.735
ma.L2	-0.2509	0.061	-4.085	0.000	-0.371	-0.131
ma.L3	0.0668	0.057	1.170	0.242	-0.045	0.179
Sigma2	119.2924	5.164	23.101	0.000	109.171	129.414

จากผลลัพธ์ตัวแบบ  $ARIMAX(2,0,3)$  พบว่ามีพารามิเตอร์ในตัวแบบบางตัวที่ไม่แตกต่างจาก 0 ที่ระดับนัยสำคัญ 0.05 จึงทำการปรับปรุงตัวแบบใหม่เป็นตัวแบบ  $ARIMAX(2,0,2)$



ตารางที่ 4.6 ผลลัพธ์ของตัวแบบ ARIMAX(2,0,2)

SARIMAX Results						
Dep.Variable: SET50			No. Observation: 380			
Model: ARIMA(2,0,2)			Log Likelihood: -1449.433			
Date: Tue, 25 Apr 2023			AIC: 2918.886			
Time: 18.27.20			BIC: 2958.267			
Sample: 0-380			HQIC: 2934.500			
Covariance Type: opg						
	coef	Std err	Z	P> Z	[0.025	0.975]
const	937.5339	35.239	26.605	0.000	868.467	1006.601
PC1	-9.6185	2.825	-3.405	0.001	-15.155	-4.082
PC2	-33.9232	2.183	-15.537	0.000	-38.203	-29.644
PC3	7.7357	2.165	3.573	0.000	3.429	11.979
PC5	-14.6701	3.501	-4.191	0.000	-21.531	-7.809
ar.L1	0.0933	0.036	2.572	0.010	0.022	0.164
ar.L2	0.8858	0.035	25.272	0.000	0.817	0.955
ma.L1	0.6317	0.048	13.046	0.000	0.537	0.727
ma.L2	-0.3302	0.037	-9.041	0.000	-0.402	-0.259
Sigma2	118.2637	4.959	23.848	0.000	108.544	127.983

ซึ่งจากผลลัพธ์ของตัวแบบ ARIMAX(2,0,2) พบว่าพารามิเตอร์ทุกตัวมีค่าแตกต่างจาก 0 ที่ระดับนัยสำคัญ 0.05 และจากตัวผลลัพธ์ของตัวแบบสามารถอธิบายเป็นสมการได้ดังนี้

$$\hat{Y}_t = 937.5339 - 9.6185PC1 - 33.9232PC2 + 7.7357PC3 - 14.6701PC5 + 0.0933Y_{t-1} + 0.8858Y_{t-2} + 0.6317\varepsilon_{t-1} - 0.3302\varepsilon_{t-2} + \varepsilon_t$$

- การตรวจสอบความถูกต้อง (Diagnostic Checking)

ก) ทดสอบความเป็นอิสระของข้อมูลอนุกรมเวลา เมื่อทำการประมวลผลผ่านการทำ Ljung and Box Test เพื่อทดสอบคุณสมบัติความเป็น White Noise ของความคลาดเคลื่อนของตัวแบบพยากรณ์ สามารถแสดงผลได้ดังต่อไปนี้

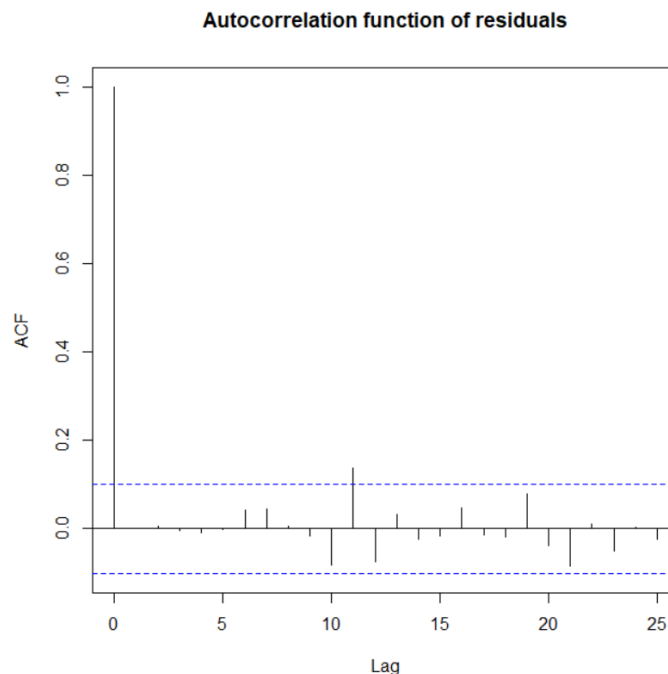
ตารางที่ 4.7 ผลการทดสอบ Ljung and Box เพื่อประเมิน White Noise ในตัวแบบ ARIMAX ที่นำมาพยากรณ์

Model	Lag	Q-Statistic	Probability
ARIMAX (2,0,2)	10	1.839	0.187

ผลจากตารางที่ 4.8 แสดงให้เห็นถึง Q-Statistic ของแบบ ARIMAX (2,0,2) ที่มีความล่าช้าของเวลาที่ 10 โดยมี Probability ที่ไม่แตกต่างจาก 0.05 อย่างมีนัยสำคัญ ซึ่งหมายความว่า ค่าความคลาดเคลื่อนของตัวแบบมีลักษณะการกระจายแบบปกติ ค่าเฉลี่ยเท่ากับศูนย์ และค่าแปรปรวนคงที่ หรือมีลักษณะเป็น White Noise อันเป็นผลให้ตัวแบบมีความเหมาะสมในการนำมาใช้พยากรณ์ต่อไป

ข) ตรวจสอบค่าเศษเหลือ

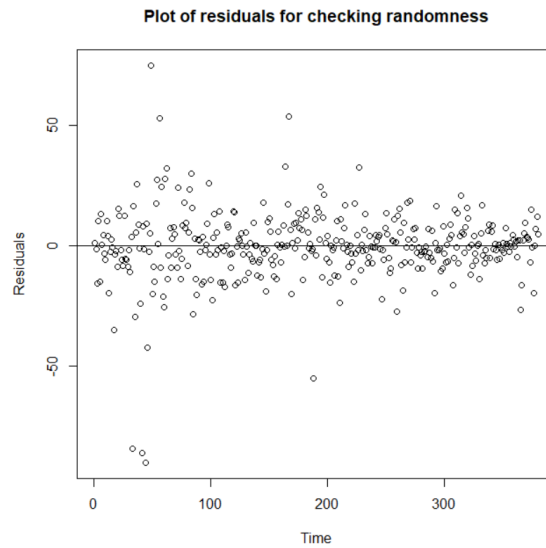
ข.1) กราฟ Autocorrelation function (ACF) ของเศษเหลือ



รูปที่ 4.6 ACF ของค่าเศษเหลือ

กราฟ ACF ที่เหมาะสม เมื่อค่าเศษเหลือเข้าเงื่อนไขเป็นอิสระจากกัน จะเป็นไปตามภาพที่ 4.2 เส้น ACF จะทะลุออกนอกเส้นประที่เป็นขอบเขตที่ Lag=0 และ Lag=11 โดยที่ Lag=11 เกินออกมาเพียงนิดเดียว แต่ที่ Lag อื่นๆ เส้น ACF จะอยู่ภายในเส้นประทั้งหมด

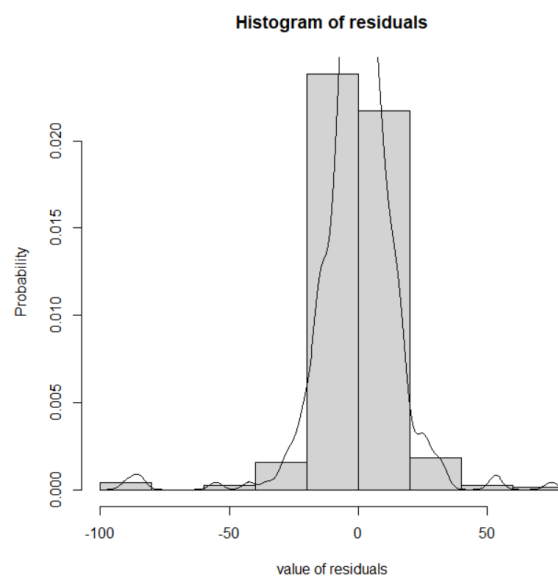
ข.2) กราฟ Scatter plot ของค่าเศษเหลือ



รูปที่ 4.7 Scatter plot ของค่าเศษเหลือ

Scatter plot ช่วยบอกให้รู้ว่าข้อมูลมีการกระจายตัวเป็นอิสระหรือไม่ ซึ่งสามารถช่วยตรวจสอบความผิดปกติของความสัมพันธ์ของค่าเศษเหลือ จากภาพที่ 4.8 จัดว่าเศษเหลือมีการกระจายตัวแบบสุ่ม ไม่ได้มีรูปแบบการกระจายตัวที่ระบุได้ชัดเจน

ข.3) Histogram ของค่าเศษเหลือ

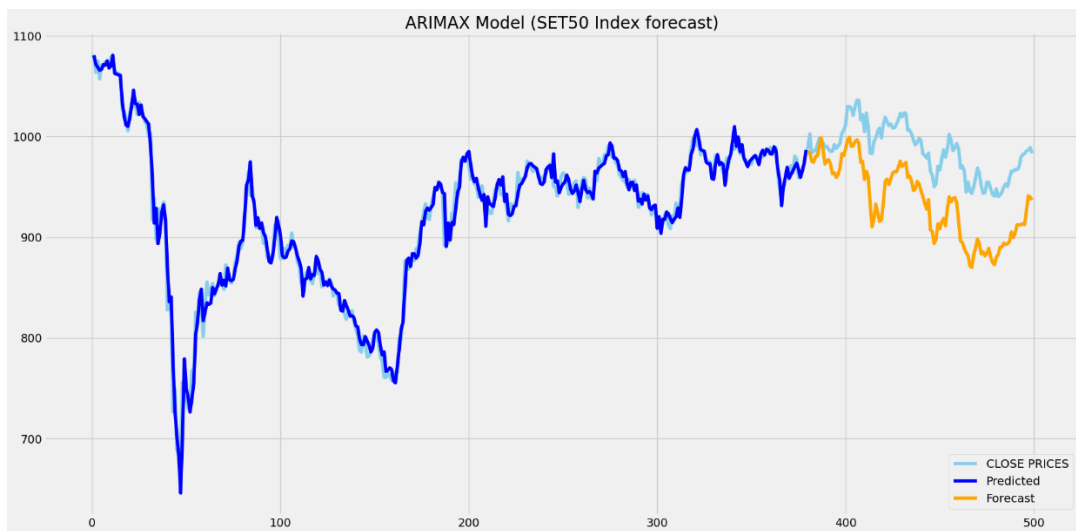


รูปที่ 4.8 Histogram ของค่าเศษเหลือ

จากภาพ Histogram ที่มีการลากเส้นโค้งตามแท่ง Histogram นั้น เป็นการตรวจสอบคุณสมบัติของค่าเศษเหลือว่ามีการแจกแจงแบบปกติหรือไม่ ซึ่งจากภาพที่ 4.8 พบว่าข้อมูลมีลักษณะการแจกแจงแบบปกติ

#### - การพยากรณ์ (Forecasting)

การพยากรณ์ราคาดัชนี SET50 โดยใช้ตัวแบบ ARIMAX(2,0,2) ล่วงหน้าเป็นระยะเวลา 120 วัน ให้ผลลัพธ์ดังภาพที่ 4.5



รูปที่ 4.9 กราฟแสดงผลการพยากรณ์ผ่านตัวแบบ ARIMAX (2,0,2) ทั้งข้อมูลชุดฝึกสอนและข้อมูลชุดทดสอบ

จากผลการพยากรณ์ข้อมูลดัชนี SET50 ด้วย ARIMAX (2,0,2) ทั้งข้อมูลช่วงฝึกสอนและข้อมูลช่วงทดสอบ โดยที่พยากรณ์ล่วงหน้าในส่วน of ข้อมูลช่วงทดสอบไปทั้งหมด 120 วัน ได้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 53.9357 บาท

#### 4.2.1.2 การสร้างตัวแบบ ANN

จากการทดลองสร้างตัวแบบ ANN แบบ Feed Forward Back Propagation จากข้อมูลชุดฝึกสอนโดยเปลี่ยนจำนวนข้อมูลเข้าและจำนวนโหนด (Node) เป็น 10, 20, 50, 80 และ 100 โหนด ในชั้นซ่อน (Hidden Layer) ทั้งหมด 2 ชั้น โดยใช้อัตราการเรียนรู้ 0.01 ฟังก์ชันถ่ายโอน ReLu ในชั้นซ่อนและฟังก์ชันเชิงเส้นในชั้นผลลัพธ์ และทำซ้ำ (epoch) จำนวน 5, 10, 50 และ 100 รอบ โดยมีการแบ่งข้อมูลเข้า (Batch) จำนวน 5, 10, 15 และ 20 ข้อมูล ซึ่งจะได้ตัวแบบรวมทั้งหมด 80 ตัวแบบ และใช้ MSE เป็นเกณฑ์ในการประเมิน

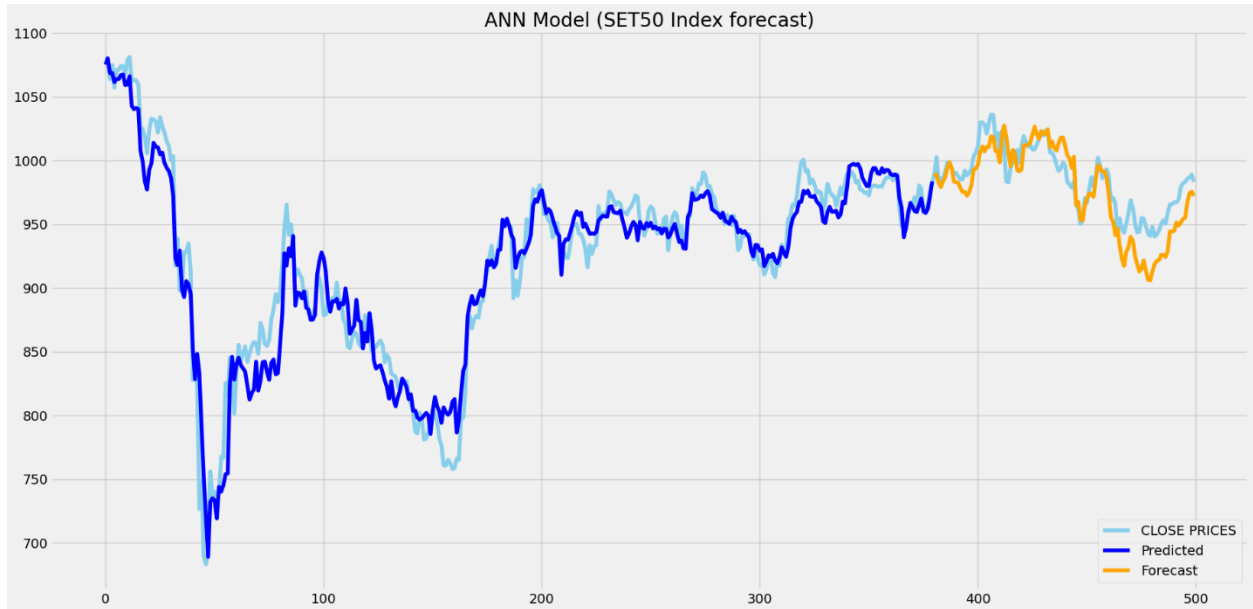
ตัวแบบ โดยทำการเลือกตัวแบบที่มีค่า MSE ต่ำสุดทั้งในชุดข้อมูลฝึกสอน และข้อมูลทดสอบ ผลการทดลองใช้ตัวแบบภายใต้เงื่อนไขดังกล่าวพยากรณ์ดังตารางที่ 4.9

**ตารางที่ 4.8** ความคลาดเคลื่อนในการพยากรณ์ตัวแบบ ANN ในแต่ละพารามิเตอร์

ตัวแบบ	โครงข่าย (Architecture)	Batch	Epochs	MSE	
				Train	Test
1	6-10-10-1	5	5	611581.4	935610.8
2	6-20-20-1	5	5	47906.37	749026
3	6-50-50-1	5	5	22255.72	466829.5
4	6-80-80-1	5	5	14129.18	300334
5	6-100-100-1	5	5	10319.24	222031.2
6	6-10-10-1	5	10	56517.09	802789.8
7	6-20-20-1	5	10	36434.94	726509.7
8	6-50-50-1	5	10	7109.109	206352.7
9	6-80-80-1	5	10	3227.543	74421.81
10	6-100-100-1	5	10	3089.567	45936.52
11	6-10-10-1	5	50	858.5405	7704.921
12	6-20-20-1	5	50	532.7755	2557.389
13	6-50-50-1	5	50	561.5124	2690.233
14	6-80-80-1	5	50	625.3553	1746.895
15	6-100-100-1	5	50	500.9991	1195.552
16	6-10-10-1	5	100	614.4652	119.0782
17	6-20-20-1	5	100	391.4654	689.1504
<b>18</b>	<b>6-50-50-1</b>	<b>5</b>	<b>100</b>	<b>386.3512</b>	<b>158.0942</b>
19	6-80-80-1	5	100	385.0472	476.1079
20	6-100-100-1	5	100	443.0436	718.7333
...	...	...	...	...	...
31	6-10-10-1	10	50	11723.03	351826.1
32	6-20-20-1	10	50	1662.139	34403.65
33	6-50-50-1	10	50	731.7002	10618.13

ตัวแบบ	โครงข่าย (Architecture)	Batch	Epochs	MSE	
				Train	Test
34	6-80-80-1	10	50	664.9447	8604.349
35	6-100-100-1	10	50	543.8333	5912.087
36	6-10-10-1	10	100	954.0899	8830.599
37	6-20-20-1	10	100	413.2957	984.418
38	6-50-50-1	10	100	463.7015	2361.607
39	6-80-80-1	10	100	351.7933	509.466
40	6-100-100-1	10	100	361.3639	815.6214
...	...	...	...	...	...
71	6-10-10-1	20	50	41063.01	724084.2
72	6-20-20-1	20	50	15833	420375.1
73	6-50-50-1	20	50	2722.777	63441.02
74	6-80-80-1	20	50	1583.726	25847.26
75	6-100-100-1	20	50	1509.157	19853.79
76	6-10-10-1	20	100	22498.21	584924.5
77	6-20-20-1	20	100	1572.788	36066.51
78	6-50-50-1	20	100	657.5922	8468.129
79	6-80-80-1	20	100	510.9195	5006.33
80	6-100-100-1	20	100	580.6265	3536.626

ผลการทดลองพบว่าการใช้ตัวแบบ ANN ที่มี Node = 50 , Batch = 5 และมีค่าของรอบการทำซ้ำ Epochs = 100 รอบ จะให้ค่า MSE ทั้งในข้อมูลฝึกสอน และข้อมูลทดสอบต่ำที่สุด เท่ากับ 386.3512 และ 158.0942 ตามลำดับ จึงเลือกใช้ตัวแบบ ANN (6-50-50-1) ที่มี batch เท่ากับ 5 ข้อมูลและ Epochs เท่ากับ 100 รอบเป็นตัวแบบที่เหมาะสม เพื่อใช้ในการพยากรณ์ข้อมูลดัชนี SET50 ต่อไป



รูปที่ 4.10 กราฟแสดงผลการพยากรณ์ผ่านตัวแบบ ANN (6-50-50-1) ทั้งข้อมูลชุดฝึกสอนและข้อมูลชุดทดสอบ

#### 4.2.1.3 การสร้างตัวแบบ SVR

จากการทดลองสร้างตัวแบบ SVM โดยค้นหาค่าพารามิเตอร์ทั้งสี่ เริ่มจากค่า  $c$  เท่ากับ 1,5,10,25,30,35,50,60,65,70,75,80,85,100,400,800,1500,2500 และ 3000 ค่า Epsilon เท่ากับ  $1e-5, 1e-4, 1e-3, 1e-2, 5e-2, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8$  และ 0.99 ค่า Gamma เป็น scale และ auto ส่วนพารามิเตอร์ kernel เป็น rbf และ linear ซึ่งจากค่าพารามิเตอร์ที่ต้องการหาตัวแบบที่เหมาะสม จะได้ตัวแบบทั้งหมดที่เป็นไปได้ 1065 ตัวแบบ โดยจะหาตัวแบบที่เหมาะสมที่สุดจากเกณฑ์ MSE ที่ต่ำที่สุดในชุดข้อมูลฝึกสอนและชุดข้อมูลทดสอบ ได้ผลลัพธ์ดังตารางที่ 4.10

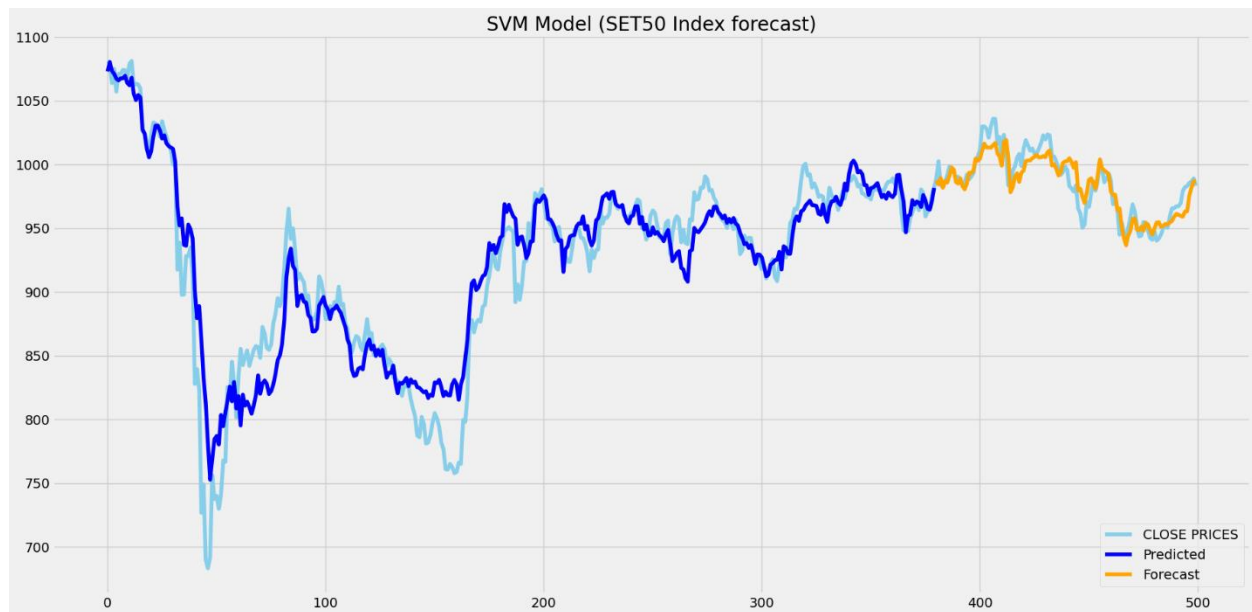
ตารางที่ 4.9 แสดงผลลัพธ์ของตัวแบบ SVR

ตัวแบบ	พารามิเตอร์				MSE	
	C	Epsilon	Gamma	Kernel	Train	Test
1	1	0.00001	scale	rbf	2699.909	539.9818
2	1	0.00001	scale	linear	679.1398	135.828
3	1	0.00001	auto	rbf	3261.549	652.3098
4	1	0.00001	auto	linear	679.1398	135.828
5	1	0.0001	scale	rbf	2699.91	539.9819
6	1	0.0001	scale	linear	679.1363	135.8273
7	1	0.0001	auto	rbf	3261.55	652.3099
8	1	0.0001	auto	linear	679.1363	135.8273
9	1	0.001	scale	rbf	2699.921	539.9841
10	1	0.00001	scale	rbf	679.1089	135.8218
...	...	...	...	...	...	...
50	1	0.8	scale	rbf	2697.759	539.5519
51	1	0.8	scale	linear	666.1167	133.2233
52	1	0.8	auto	rbf	3253.957	650.7915
53	1	0.8	auto	linear	666.1167	133.2233
54	1	0.99	scale	rbf	2695.679	539.1358
<b>55</b>	<b>1</b>	<b>0.99</b>	<b>scale</b>	<b>linear</b>	<b>630.9878</b>	<b>112.7196</b>
56	1	0.99	auto	rbf	3254.603	650.9206
57	1	0.99	auto	linear	663.6848	132.737
58	5	0.00001	scale	rbf	2020.545	404.109
59	5	0.00001	scale	linear	711.6431	142.3286
60	5	0.00001	auto	rbf	2268.649	453.7298
...	...	...	...	...	...	...
1056	3500	0.7	auto	rbf	4814.132	962.8265
1057	3500	0.7	auto	linear	734.0593	146.8119
1058	3500	0.8	scale	rbf	3539.963	707.9925



ตัวแบบ	พารามิเตอร์				MSE	
	C	Epsilon	Gamma	Kernel	Train	Test
1059	3500	0.8	scale	linear	732.5369	146.5074
1060	3500	0.8	auto	rbf	4715.94	943.188
1061	3500	0.8	auto	linear	732.5369	146.5074
1062	3500	0.99	scale	rbf	3498.105	699.621
1063	3500	0.99	scale	linear	740.1529	148.0306
1064	3500	0.99	auto	rbf	4636.69	927.3381
1065	3500	0.99	auto	linear	740.1529	148.0306

ผลการทดลองพบว่าการใช้ตัวแบบ SVR ที่มีพารามิเตอร์  $C=1$ ,  $Epsilon=0.99$ ,  $Gamma='scale'$  และ  $kernel='linear'$  จะให้ค่า MSE ที่ต่ำที่สุดในข้อมูลชุดฝึกสอนและข้อมูลชุดทดสอบ โดยมีค่า MSE เท่ากับ 630.9878 และ 112.7196 ตามลำดับ จึงเลือกใช้ตัวแบบ SVM นี้ในการพยากรณ์ราคาดัชนี SET50 ต่อไป



รูปที่ 4.10 กราฟแสดงผลการพยากรณ์ผ่านตัวแบบ SVR ที่คัดเลือกกว่าเหมาะสม ทั้งข้อมูลชุดฝึกสอนและข้อมูลชุดทดสอบ

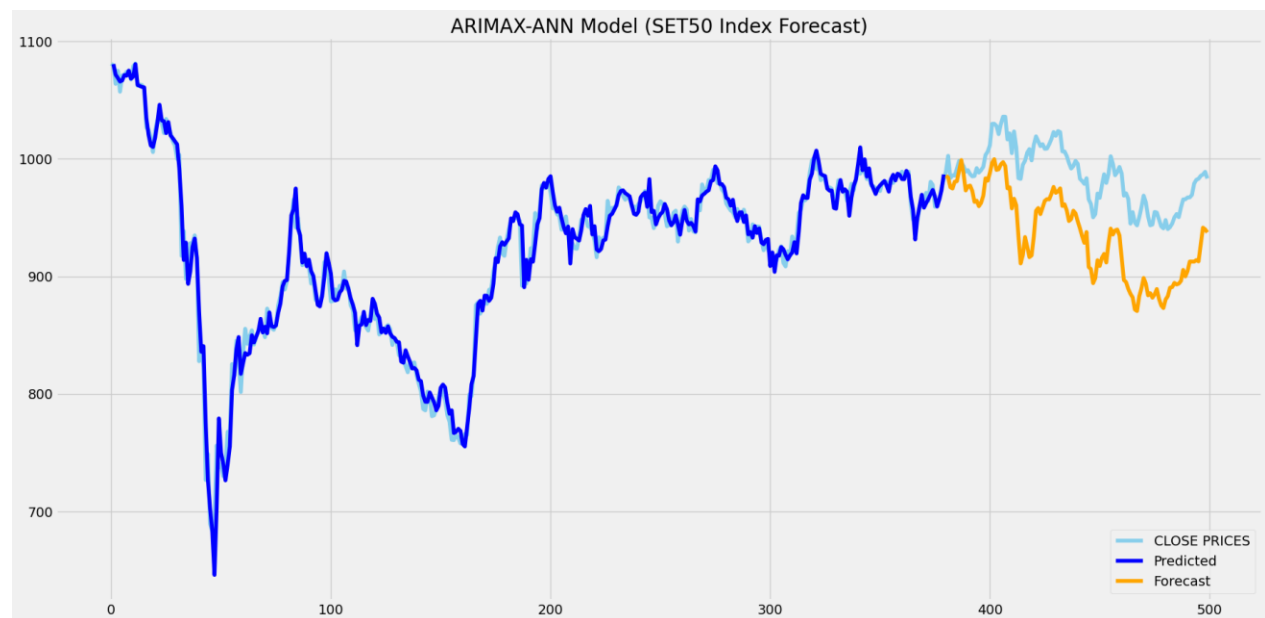
## 4.2.2 การสร้างตัวแบบผสม (Traditional Hybrid Model)

### 4.2.2.1 การสร้างตัวแบบ ARIMAX-ANN

จากการนำอนุกรมเวลาของส่วนเหลือที่ได้จากตัวแบบ ARIMAX(2,0,2) มาสร้างตัวแบบ ANN สำหรับพยากรณ์ส่วนประกอบที่ไม่เป็นเชิงเส้นของส่วนเหลือ ซึ่งจากการทดลอง พบว่าตัวแบบที่ใช้จำนวน โหนดในชั้นซ่อนแต่ละชั้นซ่อน 50 โหนด มี batch เท่ากับ 20 และทำซ้ำ 5 ครั้งจะให้ตัวแบบ ANN ที่เหมาะสม จากนั้นนำค่าพยากรณ์ส่วนที่ได้จาก ANN ไปรวมกับค่าพยากรณ์ที่ได้จาก ARIMAX และคำนวณค่าความคลาดเคลื่อน รวมถึงกราฟเปรียบเทียบระหว่างข้อมูลจริงและข้อมูลพยากรณ์ แสดงได้ผลดังตารางที่ 4.10 และรูปที่ 4.11

ตารางที่ 4.10 ความคลาดเคลื่อนในการพยากรณ์สำหรับตัวแบบผสมระหว่าง ARIMAX และ ANN

โครงข่าย (Architecture)	Batch	Epochs	MSE	
			Train	Test
6-50-50-1	20	5	125.5141	2862.3666



รูปที่ 4.11 กราฟแสดงผลการพยากรณ์ผ่านตัวแบบ ARIMAX-ANN ทั้งข้อมูลชุดฝึกสอนและข้อมูลชุดทดสอบ

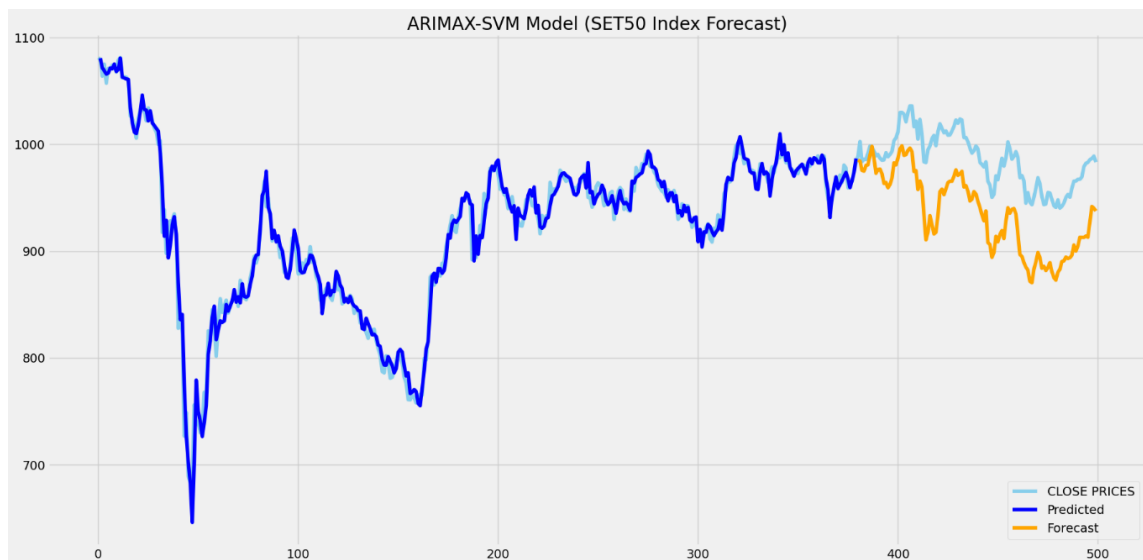
จากผลการพยากรณ์ข้อมูลดัชนี SET50 ด้วย ARIMAX-ANN ทั้งข้อมูลช่วงฝึกสอนและข้อมูลช่วงทดสอบ โดยที่พยากรณ์ล่วงหน้าในส่วนของคุณข้อมูลช่วงทดสอบไปทั้งหมด 120 วัน ได้ค่าความคลาดเคลื่อนเฉลี่ย (MSE) เท่ากับ 2862.3666 บาท<sup>2</sup> ซึ่งดีกว่าตัวแบบเดี่ยว ARIMAX เพียงเล็กน้อย

#### 4.2.2.2 การสร้างตัวแบบ ARIMAX-SVR

จากการนำอนุกรมเวลาของส่วนเหลือที่ได้จากตัวแบบ ARIMAX(2,0,2) มาสร้างตัวแบบ SVR สำหรับพยากรณ์ส่วนประกอบที่ไม่เป็นเชิงเส้นของส่วนเหลือ ซึ่งจากการทดลอง พบว่าตัวแบบที่ใช้พารามิเตอร์ C เท่ากับ 1 , Epsilon เท่ากับ 0.99 , Gamma เป็นแบบ scale และ Kernel เป็นแบบ linear จะให้ตัวแบบ SVR ที่เหมาะสมกับข้อมูลค่าส่วนเหลือที่ได้จากตัวแบบ ARIMAX จากนั้นนำค่าพยากรณ์ส่วนที่ได้จาก SVR ไปรวมกับค่าพยากรณ์ที่ได้จาก ARIMAX และคำนวณค่าความคลาดเคลื่อน รวมถึงกราฟเปรียบเทียบระหว่างข้อมูลจริงและข้อมูลพยากรณ์ แสดงได้ผลดังตารางที่ 4.11 และภาพที่ 4.12

ตารางที่ 4.11 ความคลาดเคลื่อนในการพยากรณ์สำหรับตัวแบบผสมระหว่าง ARIMAX และ SVR

พารามิเตอร์				MSE	
C	Epsilon	Gamma	Kernel	Train	Test
1	0.99	auto	linear	121.2151	2876.2779



รูปที่ 4.12 กราฟแสดงผลการพยากรณ์ผ่านตัวแบบ ARIMAX-SVR ทั้งข้อมูลชุดฝึกสอนและข้อมูลชุดทดสอบ

จากผลการพยากรณ์ข้อมูลดัชนี SET50 ด้วย ARIMAX-SVM ทั้งข้อมูลช่วงฝึกสอนและข้อมูลช่วงทดสอบ โดยที่พยากรณ์ล่วงหน้าในส่วนของคุณข้อมูลช่วงทดสอบไปทั้งหมด 120 วัน ได้ค่าความคลาดเคลื่อนเฉลี่ย (MSE) เท่ากับ 2876.2779 บาท<sup>2</sup> ซึ่งดีกว่าตัวแบบเดี่ยว ARIMAX เพียงเล็กน้อยเช่นเดียวกัน

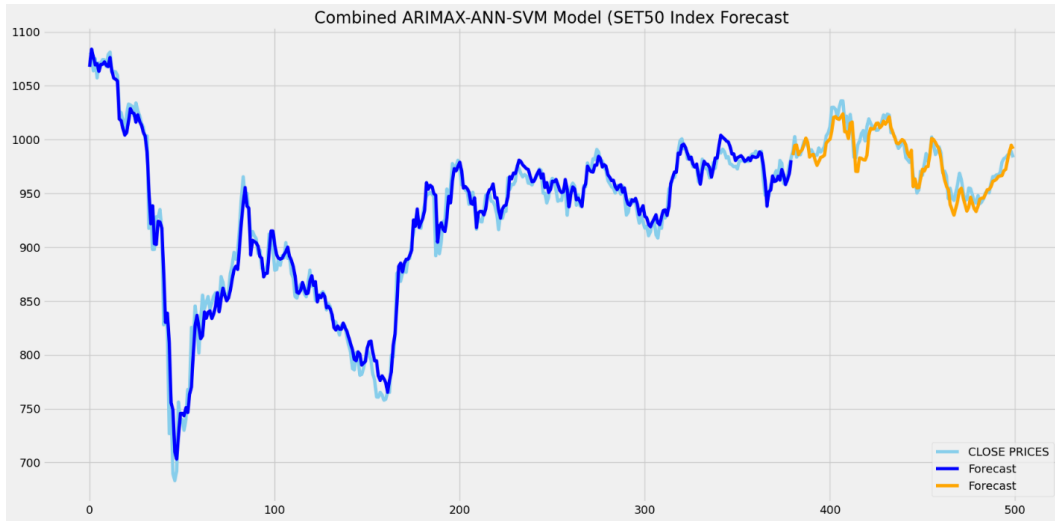
#### 4.2.3 การสร้างตัวแบบรวม (Hybrid Combined Model)

##### 4.2.3.1 การสร้างตัวแบบ ARIMAX-ANN-SVM

การสร้างตัวแบบรวมระหว่าง ARIMAX ANN และ SVM จะนำค่าพยากรณ์ของแต่ละตัวแบบพยากรณ์ได้มารวมกันโดยใส่ค่า weight ในแต่ละตัวแบบ และทำการค้นหาค่า weights ที่ดีที่สุดในข้อมูลชุดทดสอบภายใต้เงื่อนไขค่า MSE มีค่าต่ำที่สุด และค่า weight ต้องรวมกันเท่ากับ 1 พอดี โดยค่า weights ที่ใช้ในตัวแบบ Combined จะขึ้นอยู่กับจำนวนตัวแบบที่นำมาใช้ สำหรับงานวิจัยชิ้นนี้นำเสนอตัวแบบ Combined (ARIMAX-ANN-SVM) ดังนั้นจะมีค่า weights จำนวน 3 ค่าคือ ค่า Weight ของ ARIMAX, ค่า Weight ของ ANN และค่า Weight ของ SVR ทำการค้นหาชุดของค่า Weights ทั้งสามที่ดีที่สุดที่ทำให้ค่า MSE ต่ำที่สุด ผ่านวิธี Nelder-Mead หรือ "Nelder-Mead simplex method" เป็นวิธีหาค่าต่ำสุดของฟังก์ชัน ในที่นี้กำหนดเป็นค่า MSE ซึ่งวิธีนี้ไม่ต้องการการเชิงเส้น (nonlinear optimization method) ซึ่งใช้เพียงแค่ค่าของฟังก์ชันเท่านั้นในการหาค่าต่ำสุด โดยผลลัพธ์ที่ผ่านวิธี Nelder-Mead จะทำให้ได้ค่า Weight กับแต่ละตัวแบบ ที่ให้ค่า MSE ต่ำที่สุดดังแสดงในตารางที่ 4.13

ตารางที่ 4.13 ความคลาดเคลื่อนในการพยากรณ์สำหรับตัวแบบรวม ARIMAX-ANN-SVM

Weight			MSE	
ARIMAX	ANN	SVM	Train	Test
0.1780	0.4873	0.3347	162.9269	85.6583



รูปที่ 4.13 กราฟแสดงผลการพยากรณ์ผ่านตัวแบบรวม ARIMAX-ANN-SVM ทั้งข้อมูลชุดฝึกสอนและข้อมูลชุดทดสอบ

#### 4.3 การคำนวณหาค่าความคลาดเคลื่อน และการคัดเลือกตัวแบบ

เปรียบเทียบค่า RMSE, MSE, MAE และ MAPE ในข้อมูลชุดทดสอบจำนวน 120 ข้อมูล ทั้ง 6 ตัวแบบ รวมทั้งค่าพารามิเตอร์ที่ดีที่สุดในการหาค้นหาของแต่ละตัวแบบ ได้ผลสรุปดังตารางที่ 4.13

ตารางที่ 4.14 แสดงค่าพารามิเตอร์ของตัวแบบ,ค่า RMSE,ค่า MSE,ค่า MAE และ ค่า MAPE ที่เหมาะสมกับข้อมูลดัชนี SET50

Model	Parameter				RMSE	MSE	MAE	MAPE
ARIMAX	p	d	q		53.9357	2909.055	50.2114	5.1149
	2	0	2					
ANN	Node	Batch	Epochs		14.5343	211.2465	11.44	1.1733
	50	5	100					
SVM	C	Epsilon	Gamma	Kernel	14.8255	219.7949	12.7375	1.2933
	1	0.99	auto	linear				
ARIMA-ANN	Node	Batch	Epochs		53.6309	2876.278	49.9397	5.0871
	50	20	5					
ARIMA-SVM	C	Epsilon	Gamma	Kernel	53.5011	2862.367	49.7652	5.0697
	1	0.99	auto	linear				

Model	Parameter			RMSE	MSE	MAE	MAPE
Hybrid combined	Weight						
	ARIMAX	ANN	SVM	<u>9.2552</u>	<u>85.6583</u>	<u>7.3386</u>	<u>0.7432</u>
	0.178	0.4873	0.3347				

หมายเหตุ RMSE, MSE, MAE และ MAPE มีหน่วยเป็น บาท, บาท<sup>2</sup>, บาท<sup>2</sup> และ % ตามลำดับ

จากผลการทดลองสร้างตัวแบบทั้งหมด 6 ตัวแบบ โดยใช้ข้อมูลฝึกสอนจำนวน 380 ข้อมูล เพื่อหาค่าพารามิเตอร์ที่ดีที่สุดในแต่ละตัวแบบ และใช้ข้อมูลทดสอบจำนวน 120 ข้อมูล เพื่อพยากรณ์และเปรียบเทียบประสิทธิภาพของตัวแบบว่ามีความแม่นยำในการพยากรณ์มากน้อยเพียงใด

ในส่วนของตัวแบบเดี่ยว พบว่าตัวแบบ ARIMAX มีค่าความคลาดเคลื่อนในการพยากรณ์มากที่สุด หรือมีความแม่นยำต่ำที่สุด ซึ่งมีค่า RMSE, MSE, MAE และ MAPE เท่ากับ 53.9357บาท, 2909.055บาท<sup>2</sup>, 50.2114บาท<sup>2</sup> และ 5.1149% ตามลำดับ ในขณะที่ตัวแบบ ANN และ SVR มีค่าความคลาดเคลื่อนในการพยากรณ์ที่น้อยกว่า ARIMAX ทั้งคู่ โดยที่ ANN มีค่า RMSE, MSE, MAE และ MAPE เท่ากับ 14.5343บาท, 211.2465บาท<sup>2</sup>, 11.44บาท<sup>2</sup> และ 1.1733% ตามลำดับ และ SVR มีค่า RMSE, MSE, MAE และ MAPE เท่ากับ 14.8255บาท, 219.7949บาท<sup>2</sup>, 12.7375บาท<sup>2</sup> และ 1.2933% ตามลำดับ ซึ่งถ้าเปรียบเทียบระหว่างตัวแบบ ANN กับ SVR ที่เป็นตัวแบบเดี่ยวเหมือนกันจะเห็นได้ว่าตัวแบบ ANN จะมีความแม่นยำกว่าตัวแบบ SVR แต่เพียงเล็กน้อยเท่านั้น

ในส่วนของตัวแบบผสม ระหว่าง ARIMAX-ANN และ ARIMAX-SVR พบว่าตัวแบบทั้งคู่ให้ค่าความคลาดเคลื่อนที่ต่ำกว่าตัวแบบเดี่ยวเดิม หรือตัวแบบ ARIMAX เพียงเล็กน้อยเท่านั้น ซึ่งตัวแบบ ARIMAX-ANN ให้ค่า RMSE, MSE, MAE และ MAPE เท่ากับ 53.6309บาท, 2876.278บาท<sup>2</sup>, 49.9397บาท<sup>2</sup> และ 5.0871% ตามลำดับ และตัวแบบ ARIMAX-SVR ให้ค่า RMSE, MSE, MAE และ MAPE เท่ากับ 53.5011บาท, 2862.367บาท<sup>2</sup>, 49.7652บาท<sup>2</sup> และ 5.0697% ตามลำดับ ซึ่งจะเห็นได้ว่าตัวแบบผสม ARIMAX-SVR มีความแม่นยำในการพยากรณ์ดีกว่าตัวแบบผสม ARIMAR-ANN แต่เพียงเล็กน้อยเท่านั้น

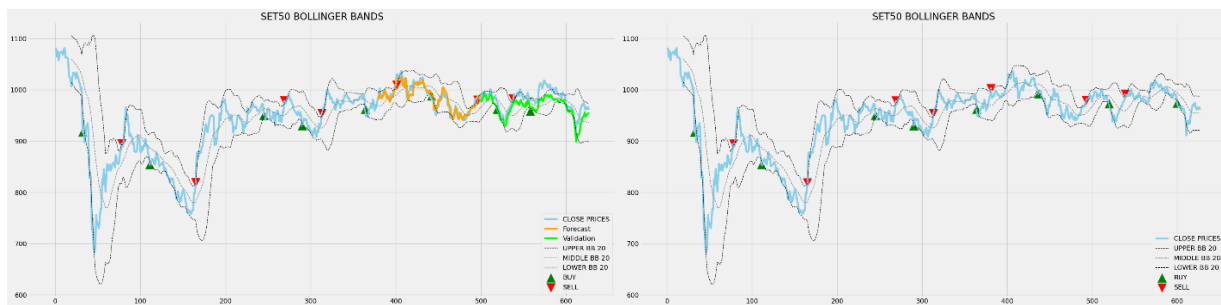
และตัวแบบสุดท้ายตัวแบบรวมระหว่าง ARIMAX-ANN-SVR พบว่าตัวแบบรวมดังกล่าวให้ค่าความคลาดเคลื่อนที่ต่ำที่สุดจากทั้งหมด 6 ตัวแบบ ซึ่งมีค่า RMSE, MSE, MAE และ MAPE เท่ากับ 9.2552บาท, 85.6583บาท<sup>2</sup>, 7.3386บาท<sup>2</sup> และ 0.7432% ตามลำดับ ซึ่งเห็นได้ว่าตัวแบบรวมเป็นตัวแบบที่มีประสิทธิภาพความแม่นยำในการพยากรณ์ข้อมูลดัชนีราคา SET50 มากที่สุด จึงเลือกตัวแบบรวมระหว่าง ARIMAX-ANN-SVR เพื่อนำมาใช้ในการพยากรณ์สัญญาณการซื้อ-ขายล่วงหน้า โดยใช้เครื่องมือ Bollinger Band ในการคำนวณหา

สัญญาณการซื้อ-ขายล่วงหน้า และนำมาเปรียบเทียบกับเครื่องมือ Exponential Smoothing Average หรือ EMA ที่ใช้ข้อมูลดัชนีราคา SET50 เดิมในขั้นตอนต่อไป

#### 4.4 การพยากรณ์สัญญาณการซื้อ-ขายล่วงหน้าและการวัดประสิทธิภาพ

##### 4.4.1 การเปรียบเทียบประสิทธิภาพด้วยอัตรากำไร

จากการทดลองจะเป็นตัวแบบที่มีความแม่นยำในการพยากรณ์มากที่สุดทำให้ค่าพยากรณ์ที่ได้ อาจมีความใกล้เคียงกับข้อมูลจริงทำให้สัญญาณการซื้อ-ขายที่ได้จากเครื่องมือ Bollinger Band จะไม่ค่อยมีความแตกต่างกันมากนัก ในเรื่องของจุดเวลาการซื้อ-ขาย ดังนั้นการเปรียบเทียบประสิทธิภาพจึงจำเป็นต้องเปรียบเทียบด้วยอัตรากำไร ซึ่งได้ผลลัพธ์ดังภาพ



รูปที่ 4.14 กราฟแสดงผลการเปรียบเทียบผลลัพธ์ระหว่างสัญญาณซื้อ-ขายของ Bollinger band ที่ได้จากข้อมูลพยากรณ์กับข้อมูลจริง

อัตรากำไรที่ได้จากตัวแบบพยากรณ์ ARIMAX-ANN-SVR ร่วมกับเครื่องมือ Bollinger Band เท่ากับ -12% ในขณะที่อัตรากำไรที่ได้จากข้อมูลดัชนี SET50 ร่วมกับเครื่องมือ Bollinger Band เท่ากับ -14% ซึ่งสรุปได้ว่าอัตรากำไรที่ได้จากตัวแบบพยากรณ์ ARIMAX-ANN-SVR ร่วมกับเครื่องมือ Bollinger Band มีอัตรากำไรมากกว่าอยู่ 2% โดยที่สัญญาณการซื้อ-ขายมีความใกล้เคียงกัน ไม่แตกต่างกันมากนัก ตามที่คาดการณ์ไว้

##### 4.4.2 การเปรียบเทียบประสิทธิภาพด้วยอัตราซื้อขายสำเร็จ

การเปรียบเทียบประสิทธิภาพด้วยอัตราซื้อขายสำเร็จจะเปรียบเทียบระหว่างสัญญาณที่ได้จากเครื่องมือ Moving Average กับข้อมูลดัชนี SET50 จริง กับสัญญาณที่ได้จากเครื่องมือ Bollinger Band กับข้อมูลดัชนี SET50 ที่ได้จากการพยากรณ์จากตัวแบบ ARIMAX-ANN-SVR



รูปที่ 4.15 กราฟแสดงผลการเปรียบเทียบผลลัพธ์ระหว่างสัญญาณซื้อ-ขายของ Bollinger band ที่ได้จากข้อมูลพยากรณ์กับสัญญาณซื้อ-ขายของเครื่องมือ Moving Average ที่ได้จากข้อมูลจริง

อัตราการซื้อขายสำเร็จที่ได้จากตัวแบบพยากรณ์ ARIMAX-ANN-SVR ร่วมกับเครื่องมือ Bollinger Band เท่ากับ 0.5714 ในขณะที่อัตราการซื้อขายสำเร็จที่ได้จากข้อมูลดัชนี SET50 เดิมร่วมกับเครื่องมือ Moving Average เท่ากับ 0.4000 ซึ่งสรุปได้ว่า สัญญาณที่ได้จากตัวแบบพยากรณ์ ARIMAX-ANN-SVR ร่วมกับเครื่องมือ Bollinger Band สามารถให้สัญญาณการซื้อ-ขายที่มีอัตราซื้อขายสำเร็จมากกว่าสัญญาณการซื้อ-ขายที่ได้จากข้อมูลดัชนี SET50 เดิม ร่วมกับเครื่องมือ Moving Average



## บทที่ 5

### สรุปผลการวิจัยและข้อเสนอแนะ

การศึกษาวิธีการพยากรณ์ ดัชนีราคาหุ้นตลาดหลักทรัพย์แห่งประเทศไทย ดัชนีราคา SET50 ในครั้งนี้ เป็นการเปรียบเทียบวิธีการพยากรณ์ 6 วิธี ได้แก่ วิธีพยากรณ์จากตัวแบบเดี่ยว ซึ่งก็คือตัวแบบ ARIMAX, ANN และ SVR รวมเป็น 3 ตัวแบบในประเภทตัวแบบเดี่ยว และวิธีพยากรณ์จากตัวแบบผสม ซึ่งก็คือตัวแบบ ARIMAX-ANN และ ARIMAX-SVR รวมเป็น 2 ตัวแบบ และอีกหนึ่งตัวแบบคือตัวแบบรวม ARIMAX-ANN-SVR ซึ่งจะรวมกันได้ทั้งหมด 6 ตัวแบบ และแบ่งข้อมูลที่ใช้ในงานวิจัยออกเป็นทั้งหมด 3 ส่วน ได้แก่ ส่วนข้อมูลฝึกสอน ส่วนข้อมูลทดสอบ และส่วนข้อมูลทำนายเพื่อเปรียบเทียบการพยากรณ์สัญญาณซื้อ-ขายล่วงหน้าระหว่างวิธี Bollinger Band และ วิธี Moving Average (MA) โดยข้อมูลส่วนที่หนึ่งรวบรวมข้อมูลตั้งแต่วันที่ 1 มกราคม พ.ศ.2563 ถึงวันที่ 28 ธันวาคม พ.ศ. 2564 ในการสร้างตัวแบบ และใช้ข้อมูลส่วนที่สองตั้งแต่วันที่ 29 ธันวาคม พ.ศ.2564 ถึงวันที่ 19 สิงหาคม พ.ศ. 2565 ในการทดสอบตัวแบบ และสุดท้ายใช้ข้อมูลส่วนที่สามตั้งแต่วันที่ 22 สิงหาคม พ.ศ.2565 ถึงวันที่ 18 เมษายน พ.ศ. 2566 ในการเปรียบเทียบสัญญาณซื้อ-ขายล่วงหน้าระหว่างวิธี Bollinger Band และ วิธี Moving Average

#### 5.1 สรุปผลการวิจัย

ตัวแบบ Hybrid combined ที่นำเสนอในงานวิจัยชิ้นนี้ให้ผลความแม่นยำเหนือกว่าตัวแบบอื่น ทั้งตัวแบบ Hybrid และตัวแบบเดี่ยวเกือบทุกชุดข้อมูล ตัวแบบ Combined ที่นำเสนอนี้ใช้ตัวแบบเดี่ยวๆ คือ ARIMA, ANN และ SVM ซึ่งเป็นตัวแบบเดี่ยวที่โดดเด่นในแง่ของของแต่ละตัวแบบ ตัวแบบ ARIMA จับลักษณะความสัมพันธ์ที่เป็นเชิงเส้นตรงของข้อมูลได้ดี ขณะที่ตัวแบบ ANN และ SVR เป็น ตัวแบบที่จับลักษณะความสัมพันธ์ที่ไม่เป็นเชิงเส้นตรงของข้อมูลได้ดีและมีวิธีการสร้างตัวแบบที่ แตกต่างกันไป เมื่อนำมา Combined กันก็ช่วยให้จับลักษณะข้อมูลทั้งที่เป็นเชิงเส้นตรงและไม่เป็นเชิง เส้นตรงได้ดียิ่งขึ้น จัดเป็นตัวแบบที่ช่วยลดความผิดพลาดในการเลือกตัวแบบ ซึ่งอาจเลือกตัวแบบที่ไม่เหมาะสม สามารถผสมตัวแบบได้มากกว่าสองตัวแบบ โดยที่กระบวนการไม่ยุ่งยากมากนัก กล่าวคือ สามารถพยากรณ์ข้อมูลทั้งสามตัวแบบพร้อมๆ กัน แล้วนำเข้า Combined จึงเหมาะสมเป็นอย่างยิ่งที่จะนำตัวแบบนี้ไปใช้ในพยากรณ์ข้อมูลที่มีความซับซ้อน ข้อมูลที่ยากแก่การพิจารณาว่าควรใช้ตัวแบบใด และต้องการความแม่นยำในการพยากรณ์สูง

ในส่วนของการวัดประสิทธิภาพการพยากรณ์สัญญาณซื้อ-ขายล่วงหน้าจากการใช้ตัวแบบพยากรณ์ที่ดีที่สุดร่วมกับเครื่องมือ Bollinger Band พบว่าให้ผลลัพธ์ที่ดีกว่าสัญญาณซื้อ-ขายที่ได้จากข้อมูลดัชนี SET50 ที่เป็น

ข้อมูลจริง ซึ่งให้สัญญาณซื้อ-ขายที่ใกล้เคียงความเป็นจริง และให้อัตรากำไรดีกว่าไม่มากนัก และในส่วนการเปรียบเทียบกับเครื่องมือ Moving Average ซึ่งเป็นเครื่องมือที่ใช้หาสัญญาณซื้อ-ขายเช่นเดียวกับ Bollinger Band ก็พบว่าสัญญาณซื้อ-ขายล่วงหน้าที่ได้จากตัวแบบพยากรณ์ที่ดีที่สุดร่วมกับเครื่องมือ Bollinger Band ให้อัตราการซื้อขายสำเร็จที่มากกว่าการใช้ข้อมูลดัชนี SET50 ที่เป็นข้อมูลจริงร่วมกับเครื่องมือ Moving Average

ซึ่งผลลัพธ์ทั้งหมดเป็นเพียงผลลัพธ์จากช่วงของข้อมูลดัชนี SET50 เพียงบางช่วงเวลา ซึ่งเป็นช่วงขาของหุ้น ถ้าหากวัดผลลัพธ์จากหลายๆช่วงของหุ้นที่เป็น ช่วงขาขึ้น หรือ ช่วง sideway ด้วยจะได้ผลลัพธ์ที่แตกต่างกันออกไป เนื่องจากเครื่องมือทางเทคนิคในงานวิจัยนี้ใช้ได้ไม่ค่อยดีในช่วงหุ้นที่เป็นขาลงทำให้เห็นว่าอัตรากำไรที่ได้มีค่าติดลบ

## 5.2 ข้อเสนอแนะและแนวทางในการทำวิจัยในขั้นถัดไป

1. ในงานวิจัยนี้ทำการเปรียบเทียบความแม่นยำของค่าพยากรณ์เพียง 6 ตัวแบบ คือตัวแบบเดี่ยว ARIMAX, ANN, SVR, ตัวแบบผสมระหว่างตัวแบบ ARIMAX-ANN, ARIMAX-SVR และตัวแบบรวม ARIMAX-ANN-SVR ซึ่งผู้วิจัยอื่นอาจทดลองทำการเปรียบเทียบตัวแบบผสมอื่นๆที่น่าสนใจ เช่น ตัวแบบผสม Pegels-ARIMAX, ตัวแบบผสม ARIMAX กับ Adaptive neuro-fuzzy inference system (ANFIS), ตัวแบบผสมระหว่าง ARIMAX-LSTM หรือตัวแบบผสมระหว่างตัวแบบ ARIMAX และการถดถอยโพลิโนเมียล เป็นต้น

2. การสร้างตัวแบบเดี่ยว ANN และตัวแบบผสม ARIMAX-ANN ซึ่งมีส่วนประกอบของตัวแบบโครงข่ายประสาทเทียม ในงานวิจัยนี้เลือกใช้เทคนิคการฝึกสอนโครงข่ายแบบ Feedforward ซึ่งผู้วิจัยอื่นๆที่สนใจศึกษาอาจเลือกใช้เทคนิคการฝึกสอนโครงข่ายแบบอื่นๆ เช่น Levenberg-Marquardt, BFGS Quasi-Newton หรือ Gradient Descent with Momentum เป็นต้น

3. การสร้างตัวแบบเดี่ยว SVR และการสร้างตัวแบบผสมระหว่างตัวแบบ ARIMAX-SVR ซึ่งเป็นส่วนประกอบของตัวแบบซัพพอร์ทเวกเตอร์แมชชีน ในงานวิจัยนี้เลือกใช้ kernel function คือ Gaussian radial basis function (RBF) กับ Linear function ซึ่งผู้วิจัยอื่นๆอาจเลือกใช้ kernel function ประเภทอื่นๆได้ เช่น Polynomial kernel หรือ Sigmoid kernel เป็นต้น ทั้งนี้ขึ้นกับลักษณะของข้อมูลด้วย หรืออาจทดลองเพิ่มจำนวนโหนดที่ใช้ในชั้นซ่อนเพื่อหาจำนวนโหนดที่เหมาะสมมากขึ้น นอกจากนี้การทดลองปรับจูนหาชุดของค่าพารามิเตอร์ของตัวแบบซัพพอร์ทเวกเตอร์แมชชีน คือ  $c$  และ  $\epsilon$  ชุดอื่นๆที่ต่างออกไป ก็อาจจะทำให้ผลลัพธ์ในการพยากรณ์ค่ามีความแม่นยำมากยิ่งขึ้น

4. ในงานวิจัยได้นั้นผลลัพธ์การพยากรณ์โดยวัดจากความแม่นยำของตัวแบบพยากรณ์ว่ามีความใกล้เคียงกับค่าจริงมากน้อยเพียงใดแล้วนำไปหาสัญญาณซื้อ-ขายล่วงหน้า ซึ่งถ้าหากต้องการเน้นผลลัพธ์ในการพยากรณ์สัญญาณการซื้อ-ขายที่ดีที่สุด ผู้วิจัยอื่นๆที่สนใจอาจเลือกการวัดประสิทธิภาพของตัวแบบจากการทำนายค่า

พยากรณ์ให้อัตราซื้อขายสำเร็จสูงสุด แล้วค่อยปรับจนพารามิเตอร์จากผลลัพธ์ของการพยากรณ์สัญญาณซื้อขายแทนให้อัตราซื้อขายสำเร็จสูงสุด เป็นต้น

## บรรณานุกรม

ภาษาไทย

### 1.หนังสือ

กฤตยา ตติรังสรรค์สุข. (2547). เศรษฐศาสตร์มหภาคเบื้องต้น (พิมพ์ครั้งที่ 4). สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย. กรุงเทพมหานคร

สนธิ อังสนากุล. (2547). มหัตถรรพ์แห่งเทคนิค : เพื่อให้เข้าถึงหัวใจ (พิมพ์ครั้งที่ 1). กรุงเทพมหานคร

### 2.วิทยานิพนธ์

ธัญชนก จันทรหอม. (2564). การพยากรณ์แบบอนุกรมเวลา เพื่อกำหนดการสั่งซื้อที่ประหยัดที่สุดของโรงงานผลิตยางซิลิโคนแห่งหนึ่ง (TIME SERIES FORECASTING FOR DETERMINING ECONOMIC ORDER QUANTITY OF SILICON RUBBER MANUFACTURER) (วิทยานิพนธ์ บริหารธุรกิจมหาบัณฑิต). มหาวิทยาลัยศิลปากร.

ภัทร วรภู. (2556). การเปรียบเทียบความแม่นยำของการพยากรณ์อนุกรมเวลาระหว่างตัวแบบผสมและตัวแบบเดี่ยว (COMPARISON OF TIME SERIES FORECASTING ACCURACY BETWEEN THE HYBRID AND INDIVIDUAL MODEL) (วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต). จุฬาลงกรณ์มหาวิทยาลัย.

วชิราภรณ์ แก้วมาตย์. (2556). การพยากรณ์ดัชนีตลาดหลักทรัพย์ของกลุ่มประเทศอาเซียนโดยใช้โครงข่ายประสาทเทียม. (วิทยานิพนธ์เศรษฐศาสตร์มหาบัณฑิต). สาขาวิชาเศรษฐศาสตร์ธุรกิจ, มหาวิทยาลัยขอนแก่น.

### 3.สารนิพนธ์

นายพงษ์ศิริ ศิริพานิช (2550). การพยากรณ์อนุกรมเวลาด้วยตัวแบบผสมระหว่าง ARIMA และเครือข่ายประสาทเทียม (สารนิพนธ์ วิทยาศาสตร์มหาบัณฑิต). มหาวิทยาลัยศิลปากร.

ภูวิศ บรรจงพัฒนา (2564). การศึกษาและเปรียบเทียบตัวแบบการพยากรณ์ ARIMA, ARIMAX, และ GARCH ต่อดัชนี SET 50 ในช่วงก่อน-หลังการเกิดวิกฤตการณ์โรคระบาดโควิด-19 (วิทยาศาสตร์มหาบัณฑิต (การบัญชีและการบริหารการเงิน)). มหาวิทยาลัยธรรมศาสตร์.

### 4.บทความวิจัย

กรอภิชิต เหม้าพันธ์ และ สมพร ปันโกษา (2562). การพยากรณ์ความผันผวนของตลาดหลักทรัพย์โดยใช้

ซัพพอร์ตเวกเตอร์แมชชีน (FORCASTING STOCK MARKET VOLATILITY WITH SUPPORT VECTOR MACHINE) วิทยาศาสตร์มหาบัณฑิต. สาขาวิศวกรรมการเงิน คณะวิทยาศาสตร์ มหาวิทยาลัยหอการค้าไทย, 11-18.

นัท กุลวานิช (2563). การเพิ่มความแม่นยำในการพยากรณ์ข้อมูลอนุกรมเวลาด้วยตัวแบบ

ผสม ARIMA-ANN และการวิเคราะห์การถดถอย (Improving Prediction Accuracy of Time Series Data Using ARIMA-ANN Hybrid Model and Regression Analysis). ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี. จุฬาลงกรณ์มหาวิทยาลัย, 2102-2111. doi: 10.14456/tstj.2020.167

วิจิตร คลังภูเขียว, ปิยภัทร บุชบาบดินทร์ และบังอร กุมพล (2560). การพยากรณ์ความผันผวนราคาทองคำด้วยโครงข่ายประสาทเทียม( Forecasting Volatility of Gold Price with Artificial Neural Networks). ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยมหาสารคาม, 126-133.

#### 4.วารสาร

วิศรุต แก้วมหา,วิริศ ปัญญาฉัตรพร (2564). การคาดการณ์ผลตอบแทนในอนาคตของตราสารทุนหุ้นสามัญ โดยการใช้ระบบคอมพิวเตอร์เรียนรู้ได้ด้วยตนเอง (Predicting Stock Return Using Machine Learning) โดยใช้ตัวชี้วัดทางเทคนิคเพื่อการลงทุนในตลาดหลักทรัพย์แห่งประเทศไทย. วารสารเทคโนโลยีสารสนเทศ 2(3): 110-119.

#### ภาษาอังกฤษ

##### 1.Thesis

Jaroenkitwatcharachai K. (2018). Artificial intelligence for forecasting wage. (Master's degree Individual Study). Thammasat University.

Jesper Groenendijk. (2021). Predicting intraday stock returns using a hybrid ARIMA and long short-term memory neural network model. (Master's degree Quantitative Finance). Erasmus School of Economics or Erasmus University.

PATTAREEYA PIRAVECHSAKUL. (2021) Predicting Stock Return Using Machine Learning (Master of Engineering (Information and Communication Technology for Embedded Systems). KASETSART UNIVERSITY.

Valeriy V. Gavrishchaka, & Supriya Banerjee. (2006). Support vector machine as an efficient

framework for stock market volatility forecasting. Computational Management Science (CMS), pp. 147–160. Doi: 10.1007/s10287-005-0005-5

## 2. Journal

Kim, K.-j. (2003). Financial time series forecasting using support vector machines.

Neurocomputing, 55(1), 307–319. doi: [https://doi.org/10.1016/S0925-2312\(03\)00372-2](https://doi.org/10.1016/S0925-2312(03)00372-2)

Nevasalmi, L. (2020). Forecasting multinomial stock returns using machine learning

methods. Journal of Finance and Data Science, 6(1), 86-106.

doi: <https://doi.org/10.1016/j.jfds.2020.09.001>

Patel, J. et al. (2014). Predicting Stock and Stock Price Index Movement Using Trend

Deterministic Data Preparation and Machine Learning Techniques. Expert Systems with Applications, 42(1), 259-268.

Wabomba, M. S., Mutwiri, M. P., Frederick, M (2016). Modelling and Forecasting Kenyan GDP

using ARIMA models, Science Journal of Applied Mathematics and Statistics, 4(2), 64-73.

ภาคผนวก

### ภาคผนวก ก

ผลลัพธ์การเปรียบเทียบตัวแบบ ARIMA ที่เหมาะสมที่สุดผ่านตัวชี้วัด Akaike Info Criterion (AIC) ที่ได้จาก Google Collaboratory run on Python version 3.9

```
Performing stepwise search to minimize aic
ARIMA(2,0,2)(0,0,0)[0] intercept : AIC=3137.437, Time=2.37 sec
ARIMA(0,0,0)(0,0,0)[0] intercept : AIC=4368.995, Time=0.12 sec
ARIMA(1,0,0)(0,0,0)[0] intercept : AIC=3143.527, Time=0.69 sec
ARIMA(0,0,1)(0,0,0)[0] intercept : AIC=3940.794, Time=0.65 sec
ARIMA(0,0,0)(0,0,0)[0] : AIC=6270.864, Time=0.09 sec
ARIMA(1,0,2)(0,0,0)[0] intercept : AIC=3137.306, Time=2.03 sec
ARIMA(0,0,2)(0,0,0)[0] intercept : AIC=3713.956, Time=1.15 sec
ARIMA(1,0,1)(0,0,0)[0] intercept : AIC=3144.420, Time=0.83 sec
ARIMA(1,0,3)(0,0,0)[0] intercept : AIC=3135.699, Time=1.14 sec
ARIMA(0,0,3)(0,0,0)[0] intercept : AIC=3529.566, Time=1.32 sec
ARIMA(2,0,3)(0,0,0)[0] intercept : AIC=3133.381, Time=1.36 sec
ARIMA(3,0,3)(0,0,0)[0] intercept : AIC=3135.647, Time=2.23 sec
ARIMA(2,0,4)(0,0,0)[0] intercept : AIC=3135.359, Time=1.34 sec
ARIMA(1,0,4)(0,0,0)[0] intercept : AIC=3134.535, Time=0.84 sec
ARIMA(3,0,2)(0,0,0)[0] intercept : AIC=3135.325, Time=0.87 sec
ARIMA(3,0,4)(0,0,0)[0] intercept : AIC=3137.033, Time=2.42 sec
ARIMA(2,0,3)(0,0,0)[0] : AIC=3139.866, Time=1.16 sec
```



## ภาคผนวก ข

ผลลัพธ์ตัวแบบ  $ARIMAX(p,d,q)$  ที่เหมาะสม ที่ได้จาก Google Collaboratory

run on Python version 3.9

$ARIMAX(2,0,3)$

```

SARIMAX Results
Dep. Variable: SET50      No. Observations: 380
Model: ARIMA(2, 0, 3)    Log Likelihood -1450.090
Date: Tue, 25 Apr 2023    AIC      2926.180
Time: 09:46:51           BIC      2977.402
Sample: 0                HQIC     2946.505
- 380

Covariance Type: opg
      coef  std err   z    P>|z| [0.025   0.975]
const  945.5704  34.940  27.063  0.000  877.090 1014.051
PC1    -12.8721   4.396  -2.928  0.003 -21.487  -4.257
PC2    -31.2620   3.444  -9.077  0.000 -38.012 -24.511
PC3     8.2498   2.719   3.034  0.002  2.920  13.580
PC5    -14.7328   3.481  -4.232  0.000 -21.556  -7.909
PC7     3.6188   3.646   0.993  0.321  -3.527  10.765
PC8     2.9125   4.167   0.699  0.485  -5.254  11.079
ar.L1   0.1247   0.054   2.296  0.022  0.018   0.231
ar.L2   0.8490   0.053  15.933  0.000  0.745   0.953
ma.L1   0.6142   0.061   9.992  0.000  0.494   0.735
ma.L2  -0.2509   0.061  -4.085  0.000 -0.371  -0.131
ma.L3   0.0668   0.057   1.170  0.242 -0.045   0.179
sigma2  119.2924   5.164  23.101  0.000 109.171 129.414

```

$ARIMAX(2,0,2)$

## SARIMAX Results

**Dep. Variable:** SET50      **No. Observations:** 380  
**Model:** ARIMA(2, 0, 2)      **Log Likelihood** -1449.433  
**Date:** Tue, 25 Apr 2023      **AIC** 2918.866  
**Time:** 18:27:20      **BIC** 2958.267  
**Sample:** 0      **HQIC** 2934.500  
 - 380

**Covariance Type:** opg

	coef	std err	z	P> z	[0.025	0.975]
<b>const</b>	937.5339	35.239	26.605	0.000	868.467	1006.601
<b>PC1</b>	-9.6185	2.825	-3.405	0.001	-15.155	-4.082
<b>PC2</b>	-33.9232	2.183	-15.537	0.000	-38.203	-29.644
<b>PC3</b>	7.7357	2.165	3.573	0.000	3.492	11.979
<b>PC5</b>	-14.6701	3.501	-4.191	0.000	-21.531	-7.809
<b>ar.L1</b>	0.0933	0.036	2.572	0.010	0.022	0.164
<b>ar.L2</b>	0.8858	0.035	25.272	0.000	0.817	0.955
<b>ma.L1</b>	0.6317	0.048	13.046	0.000	0.537	0.727
<b>ma.L2</b>	-0.3302	0.037	-9.041	0.000	-0.402	-0.259
<b>sigma2</b>	118.2637	4.959	23.848	0.000	108.544	127.983