การทำนายหลักทรัพย์ด้วยซีอริมา

PREDICTING STOCK PRICES USING CARIMA (CROSS CORRELATION ARIMA)

สรพล วิชัยดิษฐ

SORAPON WICHAIDIT

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมคอมพิวเตอร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2559

KMITL-2016-EN-M-070-058

การทำนายหลักทรัพย์ด้วยซีอริมา

PREDICTING STOCK PRICES USING CARIMA

(CROSS CORRELATION ARIMA)

ศรพล วิชัยดิษฐ

SORNPON WICHAIDIT

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมคอมพิวเตอร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ.2559

KMITL-2016-EN-M-070-058

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

PREDICTING STOCK PRICES USING CARIMA

(CROSS CORRELATION ARIMA)

SORNPON WICHAIDIT

A THESIS SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENT FOR THE DEGREE OF

MASTER OF ENGINEERING IN COMPUTER ENGINEERING

FACULTY OF ENGINEERING

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2016

KMITL-2016-EN-M-070-058

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

COPYRIGHT 2016

FACULTY OF ENGINEERING

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ใบรับรองวิทยานิพนธ์

หัวข้อวิทยานิพนธ์ การทำนายหลักทรัพย์ด้วยซีอริมา
Thesis Title Predicting Stock Prices Using CARIMA (Cross Correlation ARIMA)
นักศึกษา นายศรพล วิชัยดิษฐ
รหัสประจำตัว 56601133
ปริญญา วิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชา วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์ ผศ.ดร.สุรินทร์ กิตติธรกุล
หมายเลขวิทยานิพนธ์ KMITL-2016-EN-M-070-058

คณะกรรมการสอบวิทยานิพนธ์		ลายมือชื่อ
รศ.ดร.บุญธีร์	เครือตราชู	
รศ.ดร.เกียรติกูล	เจียรนัยธนะกิจ	
ผศ.ดร.ภุชงค์	อุทโยภาศ	
ผศ.ดร.ชุติเมษฏ์	ศรีนิลทา	
ผศ.ดร.สุรินทร์	กิตติธรกุล	

วัน / เดือน/ ปี ที่สอบ วันพุธที่ 15 มิถุนายน พ.ศ. 2559 เวลา 13.00-15.00 น.
สถานที่สอบ ณ อาคาร A ชั้น 5 ห้องประชุม 3

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

คณะวิศวกรรมศาสตร์ รับรองแล้ว

(รองศาสตราจารย์ ดร. คมสัน มาลีสี)
คณบดี คณะวิศวกรรมศาสตร์
วันที่ 15 มิถุนายน พ.ศ. 2559

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารนี้ทุกครั้งที่มีการนำไปใช้

หัวข้อวิทยานิพนธ์	การทำนายหลักทรัพย์ด้วยซีอริมา
นักศึกษา	นายศรพล วิชัยดิษฐ
รหัสประจำตัว	56601133
ปริญญา	วิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชา	วิศวกรรมคอมพิวเตอร์
พ.ศ.	2558
อาจารย์ที่ปรึกษาวิทยานิพนธ์	ผศ.ดร.สุรินทร์ กิตติธรกุล

บทคัดย่อ

วิทยานิพนธ์ฉบับนี้ นำเสนอโมเดลการทำนายหลักทรัพย์ในระยะสั้นด้วย วิธีการสมการเชิงเส้น ชื่อ CARIMA (Cross Correlation Autoregressive Integrated Moving Average). คือการผสมผสานวิธีการ ARIMA ด้วยหลักการสหสัมพันธ์แบบข้าม(Cross Correlation) โดยใช้หลักการใช้หลักทรัพย์อื่นมาช่วยในการทำนายกับวิธีการ ARIMA ดั้งเดิม โดยใช้ทั้งแบบ 1.การไม่ใช้หลักทรัพย์อื่นช่วยในการทำนาย 2. ใช้หลักทรัพย์ที่มีการนำหลักทรัพย์เป้าหมายมากที่สุดมาช่วยในการทำนาย 3.การใช้หลักทรัพย์อื่นมากกว่าหนึ่งหลักทรัพย์มาช่วยในการทำนาย และกระทำการวิเคราะห์ผลลัพธ์ที่ได้มาจากการทำนายมาเปรียบเทียบกัน เพื่อนำไปสู่ระบบการทำนายที่ช่วยนักลงทุนตัดสินใจที่มีประสิทธภาพต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Thesis	PREDICTING STOCK PRICES USING CARIMA (CROSS CORRELATION ARIMA)
Student	Mr.Sornpon Wichaidit
Student ID.	56601133
Degree	Master of Engineering
Program	Computer Engineering
Year	2015
Thesis Advisor	Asst.Prof.Dr.Surin Kittornkun

ABSTRACT

This Thesis presents the novel prediction model, CARIMAR (Cross Correlation Autoregressive Integrated Moving Average). This model has applied the linear equation in order to predict stock prices. The aim is enhancing the effectiveness of the prediction model by adding cross-correlation value to the ARIMAR model. To predict the price of a target stock, this model will select another stock with strong cross-correlation as an assistant stock. In the experimental results, the researchers have compared the efficiency of three different methods: 1) predict by using only the ARIMA model, 2) predict by using the help of another stock with high cross-correlation and 3) predict by using the help of a group of stocks with high cross-correlations. Finally, we have illustrated our analytical results in order to apply our model in the creation of an effective prediction model.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้าที่

บทคัดย่อภาษาไทย I

บทคัดย่อภาษาอังกฤษ II

กิตติกรรมประกาศ III

สารบัญ IV

สารบัญตาราง VI

สารบัญรูป VII

บทที่ 1 บทนำ 1

1.1 ความเป็นมาของปัญหา 1

1.2 วัตถุประสงค์ของการศึกษา 2

1.3 สมมุติฐานของการศึกษา 2

1.4 ขอบเขตงานวิจัย 2

1.5 ขั้นตอนของการศึกษา 2

1.6 โครงสร้างงานวิจัย 2

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 4

2.1 อนุกรมเวลา 4

2.2 ตัวแบบ Autoregressive Integrated Moving Average Model (ARIMA) 4

2.2.1 แบบจำลอง Auto Regressive (AR(p)) 5

2.2.2 แบบจำลอง Moving Average (MA(q 6

2.2.3 แบบจำลอง Auto Regressive Moving Average (ARMA(p,q)) 6

2.2.4 กระบวนการ Integrated (I(d)) 7

2.2.5 แบบจำลอง Autoregressive Integrated Moving Average model (ARIMA) 8

2.2.6 Autocorrelation Function (ACF) 9

2.2.7 Partial Autocorrelation Function (PACF) 9

2.3 การกำหนดสมบัติความคงที่ของอนุกรมเวลา (Stationary) 10

2.4 การกำหนดพารามิเตอร์ ARIMA(p,d,q) 12

2.5 ค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Percentage Error) 13

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

2.6 สหสัมพันธ์ (Correlation Analysis) และสหสัมพันธ์แบบข้าม (Cross Correlation Analysis).....13
2.6.1 สหสัมพันธ์ (Correlation Analysis).....13
2.6.2 สหสัมพันธ์แบบข้าม (Cross Correlation Analysis).....14
2.6.3 สหสัมพันธ์แบบระยะทาง (Distance Correlation Analysis).....14
2.7 การคิดเปอร์เซ็นความถูกต้อง.....15
2.8 การคิดราคาปิด(Closing Price).....16
2.9 อัตราส่วนชาร์ป (Sharpe ratio).....16
2.10 งานวิจัยที่เกี่ยวข้อง.....17

บทที่ 3 Cross Correlation Autoregressive integrated moving average (CARIMA).....19
3.1 ตรวจสอบสมบัติความคงที่ของอนุกรมเวลาด้วย ADF Test.....19
3.2 กำหนดพารามิเตอร์ ARIMA(p,d,q) โดยพิจารณา จาก ACF และ PACF.....19
3.3 การหา Lead/Lag ด้วย Cross Correlation ระหว่างหลักทรัพย์แต่ละตัว.....20
3.4 รวมผล Cross Correlation กับ ARIMA.....21
3.4.1 การรวมผลตัวแปรเดียว CARIMA-1.....21
3.4.2 การรวมผลหลายตัวแปร CARIMA-n.....21
3.5 อัลกอริทึม CARIMA-NF และ CARIMA-F.....22
3.5.1 เส้นค่าเฉลี่ยเคลื่อนที่แบบปรกติของสหสัมพันธ์ (Simple Moving Average of Correlation).....22
3.5.2 อัลกอริทึมการเทรดด้วยโมเดล CARIMA (Trading Algorithm with CARIMA).....23
3.5.2.1 โมเดลเทรดด้วยโมเดลแบบไม่มีการกรองทิศทาง CARIMA-NF.....23

3.5.2.1 โมเดลเทรดด้วยโมเดลแบบไม่มีการกรองทิศทาง CARIMA-NF.....23

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

บทที่ 4 ผลทดลองและอภิปรายผลการทดลอง.....25

4.1. ข้อมูลและแหล่งข้อมูล.....25

4.2. อภิปรายผลการพยากรณ์โดยโมเดล ARIMA และ โมเดล CARIMA.....27

4.2.1 ค่าร้อยละความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Percentage Error).33

4.2.2 ค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เปรียบเทียบโดยระยะเวลาการทำนาย.....37

4.2.3 ค่าสหสัมพันธ์ (Correlation).....38

4.2.4 ค่าสหสัมพันธ์เปรียบเทียบโดยระยะเวลาการทำนาย.....41

4.2.5 ค่าสหสัมพันธ์แบบระยะทาง.....42

4.2.6 ค่าสหสัมพันธ์แบบระยะทางเปรียบเทียบโดยระยะเวลาการทำนาย.....45

4.2.7 อภิปรายผลการเทรดด้วยโมเดล ARIMA, CARIMA-1 และ CARIMA-n.....46

4.3. อภิปรายผลการเทรดด้วยโมเดล ARIMA, CARIMA-NF และ CARIMA-F.....50

4.3.1 ข้อมูลที่ใช้ในการซิมมูเลชั่นการเทรด.....50

4.3.2 อภิปรายผลซิมมูเลชั่นของโมเดล CARIMA-NF.....52

4.3.2 อภิปรายผลซิมมูเลชั่นของโมเดล CARIMA-F.....56

4.4. สรุปงานวิจัย.....59

บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ.....61

5.1 สรุปผลการวิจัย.....61

5.2 ข้อเสนอแนะ.....61

5.3 แนวทางการพัฒนาต่อ.....62

เอกสารอ้างอิง.....63

ภาคผนวก.....64

ประวัติผู้เขียน.....69

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่ หน้าที่

2.1 ตารางพิจรณา ACF และ PACF....12

2.2 สรุปข้อมูลงานวิจัยที่เกี่ยวข้อง....18

3.1 ตารางค่า ค่า Akaike Information Criterion (AIC) และ Bayesian information criterion (BIC) ของโมเดลหลักทรัพ BAY....20

4.1 ตัวอย่างข้อมูลหลักทรัพย์ BAY จากวันที่ 1 มกราคม 2014 จึงวันที่ 14 มกราค2014....26

4.2 ผลการทดลองเปรียบเทียบค่าร้อยละความคลาดเคลื่อนสัมบูรณ์ของ ARIMA, CARIMA-1, CARIMA-n จากการทำนายเป็นเวลา 10 วัน....33

4.3 ผลการทดลองเปรียบเทียบค่าร้อยละความคลาดเคลื่อนเฉลี่ยของการทำนายหลักทรัพย์ทั้งหมดระหว่าง ARIMA, CARIMA-1 และ CARIMA–n เปรียบเทียบโดยระยะเวลาการทำนายจาก 10 วัน จนถึง 1 วัน....37

4.4 ผลการทดลองเปรียบเทียบค่าสหสัมพันธ์เฉลี่ยของการทำนายหลักทรัพย์ทั้งหมดระหว่าง ARIMA, CARIMA-1 และ CARIMA-n จากการทำนายเป็นเวลา 10 วัน....38

4.5 ผลการทดลองเปรียบเทียบค่าสหสัมพันธ์ระหว่าง ARIMA CARIMA-1 และ CARIMA–n เปรียบเทียบโดยระยะเวลาการทำนายจาก 10 วัน จนถึง 6 วัน....41

4.6 เปรียบเทียบค่าสหสัมพันธ์แบบระยะทางระหว่าง ARIMA CARIMA-1 และ CARIMA–2 โดยจากการทำนายเป็นเวลา 10 วัน....42

4.7 ผลการทดลองเปรียบเทียบค่าสหสัมพันธ์แบบระยะทางระหว่าง ARIMA CARIMA-1 และ CARIMA–n เปรียบเทียบโดยระยะเวลาการทำนายจาก 10 วัน จนถึง 6 วัน....45

4.8 ตารางเปอร์ความถูกต้องของการทำนายจากการซื้อขายตามการทำนาย ด้วย ARIMA, CARIMA-1 และ CARIMA-n จากผลการทำนายเป็นเวลา 10 วัน....46

4.9 ตัวอย่างข้อมูลดัชนีหลักทรัพย์ SET50....50

4.10 ตารางสรุปประสิทธภาพของโมเดล ARIMA CARIMA-NF....56

4.11 ตารางสรุปประสิทธภาพของโมเดล ARIMA CARIMA-F....59

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่ หน้าที่

2.1 ตัวอย่างข้อมูลที่มีคุณสมบัติคงที่ และไม่คงที่....12

3.1 ผลการทดสอบหลักทรัพย์ BAY ด้วย ADF Test....19

4.1 กราฟราคาปิดของหลักทรัพย์ในช่วงเวลาที่ใช้ในการทดลองของหลักทรัพย์ BAY โดยเริ่มตั้งแต่ วันที่ 1 มกราคม 2014 จนถึง วันที่ 31 ธันวาคม 2014....26

4.2 กราฟราคาปิดของหลักดัชนี SET50 ในช่วงเวลาที่ใช้ในการทดลองโดยเริ่มตั้งแต่ วันที่ 1 มกราคม 2014 จนถึง วันที่ 31 ธันวาคม 2014....27

4.3 กราฟเปรียบเทียบผลการทำนายหลักทรัพย์ BAY เป็นเวลา 10 วันจากวันที่ 1 มกราคม 2015 จนถึงวันที่ 13 มกราคม 2015....28

4.4 กราฟเปรียบเทียบผลการทำนายหลักทรัพย์ BDMS เป็นเวลา 10 วันจากวันที่ 1 มกราคม 2015 จนถึงวันที่ 13 มกราคม 2015....29

4.5 กราฟเปรียบเทียบผลการทำนายหลักทรัพย์ M เป็นเวลา 5 วันจากวันที่ 1 มกราคม 2015 จนถึงวันที่ 7 มกราคม 2015....30

4.6 กราฟเปรียบเทียบผลการทำนายหลักทรัพย์ HMPRO เป็นเวลา 5 วันจากวันที่ 1 มกราคม 2015 จนถึง วันที่ 7 มกราคม 2015....31

4.7 กราฟเปรียบเทียบผลการทำนายหลักทรัพย์ BH เป็นเวลา 5 วันจากวันที่ 1 มกราคม 2015 จนถึง วันที่ 7 มกราคม 2015....32

4.8 ผลการทดลองเปรียบเทียบค่าร้อยละความคลาดเคลื่อนสัมบูรณ์ของ ARIMA, CARIMA-1 และ CARIMA-n....36

4.9 ผลการทดลองเปรียบเทียบค่าสหสัมพันธ์เฉลี่ยของการทำนายหลักทรัพย์ทั้งหมดระหว่าง ARIMA, CARIMA-1 และ CARIMA-n จากการทำนายเป็นเวลา 10 วัน....41

4.10 ผลการทดลองเปรียบเทียบค่าสหสัมพันธ์แบบระยะทางระหว่าง ARIMA CARIMA-1 และ CARIMA-n จากการทำนายเป็นเวลา 10 วัน....45

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

4.11 กราฟเปอร์เซ็นต์กำไรจากการซื้อขายตามการทำนาย ด้วย ARIMA, CARIMA-1 และ CARIMA-n จากการทำนายเป็นเวลา 10 วัน....49

4.12 ตัวอย่างข้อมูลราคาปิดของ SET50 และ MSCI USA ตั้งแต่ วันที่ 1 มกราคม 2014 จนถึง วันที่ 31 ธันวาคม 2015....51

4.13 ตัวอย่างข้อมูลค่าสหสัมพันธ์ของ SET50 และ MSCI USA ตั้งแต่ วันที่ 1 มกราคม 2014 จนถึง วันที่ 31 ธันวาคม 2015....51

4.14 ผลเปรียบเทียบซิมมูเลชั่นโมเดล ARIMA และ CARIMA-NF ตั้งแต่ วันที่ 1 มกราคม 2015 จนถึง วันที่ 31 ธันวาคม 2015....52

4.15 ผลเปรียบเทียบซิมมูเลชั่นโมเดล ARIMA และ CARIMA-NF วันที่ 1 มกราคม 2015 จนถึง วันที่ 12 มกราคม 2015....53

4.16 ค่าสหสัมพันธ์ของ SET50 และ MSCI USA (บน) และ ราคาเปลี่ยนแปลงของ MSCI USA (ล่าง) ตั้งแต่ วันที่ 1 มกราคม 2015 จนถึง วันที่ 12 มกราคม 2015....53

4.17 ผลเปรียบเทียบซิมมูเลชั่นโมเดล ARIMA และ CARIMA-NF ตั้งแต่วันที่ 1 กันยายน ถึง 16 กันยายน 2015....54

4.18 ค่าสหสัมพันธ์ของ SET50 และ MSCI USA (บน) และ ราคาเปลี่ยนแปลงของ MSCI USA (ล่าง) ตั้งแต่ วันที่ 1 กันยายน 2015 จนถึง วันที่ 16 กันยายน 2015....55

4.19 ผลเปรียบเทียบซิมมูเลชั่นโมเดล ARIMA และ CARIMA-F ตั้งแต่ วันที่ 1 มกราคม 2015 จนถึง วันที่ 31 ธันวาคม 2015....56

4.20 ผลเปรียบเทียบซิมมูเลชั่นโมเดล ARIMA และ CARIMA-F เดือน 1 กรกฎาคม จนถึง 30 ธันวาคม 2015....57

4.21 ผลเปรียบเทียบซิมมูเลชั่นโมเดล ARIMA และ CARIMA-NF ตั้งแต่วันที่ 10 มิถุนายน ถึง 30 มิถุนายน 2015....58

4.22 ค่าสหสัมพันธ์ของ SET50 และ MSCI USA (บน) และ ราคาเปลี่ยนแปลงของ MSCI USA (ล่าง) ตั้งแต่วันที่ 10 มิถุนายน ถึง 30 มิถุนายน 2015....58

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

บทนำ

1.1ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันการลงทุนในตลาดหลักทรัพย์มีผลกระทบต่อเศรษฐกิจของประเทศเป็นอย่างมากกล่าวคือ มันจะมีส่วนช่วยส่งเสริมภาคการลงทุนในประเทศ ทำให้เศรษฐกิจอุตสาหกรรมจะเกิดการขยายตัวอันเนื่องมาจากการเพิ่มทุนลงไปในระบบอุตสหกรรมทำให้บริษัทในประเทศมีทุนในการขยายงานและการลงทุนและผลที่ตามมาคือ อัตราการจ้างงานสูงขึ้น ฐานะความเป็นอยู่ของคนในประเทศก็จะดีขึ้น จะเกิดการจับจ่ายใช้สอยภายในประเทศมากขึ้น รัฐบาลจะสามารถจัดเก็บภาษีได้มากขึ้น ทำให้รัฐบาลมีเงินทุนในการพัฒนาประเทศมากขึ้นตามไปด้วย

ตลาดหลักทรัพย์แห่งประเทศไทยเป็นแหล่งระดมเงินทุนสำหรับบริษัทที่ต้องการระดมเงินทุนสำหรับธุรกิจที่ต้องการระดมเงินทุนเพื่อดำเนินการพัฒนาธุรกิจ และสามารถทำให้ประชาชนใช้เป็นช่องทางการออมเงินและลงทุนได้และเป็นเจ้าของกิจการ ซึ่งถ้าบริษัทเติบโตดำเนินกิจการมีกำไร นักลงทุนก็จะมีสิทธิได้รับเงินปันผล หรือ ได้กำไรจากส่วนต่างมูลค่าหลักทรัพย์ระหว่างราคาซื้อและขาย ซึ่งสามารถให้ผลกำไรมากกว่าการออมเงินโดยการฝากธนาคารพาณิชย์ แต่การลงทุนในตลาดหลักทรัพย์นั้นมีความเสี่ยงสูง หากบริษัทที่ลงทุนนั้นเกิดภาวะขาดทุน หรือ ราคาของหลักทรัพย์บริษัทนั้นๆ เกิดการเคลื่อนไหนไหวที่ผันผวน นักลงทุนจึงต้องศึกษาหาวิธีลงทุนให้ลดความเสี่ยงให้มากที่สุด

การทำนายอนุกรมเวลา (Time Series) เป็นวิธีที่ใช้สถิติที่ใช้ข้อมูลทางสถิติในอดีตมาเป็นตัวแบบในการอธิบายความสัมพันธ์ระหว่างค่าที่เราสังเกตตามลำดับของเวลา และใช้มันเป็นตัวแบบในการทำนายอนาคต วิธีทำนายข้อมูลในอนาคตสามารถแบ่งเป็นสองกลุ่มใหญ่ๆ คือ วิธีการแบบไม่เชิงเส้น(None Linear approach) ซึ่งพยายามอธิบายข้อมูลที่ไม่ใช่ความสัมพันธ์เชิงเส้น และวิธีการแบบเชิงเส้น (Linear approach) ซึ่งมีข้อสมมุติฐานว่าลักษณะข้อมูลเป็นความสัมพันธ์ของค่าสังเกตในปัจจุบันเป็นค่าที่สามารถอธิบายได้ด้วยสมการเชิงเส้นจากอดีต ตัวแบบที่แพร่หลาย ก็คือ วิธีแบบ ARIMA (Autoregressive Integrated Moving Average)[4] ซึ่งจะเป็นตัวแบบในงานวิจัยครั้งนี้

การศึกษาครั้งนี้จะเป็นการเพิ่มประสิทธิภาพในการทำนายหลักทรัพย์ราคาหลักทรัพย์ของบริษัทในกลุ่มหลักทรัพย์ SET50 ด้วยวิธีการ ARIMA ด้วยการรวมสหสัมพันธ์ไขว้(Cross Correlation) ของหลักทรัพย์ต่างๆใน SET50 เพื่อเพิ่มประสิทธิภาพการทำนาย โดยให้ชื่อว่า CARIMA โดยจะมีการแบ่งเป็น 2 ประแภทคือ 1. CARIMA แบบใช้การรวมโดยใช้สหสัมพันธ์ไขว้โดยเลือกจากหลักเกณฑ์ที่มีค่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สัมประสิทธิ์สหสัมพันธ์มากที่สุดและมากกว่าเงื่อนไข เรียกว่า CARIMA-1 ใช้หลักทรัพย์ที่มีค่าสหสัมพันธ์มากที่สุดเพียง 1 ตัว และ CARIMA-n ใช้หลักทรัพย์ที่มีสหสัมพันธ์มากกว่าเงื่อนไขที่กำหนด 2. CARIMA แบบคำนึงถึงเทรนด์ของค่าสัมประสิทธิ์ของสหสัมพันธ์โดยแบ่งเป็น CARIMA-NF คือ การใช้เทรนด์ของดัชนีประเทศสหรัฐอเมริกามาช่วยในการทำนายดัชนี SET50 โดยใช้อัลกอริทึ่มครอสโอเวอร์ของค่าสัมประสิทธิ์สหสัมพันธ์ และ CARIMA-F คือการใช้การใช้เทรนด์ของดัชนีประเทศสหรัสอเมริกามาช่วยในการทำนายดัชนี SET50 โดยใช้อัลกอริทึ่มครอสโอเวอร์ของค่าสัมประสิทธิ์สหสัมพันธ์โดยทำการคัดกรองทิศทางของค่า ณ วันที่มีการครอสโอเวอรนั้น

1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา

1.2.1 เพื่อศึกษาวิธีการพยากรณ์ราคาหลักทรัพย์ SET50 ด้วยโมเดล ARIMA

1.2.2 เพื่อเปรียบเทียบวิธีการพยากรณ์ราคาหลักทรัพย์ SET50 ด้วยวิธี ARIMA และ CARIMA

1.2.3 เพื่อเพิ่มประสิทธิภาพในการการพยากรณ์โดยใช้ Cross Correlation

1.2.4 เพื่อลดค่า Drawdown ของการซื้อขายหลักทรัพย์

1.2.5 เพื่อศึกษาความสัมพันธ์ระหว่างหลักทรัพย์ต่อหลักทรัพย์และดัชนีต่อดัชนี

1.3 สมมุติฐานของการศึกษา

หลักทรัพย์ใน SET50 ที่มีการนำ (Lead) เป็นจำนวนวันที่เพียงพอ การศึกษาสามารถนำหลักทรัพย์เหล่านั้นมาช่วยในการทำนายการเคลื่อนไหวของราคาหลักทรัพย์ตัวอื่นใน SET50 ได้ ซึ่งทำให้ได้ประสิทธิภาพที่ดีขึ้นได้ และ ดัชนีหลักทรัพย์จากประเทศอื่นมีส่วนช่วยในการทำนายราคาหลักทรัพย์ SET50 ได้ และสามารถนำมาช่วยลดค่า Drawdown ได้

1.4 ขอบเขตการวิจัย

1.4.1 งานวิจัยชิ้นนี้เพื่อเพิ่มประสิทธิภาพในการทำนายทิศทางราคาของหลักทรัพย์ตระกูล SET50 ด้วยวิธี ARIMA และ CARIMA (ARIMA และ Cross Correlation) ทั้ง 4 โมเดลย่อย

1.4.2 งานวิจัยชิ้นนี้เพื่อศึกษาความสัมพันธ์ของราคาหลักทรัพย์ในตระกูล SET50 และ ดัชนีหลักทรัพย์ SET50 และ ดัชนีหลักทรัพย์ MSCI USA

1.4.3 ข้อมูลในงานวิจัยชิ้นนี้ได้ศึกษาการพยากรณ์ราคาหลักทรัพย์โดยใช้ราคาปิดของบริษัทในครือ SET50 จำนวน 50 หลักทรัพย์ และราคาปิดของดัชนีหลักทรัพย์ SET50 และ ดัชนี MSCI USA

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.5 ขั้นตอนของการศึกษา

1.5.1 รวบรวมข้อมูลราคาหลักทรัพย์ในตระกูล SET50 จาก http://www.setsmart.com/ และข้อมูลดัชนี MSCI USA จาก http://www.yahoo.com/

1.5.2 สร้างตัวแบบ ARIMA และ CARIMA-1, CAIMA-n, CARIMA-NF, CARIMA-F

1.5.3 เปรียบเทียบประสิทธิภาพของตัวแบบการทดลอง ARIMA และ CARIMA-1, CARIMA-n จากการทำนายหลักทรัพย์เป็นเวลา 1 ถึง 10 วัน และ 3 เปรียบเทียบประสิทธิภาพของตัวแบบการทดลอง ARIMA และ CARIMA-NF, CARIMA-F จากการทำนายหลักทรัพย์ SET50 เป็นเวลา 1 ปี

1.5.4 สรุปผลการทดลองและข้อเสนอแนะ

1.6 โครงสร้างของการวิจัย

1.6.1 บทที่ 2 จะอธิบายเกี่ยวกับสมการต่างๆที่เกี่ยวข้องและทษฎีของนักวิจัยคนอื่นๆที่เคยทำวิจัยทางด้านนี้มาก่อน

1.6.2 บทที่ 3 จะอธิบายเกี่ยวกับวิธีการวิธีการของ CARIMA

1.6.3 บทที่ 4 จะเป็นการอภิปรายผลการทำนานด้วย ARIMA และ CARIMA เปรียบเทียบกัน

1.6.4 บทที่ 5 จะเป็นการสรุปผลการทดลองและข้อแนะนำและแนวทางการพัฒนาต่อ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 อนุกรมเวลา (Time Series)

อนุกรมเวลา คือ เซตของข้อมูลเชิงปริมาณที่จัดเก็บในช่วงเวลาหนึ่ง ตัวอย่างเช่น ดัชนีตลาดหลักทรัพย์ในแต่ละวันเมื่อปิดทำการซื้อขายในแต่ละวัน รายได้ประชาชาติ (GNP) รายไตรมาส รายรับในแต่ละปีของบริษัทแห่งหนึ่ง

ข้อมูลอนุกรมเวลา (Time Series data) คือ ชุดของข้อมูลที่เก็บรวบรวมตามระยะเวลาเป็นช่วงๆ อย่างต่อเนื่องกัน เช่น ข้อมูลยอดขายสินค้าที่เก็บรวบรวมต่อเนื่องกันไปเป็นระยะเวลาหลายๆเดือน ข้อมูลรายได้ประชาชาติปีต่างๆ ที่เก็บรวบรวมต่อเนื่องกันไปเป็นระยะเวลาหลายๆ ปี เป็นต้น ข้อมูลอนุกรมเวลาอาจอยู่ในลักษณะที่เป็นข้อมูลรายปี รายไตรมาส หรือรายเดือนก็ได้ ทั้งนี้ขึ้นอยู่กับความเหมาะสมในการนำไปใช้ประโยชน์

เนื่องจากข้อมูลทางธุรกิจมีการเปลี่ยนแปลงตลอดเวลา ผู้นำทางธุรกิจหรือองค์กรต้องหาวิธีพัฒนาต่างๆ ที่สามารถนำไปใช้ประกอบการตัดสินใจวางแผน เกี่ยวกับผลที่เกิดจากความเปลี่ยนแปลงในการดำเนินการอย่างใดอย่างหนึ่ง ดังนั้นการวิเคราะห์อนุกรมเวลาจึงเข้ามามีบทบาทช่วยในการตัดสินใจเทคนิคอย่างหนึ่งที่ใช้ช่วยในการควบคุมการดำเนินการในปัจจุบันและในการวางแผนความต้องการในอนาคต คือ การทำนาย(forecasting) ซึ่งการทำนายนั้นทำได้หลายวิธี แต่ละวิธีต่างมีเป้าหมายเดียวกัน คือ ทำนายเหตุการณ์ในอนาคต องค์ประกอบของอนุกรมเวลามีอะไรบ้าง

2.2 ตัวแบบ Autoregressive Integrated Moving Average Model (ARIMA)[6]

แบบจำลอง ARIMA[1] เป็นแบบจำลองทางคณิตศาสตร์ที่ได้รับความนิยมอย่างมาก และเป็นวิธีที่ให้ค่าทำนายในระยะสั้นที่ดี เนื่องจากโมเดลให้ค่าเฉลี่ยของความคลาดเคลื่อน (Mean Square Error : MSE) ของการทำนายที่ได้จะต่ำกว่าวิธีอื่นๆ เช่น วิธีถดถอยเชิงพหุ วิธีการปรับเรียบแบบเอ็กโพเน็นเชียล และ การวิเคราะห์แนวโน้ม ป็นต้น อีกทั้งในการจัดทำสมการและการทำนายยังมีขั้นตอนที่ยุ่งยาก และซับซ้อนน้อยกว่าแบบมหภาคที่อยู่ในลักษณะระบบสมการหลายชั้น สำหรับแบบจำลอง ARIMA เป็นโมเดลที่พัฒนาโดย George E.P.Box และ Gwilym M. Jenkins ในปี ค.ศ. 1970 โดยพื้นฐานแล้วแบบจำลอง ARIMA เป็นวิธีที่ให้ค่าการทำนายในระยะสั้นที่ดี หรือเหมาะกับการทำนายไปข้างหน้าในช่วงเวลาสั้นๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

และต้องมีช่วงของข้อมูลที่ยาวนานพอสมควร แบบจำลอง ARIMA(p,d,q) ประกอบด้วย 3 ส่วน อันได้แก่ แบบจำลอง AutoRegressive ($AR(p)$) กระบวนการ Integrated ($I(d)$) และแบบจำลอง Moving Average ($MA(q)$) โดยรายละเอียดของแต่ละส่วนมีดังต่อไปนี้

2.2.1 แบบจำลอง Auto Regressive ($AR(p)$)

แบบจำลองโมเดล Auto Regressive เป็นรูปแบบที่แสดงว่าค่าสังเกต Y_t ถูกกำหนดจากค่าของ X_{t-1},..., X_{t-p}หรือ ค่าสังเกตที่ได้ปรากฏขึ้นก่อนหน้าจำนวน p โดยกระบวนการหรือระบบ $AR(p)$ คือ กระบวนการหรือระบบ Auto Regressive ที่มีอันดับที่ p ซึ่งเขียนอยู่ในรูปสมการได้ดังต่อไปนี้

$$\text{AR(p)} \text{ คือ } X_t = \mu + \emptyset_1 X_{t-2} + \ldots + \emptyset_p X_{t-p} + \varepsilon_t \quad (2.1)$$

โดยที่

μ คือ ค่าคงที่ (Constant Term)

$\emptyset_p$ คือ พารามิเตอร์ตัวที่ p

ε_t คือ ความคลาดเคลื่อนเวลา ณ วันที่ t

ในกรณี ของ $AR(2)$ สามารถเขียนรูปแบบสมการได้ดังนี้

$$X_t = \mu + \emptyset_1 X_{t-1} + \varepsilon_t \quad (2.2)$$

หรือ

$$X_t - \emptyset_1 X_{t-1} = \mu + \varepsilon_t \quad (2.3)$$

หรือ

$$(1 - \emptyset_1 B) X_t = \mu + \varepsilon_t \quad (2.4)$$

เมื่อ B คือ Backward shift operation

และในกรณี ของ $AR(2)$ สามารถเขียนรูปแบบสมการได้ดังนี้

$$X_t = \mu + \emptyset_1 X_{t-1} + \emptyset_2 X_{t-2} + \varepsilon_t \quad (2.5)$$

หรือ

$$X_t - \emptyset_1 X_{t-1} - \emptyset_2 X_{t-2} = \mu + \varepsilon_t \quad (2.6)$$

หรือ

$$(1 - \emptyset_1 B - \emptyset_2 B) X_t = \mu + \varepsilon_t \quad (2.7)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.2 แบบจำลอง Moving Average ($MA(q)$)

แบบจำลองโมเดล Moving Average (MA) เป็นรูปแบบที่แสดงว่าค่าสังเกตการณ์ Y_t เป็นผลมาจากค่าความคลาดเคลื่อน $\varepsilon_{t-1},\ldots,\varepsilon_{t-q}$ หรือค่าความคลาดเคลื่อนที่เกิดขึ้นอยู่ก่อนหน้านั้น โดยกระบวนการหรือโมเดล $MA(q)$ คือกระบวนการหรือโมเดล Moving Average ค่าเฉลี่ยเคลื่อนที่ของกระบวกการคลาดเคลื่อนที่มีอันดับ q ซึ่งเขียนในรูปของ $MA(q)$ ได้ดังนี

$$MA(q) \text{ คือ } X_t = \sigma + \varepsilon_t - \emptyset_1\varepsilon_{t-1} - \emptyset_2\varepsilon_{t-2} - \ldots - \emptyset_q\varepsilon_{t-q} \quad (2.8)$$

โดยที่

σ คือ ค่าคงที่ (Constant Term)

$\emptyset_q$ คือ พารามิเตอร์ตัวที่ q

ε_t คือ ความคลาดเคลื่อนเวลา ณ วันที่ t

ในกรณี $MA(1)$ สามารถเขียนรูปแบบสมการได้ดังต่อไปนี

$$X_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} \quad (2.9)$$

หรือ

$$X_t = \mu + (1 - \theta_1 B)\varepsilon_t \quad (2.10)$$

และในกรณี $MA(2)$สามารถเขียนรูปแบบสมการได้ดังต่อไปนี้

$$X_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} \quad (2.11)$$

หรือ

$$X_t = \mu + (1 - \theta_1 B - \theta_1 B^2)\varepsilon_t \quad (2.12)$$

2.2.3 แบบจำลอง Auto Regressive Moving Average ($ARIMA(p,q)$)

แบบจำลองโมเดล Auto Regressive Moving Average (ARMA) เป็นแบบจำลองที่นำเอากระบวนการ Auto Regressive และ Moving Average มารวมกันโดยกระบวนการหรือโมเดล $ARIMA(p,q)$ คือกระบวนการหรือโมเดล Auto Regressive ที่มีอันดับที่ p และ Moving Average ที่มีอันดับ q ซึ่งเขียนอยู่ในรูปสมการได้ดังนี้ต่อไปนี้

แบบจำลอง $ARIMA(p,q)$

$$X_t = \sigma + \emptyset X_{t-1} + \emptyset X_{t-2} + \cdots + \emptyset X_{t-p} + \varepsilon_t + \theta_1\varepsilon_{t-1} - \ldots - \theta_q\varepsilon_{t-q} \quad (2.13)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยที่

X_t คือ ค่าสังเกตในอนุกรมเวลา ณ วันที่ t

p คือ อันดับของ Auto Regressive

q คือ อันดับของ Moving Average

σ คือ ค่าคงที่ (Constant Term)

ε_t คือ กระบวนการ white noise ซึ่งก็คือ ค่าความคลาดเคลื่อนเวลา ณ วันที่ t

$\emptyset$ คือ พารามิเตอร์ของ Auto Regressive

θ คือ พารามิเตอร์ของ Moving Average

2.2.4 กระบวนการ Integrated $(I(d))$

กระบวนการอินทีเกรด Integrated $I(d)$ เป็นการหาผลต่างของอนุกรมเวลาระหว่างข้อมูล ณ ปัจจุบันกับข้อมูลถอยหลังไป d เวลา โดยสาเหตุที่ต้องทำการหาผลต่างของอนุกรมเวลานั้นเพราะเนื่องจากแบบจำลองโมเดล ARIMA ต้องใช้ในการวิเคราะห์ข้อมูลอนุกรมเวลาทีมีคุณสมบัติคงที่ (Stationary) เท่านั้น โดยในกรณีข้อมูลอนุกรรมเวลาที่เรานำใช้ในการวิเคราะห์มีคุณสมบัติไม่คงที่ (Nonstationary) เราจำเป็นจะต้องทำการแปลงข้อมูลดังกล่าวให้เป็นข้อมูลที่มีคุณสมบัติคงที่ก่อน โดยการใช้วิธีหาผลต่างของข้อมูลอนุกรมเวลาก่อนที่นำไปสร้างแบบจำลอง ARIMA ซึ่งโดยทั่วไปแล้วถ้าต้องหาผลต่างอันดับที่ d สามารถเขียนในรูปของ $I(d)$ ได้ดังต่อไปนี้

$$I(d) \quad \text{คือ } \Delta_d X_t = \Delta_{d-1}(X_t - X_{t-1}) \text{ หรือ } (1-B)^d X_t \tag{2.14}$$

ในกรณี $I(1)$ สามารถเขียนรูปแบบได้ดังต่อไปนี้

$$I(1) \quad \text{คือ } \Delta X_t = (X_t - X_{t-1}) \text{ หรือ } (1-B)X_t \tag{2.15}$$

ในกรณี $I(2)$ สามารถเขียนรูปแบบได้ดังต่อไปนี้

$$I(2) \quad \text{คือ } \Delta_2 X_t = \Delta(X_t - X_{t-1}) \text{ หรือ } (1-B)^2 X_t \tag{2.16}$$

โดยที่

$(1-B)^d X_t$ คือ ผลต่างอันดับที่ d

B คือ Backward shift operation

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.5 แบบจำลอง Autoregressive Integrated Moving Average model (ARIMA)

จากโมเดลต่างๆและรายละเอียดที่เราได้กล่าวในข้างต้นถ้านำแบบจำลองโมเดล Auto Regressive และแบบจำลองโมเดล Moving Average และ กระบวนการ Integrated มาพิจารณารวมกันสามารถนำมากำหนดเป็นรูปแบบทั่วไปของแบบจำลอง ARIMA ที่ใช้ในการประมาณการคือ

แบบจำลอง $ARIMA(p,d,q)$

$$\hat{X}_t = \sigma + \emptyset\Delta_d X_{t-1} + \emptyset\Delta_d X_{t-2} + \cdots + +\emptyset\Delta_d X_{t-p} + \varepsilon_t - \theta_1\varepsilon_{t-1} - \ldots - \theta_1\varepsilon_{t-q} \quad (2.17)$$

โดยที่

X_t คือ ค่าสังเกตในอนุกรมเวลา ณ เวลา t

d คือ จำนวนครั้งของการหาผลต่างเพื่อให้อนุกรมเวลามีคุณสมบัติคงที่ (Stationary)

p คือ อันดับของ Autoregressive

q คือ อันดับของ Moving Average

σ คือ ค่าคงที่ (Constant Term)

Δ คือ ผลต่างอนุกรมเวลา

$\emptyset_1, ., \emptyset_p$ คือ พารามิเตอร์ของ Auto Regressive

$\theta_1, ., \theta_q$ คือ พารามิเตอร์ของ Moving Average

ε_t คือ กระบวนการ White Noise ซึ่งก็คือ ค่าความคลาดเคลื่อน ณ เวลา t ภายใต้ข้อสมมติที่ว่าความคลาดเคลื่อนที่คนละเวลาเป็นตัวแปรสุ่มที่เป็นอิสระต่อกัน โดยมี การแจกแจงปกติที่มีค่าเฉลี่ยเป็นศูนย์ และความแปรปรวนคงที่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.6 Autocorrelation Function (ACF)

สหสัมพันธ์ของตัวเอง Autocorrelation Function เป็นฟังก์ชันของการวัดสหสัมพันธ์ระหว่างข้อมูลตัวมันเอง ณ เวลา $t(X_t)$ และ ข้อมูล ณ เวลา $t-i(X_{t-i})$ ของช่วงเวลาห่างกัน i หน่วย ซึ่งแทนด้วยสัญลักษณ์ p_i หรือ r_iในกรณีสหสัมพันธ์ในตัวเองของตัวอย่าง ซึ่งสามารถคำนวณได้ดังนี้

$$p_k \text{ หรือ } r_i = \frac{\sum_{t=1}^{n}(X_t-\bar{X})(X_{t-i}-\bar{X})}{\sum_{t=1}^{n}(X_t-\bar{X})^2} \tag{2.18}$$

เมื่อ $\bar{X} = \frac{1}{n}\sum_{t=1}^{n} X_t$และ $i = 0,1,2,3 \ldots$

โดยความคลาดเคลื่อนมาตรฐานของ r_i (Standard error of r_i) ซึ่งมีสูตรในการคำนวณดังนี้

$$se_i = \frac{1}{\sqrt{n}} \tag{2.19}$$

สหสัมพันธ์ในตัวเองของข้อมูลสุ่ม (random data) มีการแจกแจงเชิงตัวอย่างที่สามารถประมาณได้ โดยการแจกแจงปกติด้วยค่าเฉลี่ย (mean) เท่ากับศูนย์ และความคลาดเคลื่อนมาตรฐานเท่ากับ $\frac{1}{\sqrt{n}}$

ในการศึกษาจะใช้สหสัมพันธ์ในตัวเองเป็นเครื่องมือที่สำคัญสำหรับสืบค้นคุณสมบัติของอนุกรมเวลาเชิงประจักษ์ โดยมี 2 วิธีสำหรับทดสอบว่าค่า r_i มีค่าแตกต่างไปจาก 0 หรือไม่โดย ใช้การแจกแจงปกติมาตรฐาน (Standard Normal Distribution) หรือ ใช้ค่าสถิติ Box-Pierce Q statistic ซึ่งมีรูปแบบดังนี้

การแจกแจงปกติมาตรฐาน (Standard Normal Distribution)

$$\mathrm{r_i} \sim N(0, \frac{1}{\sqrt{n}}) \tag{2.20}$$

ค่าสถิติ Box-Pierce Q statistic

$$Q = n\sum_{i=1}^{m} r^2 \sim X^2(m-p-q) \tag{2.21}$$

โดยที่ m คือค่าล่าหรือค่าล้าหลังสูงสุด (Maximum Lag) ที่พิจารณา

2.2.7 Partial Autocorrelation Function (PACF)

การพิจารณาสหสัมพันธ์ระหว่างตัวแปร X_t กับ X_{t-1} อาจเป็นไปได้ว่าสหสัมพันธ์ดังกล่าวเป็นผลเนื่องมาจากสหสัมพันธ์ระหว่าง 2 ตัวแปรนี้กับตัวแปร X_{t-1}อาจ,..., X_{t-i+1}อาจดังนั้นเพื่อที่จะได้สหสัมพันธ์ระหว่าง X_t กับ X_{t-1} ที่ได้ขจัดความเกี่ยวข้องระหว่างตัวแปรทั้งสองตัวนี้กับตัวแปร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

X_{t-1},..., X_{t-i+1}ดังกล่าว จึงต้องทำการวัดสหสัมพันธ์ของทั้งสองตัวแปรในรูปแบบของการสหสัมพันธ์แบบมีเงื่อนไข $Corr(X_t, X_{t-i}|X_{t-1}, \dots, X_{t-i+1})$ซึ่งเรียกว่า Partial Autocorrelation โดยแทนด้วยสัญลักษณ์ $\emptyset_{ii}$ แต่ถ้านำสหสัมพันธ์ในตัวเองบางส่วนมาพิจารณาในรูปแบบฟังก์ชัน จะเรียกว่า Partial Autocorrelation Function (PACF) ซึ่ง$\emptyset_{kk}$สามารถคำนวณได้ดังนี้

$$\emptyset_i = \frac{Cov[(X_t - \hat{X}_t),[(X_{t-i} - \hat{X}_{t-i})]}{\sqrt{Var(X_t - \hat{X}_t)}\sqrt{Var(X_{t-i} - \hat{X}_{t-i})}} \quad (2.22)$$

โดยที่ $\hat{x}_t = \beta_1 X_{t-1} + \beta_2 X_{t-2} + \dots + \beta_k X_{t-i+1}$

2.3. การกำหนดสมบัติความคงที่ของอนุกรมเวลา (Stationary)[6]

การทดสอบ (Unit Root Test) คือการทดสอบความนิ่งของข้อมูลที่จะนำมาศึกษาสามารถทำได้โดยวิธีการทดสอบ DF (Dickey-Fuller Test) ซึ่งเสนอโดย Dickeyและ Fuller ในปี คศ.1981 และวิธีการทดสอบ ADF (Augmented Dickey-Fuller Test) ซึ่งได้รับการพัฒนาโดย Said และ Dickey ในปี คศ. 1984 ข้อมูลที่มีลักษณะนิ่ง (Stationary) หมายถึง ข้อมูลอนุกรมเวลาที่มีค่าเฉลี่ย (Mean) และความแปรปรวน (Variance) เท่ากันตลอดระยะเวลาที่ศึกษา และข้อมูลที่มีลักษณะไม่นิ่ง (Non-Stationary) หมายถึง ข้อมูลอนุกรมเวลาที่มีค่าเฉลี่ย (Mean) และความแปรปรวน (Variance) ไม่เท่ากันตลอดระยะเวลาที่ศึกษา ข้อมูลอนุกรมเวลามพบว่ามีลักษณะไม่นิ่ง ซึ่งต้องทำให้ข้อมูลมีลักษณะนิ่งก่อน โดยอาจทดสอบหาความสัมพันธ์ของตัวแปรในระยะยาว (Cointegration) การแปลงให้อยู่ในรูป Logarithm หรือการหาผลต่างของข้อมูล (Differencing) เป็นต้น ข้อมูลอนุกรมเวลาที่มีลักษณะนิ่ง (Stationary) คือ ข้อมูลที่ค่าเฉลี่ยและความแปรปรวนของ กระบวนการเชิงสุ่มนั้นมีค่าคงที่เมื่อเวลาได้เปลี่ยนไปและค่าความแปรปรวนระหว่างสองคาบเวลา ขึ้นอยู่กับความล่า (Lag) ระหว่างสองคาบเวลานั้น

การทดสอบว่าข้อมูลอนุกรมเวลาที่จะนาใช้มีลักษณะนิ่งหรือไม่นั้น เราจะใช้การทดสอบด้วยวิธี DF (Dickey-Fuller Test)และ ADF (Augmented Dickey-Fuller Test) กำหนดโดยสมการ

$$X_t = \rho X_{t-1} + \varepsilon_t \quad (2.23)$$

โดยที่

X_t คือ ตัวแปรอิสระหรือตัวแปรอนุกรมเวลาเวลา ณ วันที่ t

ρ คือ สัมประสิทธิ์อัตสหสัมพันธ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ε_t คือ ค่าความคลาดเคลื่อนเวลา ณ วันที่ t

โดยกำหนดสมมุติฐานหลัก $H_0: \rho = 1$

และสมมุติฐานรอง $H_1: |\rho| < 1$

ถ้าไม่ปฏิเสธ H_0 ข้อมูลมีลักษณะไม่นิ่งถ้าปฏิเสธ H_0 แสดงว่าข้อมูลมีลักษณะนิ่ง และการทดสอบนี้สามารถแปลงสมการได้ดังนี้แสดงว่า

กรณีไม่มีค่าคงที่และแนวโน้มเวลา

$$\Delta X_t = \theta X_{t-1} + \varepsilon_t \quad (2.24)$$

กรณีมีเฉพาะค่าคงที่

$$\Delta X_t = \propto + \theta X_{t-1} + \varepsilon_t \quad (2.25)$$

กรณีมีทั้งค่าคงที่และแนวโน้มเวลา

$$\Delta X_t = \propto + \beta_t + \theta X_{t-1} + \varepsilon_t \quad (2.26)$$

โดยกำหนดสมมุติฐานหลัก $H_0: \theta = 0$

และสมมุติฐานรอง $H_1: \theta < 0$

หากไม่ปฏิเสธ H_0แสดงว่าข้อมูลมีลักษณะไม่นิ่งถ้าปฏิเสธ H_0 แสดงว่าข้อมูลมีลักษณะนิ่ง เมื่อนำมาเข้าเป็นอัตถดถอยเป็น Augmented Dickey – Fuller Test ได้ดังนี้

กรณีไม่มีค่าคงที่และแนวโน้มเวลา

$$\Delta X_t = \theta X_{t-1} + \sum_{i=1}^{\rho} \emptyset_i \Delta X_{t-1} + \varepsilon_t \quad (2.27)$$

กรณีมีเฉพาะค่าคงที่

$$\Delta X_t = \propto \theta X_{t-1} + \sum_{i=1}^{\rho} \emptyset_i \Delta X_{t-1} + \varepsilon_t \quad (2.28)$$

กรณีมีทั้งค่าคงที่และแนวโน้มเวลา

$$\Delta X_t = \propto + \beta_t + \theta X_{t-1} + \sum_{i=1}^{\rho} \emptyset_i \Delta X_{t-1} + \varepsilon_t \quad (2.29)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

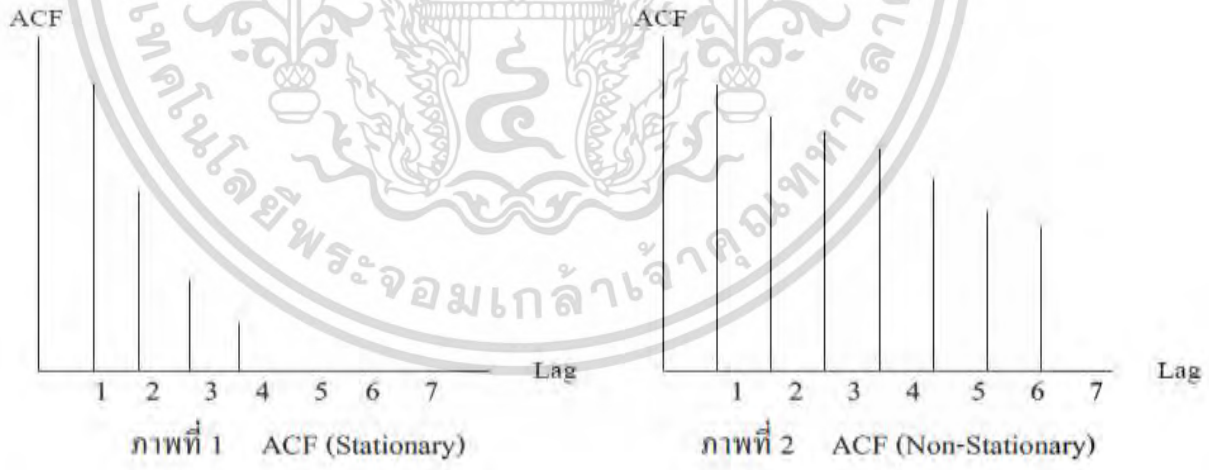
2.4 การกำหนดพารามิเตอร์ ARIMA(p,d,q)[6]

เป็นขั้นตอนที่พิจารณาว่าตัวแบบใดที่เหมาะสม กับข้อมูลที่นำมาวิเคราะห์ โดยพิจารณาจากค่า Autocorrelation Function (ACF) และค่า Partial Autocorrelation Function (PACF) ในตารางที่ 1

ตารางที่ 2.1 ตารางพิจรณา ACF และ PACF

ตัวแบบ	**ACF**	**PACF**
AR(p)	ลดลงเข้าหา 0 อย่างรวดเร็ว	หลัง lag p มีค่าเท่ากับ 0
MA(q)	หลัง lag q มีค่าเท่ากับ 0	ลดลงเข้าหา 0 อย่างรวดเร็ว
ARMA(p,q)	ลดลงเข้าหา 0 อย่างรวดเร็ว	ลดลงเข้าหา 0 อย่างรวดเร็ว

ข้อมูลอนุกรมเวลาที่นำมาสร้างตัวแบบต้องมีคุณสมบัติคงที่ (Stationary) การที่จะพิจรณาว่าข้อมูลมีคุณสมบัติคงที่หรือไม่เราจะดูจากค่าสหสัมพันธ์ในตัวเอง ACF กล่าวคือค่าของ ACF ที่คงที่จะต้องมีลักษณะที่ลดลงเข้าหาศูนย์อย่างรวดเร็วหลังจาก Lag 0 ณ จุดเวลาใดเวลาหนึ่ง ถ้ามันเข้าหาศูนย์หลังจาก Lag 0 อย่างช้าๆ แปลว่าข้อมูลเป็นข้อมูลที่ไม่คงที่ (None Stationary) ถ้าข้อมูลไม่คงที่เราต้องผ่านกระบวนการให้มันคงที่ก่อนด้วยการหาผลต่างจึงจะสามารถดำเนินต่อได้ ดังรูปที่ 2.1 คือ ตัวอย่างข้อมูลที่มีคุณสมบัติคงที่ และไม่คงที่



รูปที่ 2.1 ตัวอย่างข้อมูลที่มีคุณสมบัติคงที่ และไม่คงที่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในตัวแบบ $ARIMA(p,d,q)$ p คือ จำนวนเทอมที่ถอยในตัวเอง d คืออันดับของผลต่างที่ทำให้ข้อมูลนิ่ง q คือจำนวนเทอมของค่าเฉลี่ยเคลื่อนที่ ตัวอย่างเช่นกระบวนการ $ARIMA(2,1,2)$ มีผลต่างอันดับที่ 1$(d=1)$ ที่ทำให้ข้อมูลนิ่ง และมีเทอมการถดถอย และเทอมค่าเฉลี่ยเคลื่อนที่ 2 เทอมเท่ากัน ถ้า $(d=1)$ กระบวนการ $ARIMA(p,0,q)$ หมายถึง $ARIMA(p,q)$ ข้อสังเกตกระบวนการ $ARIMA(p,0,0)$ หมายถึงกระบวนการ $AR(p)$ และกระบวนการ $ARIMA(0,0,q)$ หมายถึงกระบวนการ $MA(q)$

ประมาณค่าพารามิเตอร์ เป็นขั้นตอนการประมาณค่าพารามิเตอร์ที่มีอยู่ในตัวแบบอนุกรมเวลา โดยใช้การประมาณค่าพารามิเตอร์ด้วยวิธีความควรจะเป็นสูงสุด (Maximum likelihood)

2.5 ค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Percentage Error)

ค่าเฉลี่ยความคลาดเคลื่อนสมบูรณ์ เป็นวิธีการวัด ค่าความแตกต่างระหว่างค่าจริงและค่าที่ประมาณได้ จากแบบจำลอง โดยหากค่า MAPE มีค่าน้อย แสดงว่า แบบจำลองสามารถประมาณค่าได้ใกล้เคียงกับค่าจริง ดังนั้นหากค่านี้มีค่าเท่ากับศูนย์แล้วจะหมายความว่า ไม่ เกิดความคลาดเคลื่อนในแบบจำลองนี้เลย ค่า MAPE สามารถคำนวณได้ดัง

$$MAPE = \left(\frac{1}{n}\sum_{t=1}^{n}\frac{|X_t-\hat{X}_t|}{X_t}\right) * 100 \qquad (2.30)$$

โดยที่

n คือ จำนวนตัวแปร

X_t คือ ค่าอนุกรมเวลาจริงเวลา ณ วันที่ t

$\hat{X}_t$ คือ ค่าที่ทำนายเวลา ณ วันที่ t

2.6 สหสัมพันธ์ (Correlation Analysis) และสหสัมพันธ์แบบข้าม (Cross Correlation Analysis)

สหสัมพันธ์เป็นสมการเป็นสมการที่บอกว่าตัวแปร 2 ตัว สหสัมพันธ์ ก็มีหลายประเภทดังนี้

2.6.1 สหสัมพันธ์ (Correlation Analysis)

เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปร 2 ตัว ว่ามีความสัมพันธ์ต่อกันหรือไม่ มากน้อยเพียงใด โดยนำข้อมูลของตัวแปรทั้ง 2 ตัว ณ ช่วงเวลา t มาศึกษาโดยใชสมการดังนี้

$$\rho_{X_tY_t} = \frac{\sum_{t=1}^{n}(X_t-\bar{X})(Y_t-\bar{Y})}{\delta_X\delta_Y} \qquad (2.31)$$

โดยที่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

X_t คือ ค่าของอนุกรมเวลา X เวลา ณ วันที่ t

$\bar{X}$ คือ ค่าเฉลี่ยของอนุกรมเวลา X

Y_t คือ ค่าของอนุกรมเวลา Y เวลา ณ วันที่ t

$\bar{Y}$ คือ ค่าเฉลี่ยของอนุกรมเวลา Y

δ_X คือ ส่วนเบี่ยงเบนมาตรฐานของอนุกรมเวลา X

δ_Y คือ ส่วนเบี่ยงเบนมาตรฐานของอนุกรมเวลา Y

n คือ จำนวนของข้อมูล

2.6.2 สหสัมพันธ์แบบข้าม(Cross Correlation Analysis)

เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปร 2 ตัว ว่ามีความสัมพันธ์ต่อกันหรือไม่มากน้อยเพียงใด โดยนำข้อมูลของตัวแปรทั้ง 2 ตัว ณ ช่วงเวลา t และ $t-i$ มาศึกษาโดยใช้สมการดังนี้

$$\rho_{X_t Y_t(i)} = \frac{\sum_{t=1}^{n}(X_t-\bar{X})(Y_{t-i}-\bar{Y})}{\delta_X \delta_Y} \quad (2.32)$$

โดยที่

X_t คือ ค่าของอนุกรมเวลา X เวลา ณ วันที่ t

$\bar{X}$ คือ ค่าเฉลี่ยของอนุกรมเวลา X

Y_{t-i} คือ ค่าของอนุกรมเวลา Y เวลา เวลา ณ วันที่ $t-i$

$\bar{y}$ คือ ค่าเฉลี่ยของอนุกรมเวลา Y

δ_x คือ ส่วนเบี่ยงเบนมาตรฐานของอนุกรมเวลา X

δ_y คือ ส่วนเบี่ยงเบนมาตรฐานของอนุกรมเวลา Y

n คือ จำนวนของข้อมูล

2.6.3 สหสัมพันธ์แบบระยะทาง (Distance Correlation)[2]

เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปร 2 ตัว ว่ามีความสัมพันธ์ต่อกันหรือไม่ มากน้อยเพียงใด แต่มันไม่จำเป็นต้องมีมิติของข้อมูลเท่ากันทั้ง 2 ตัวแปร และค่าสัมประสิทธิ์ของมันเริ่มต้นตั้งแต่ 0 ถึง 1 โดยไม่นับทิศทาง โดยนำข้อมูลของตัวแปรทั้ง 2 ตัว ณ ช่วงเวลา t มาศึกษาโดยใช้สมการดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$A_{t,j} = ||X_t - X_j|| \quad (2.33)$$

$$B_{t,j} = ||Y_t - Y_j|| \quad (2.34)$$

$$Dcov(X,Y) = \frac{1}{n^2}\sum_{t,j=1}^{n} A_{t,j} * B_{t,j} \quad (2.35)$$

$$Dvar(x)^2 = \frac{1}{n^2}\sum_{t,j=1}^{n} A_{t,j} \quad (2.36)$$

$$\frac{Dcov(X,Y)}{\sqrt{Dvar(X)Dvar(Y)}} \quad (2.37)$$

โดยที่

X คือ ค่าเฉลี่ยของอนุกรมเวลา X

Y คือ ค่าเฉลี่ยของอนุกรมเวลา Y

$Dcov(X,Y)$ คือ ความแปรปรวนเกี่ยวพันธ์แบบระยะทางของอนุกรมเวลา X และ Y

$Dvar(X)$ คือ ความแปรปรวนแบบระยะทางของอนุกรมเวลา X

$Dvar(Y)$ คือ ความแปรปรวนแบบระยะทางของอนุกรมเวลา Y

t คือ เวลา ณ วันที่ t

j คือ เวลา ณ วันที่ j

2.7 การคิดเปอร์เซ็นต์ความถูกต้อง

คิดเปอร์เซ็นต์ความแม่นยำของโมเดลการทำนาย โดยถ้าทำนายการเปลี่ยนแปลงขึ้นหรือลงตามค่าจริง

$$\frac{T}{T+F} \quad (2.38)$$

โดยที่

T คือ จำนวนการทำนายการขึ้นลงที่ถูก

F คือ จำนวนการทำนายการขึ้นลงที่ผิด

2.8 การคิดราคาปิด (Closing Price)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เพื่อเป็นการลดโอกาสในการสร้างราคาปิดของผู้ไม่ประสงค์ดีและทำให้ราคาปิดของหลักทรัพย์มีเสถียรภาพยิ่งขึ้น ปัจจุบันตลาดหลักทรัพย์จึงคำนวณราคาปิดของหลักทรัพย์โดยใช้วิธีผสมผสานระหว่างการสุ่มเลือกเวลา (Random time) และวิธี Call market ในการคำนวณราคาปิด ตลาดหลักทรัพย์จะเปิดให้ซื้อขายได้ตามปกติจนถึงเวลา 16.30 น. จากนั้นระบบการซื้อขายจะหาเวลาปิดโดยการสุ่มเลือก (Random) เวลาในช่วง 16.35-16.40 น. และเปิดโอกาสให้บริษัทสมาชิกสามารถส่งคำสั่งซื้อขายเพิ่มเติมได้อีกอย่างน้อย 5 นาที ตั้งแต่ 16.30 น. ถึงเวลาปิดที่ได้จากการสุ่มเลือก และจะยังไม่จับคู่การซื้อขายจนกว่าจะถึงเวลาปิด จึงนำคำสั่งซื้อขายทั้งหมดที่ค้างอยู่ในระบบจนกระทั่งถึงเวลาปิด มาคำนวณหาราคาปิดของแต่ละหลักทรัพย์ โดยวิธี Call market (วิธีเดียวกับการคำนวณหาราคาเปิดในช่วง Pre-open) ตามหลักเกณฑ์ดังนี้

1. เป็นราคาที่ทำให้เกิดการซื้อขายได้ในปริมาณมากที่สุด
2. ถ้ามีราคาที่ทำให้เกิดปริมาณซื้อขายมากที่สุด มากกว่า 1 ราคา ให้ใช้ราคาที่ใกล้เคียงกับราคาซื้อขายครั้งสุดท้ายก่อนหน้านั้นมากที่สุด
3. ถ้ามีราคาที่ใกล้เคียงกับราคาซื้อขายครั้งสุดท้ายก่อนหน้านั้น มากกว่า1 ราคา ให้ใช้ราคาที่สูงกว่าเป็นราคาปิด
4. อย่างไรก็ตาม หากไม่สามารถคำนวณหาราคาปิดตามหลักเกณฑ์ข้างต้นได้ให้ถือว่าราคาซื้อขายครั้งสุดท้ายในวันนั้นเป็นราคาปิด

2.8 อัตราส่วนชาร์ป (Sharpe ratio)

Sharpe ratio คือ เครื่องวัดประสิทธิภาพในการบริหารพอร์ตโฟลิโอในทางการเงิน หรือ ตัววัดความเสี่ยง(Risk) ต่อการลงทุนหนึงๆ ว่าการลงทุนในรอบช่วงเวลานั้นๆ ได้กำไรต่อความเสี่ยงที่เราต้องรับมา ความเสี่ยงในที่นี้ก็คือความผันผวน(Volatility) หรือก็คือ ระหว่างที่ผลตอบแทนของกองทุนหรือหลักทรัพย์นั้นๆจะไปถึงเป้าหมาย เช่น Return 18 % ต่อปีนั้นมีมากแค่ไหน และ ผลตอบแทนคุ้มกับความเสี่ยงนั้นๆหรือไม่ โดย Sharpe Ratio ได้ถูกคิดค้นโดยWilliam Forsyth Sharpe และมันได้กลายเป็น Standard ของการวัด Equity Curve ที่โด่งดังที่สุดตัวหนึ่ง โดยจะมีสมการดังนี้ โดยค่าที่ได้ถ้ามากกว่า 1 ก็จะถือว่าความเสี่ยง ณ เวลานั้น คุ้มค่ากับการลงทุน

$$R = \sum_{t=1}^{n=250} R_t \tag{2.39}$$

$$\frac{R-RF}{\delta_R} \tag{2.40}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยที่

R คือ ผลตอบแทนของการลงทุนในพอร์ตโฟลิโอเป็นเวลา 1 ปี

R_t คือ ผลตอบแทนการลงทุนของพอร์ตโพลิโอ ณ วันที่ t

RF คือ ผลตอบแทนจากการลงทุนที่ไม่มีความเสี่ยง(Risk free rate)

δ_R คือ ส่วนเบี่ยงเบนมาตรฐานของผลตอบแทนการลงทุนในช่วงเวลา 1 ปี

2.10 งานวิจัยที่เกี่ยวข้อง

C. Fonseka & L. Liyanage (2008) [3] ได้ศึกษาความสัมพันธ์ของหลักทรัพย์ทั้งหมดในตลาดหลักทรัพย์ประเทศออสเตเรีย Australian Stock Exchange(ASX) ตั้งแต่ช่วงเวลาวันที่ 1 กรกฎาคม 2005 จนถึงวันที่ 1 กรกฎาคม 2007 โดยผู้ศึกษาได้จัดทำอัลกอริทึ่มสำหรับการทำนายการเคลื่อนไหวของราคาหลักทรัพย์โดยใช้ราคาเปิดในแต่ละวันในการศึกษา โดยใช้สหสัมพันธ์แบบใช้เวลา Lag (Cross Correlation) มาเป็นตัวชี้วัดว่า ถ้ามีหลักทรัพย์ที่มีการนำหลักทรัพย์เป้าหมาย มากกว่า 0.95 หรือน้อยกว่า -0.95 ผู้จัดทำจะนำการเคลื่อนไหวในวันต่อไปของหลักทรัพย์นั้นมาทำนายหลักทรัพย์เป้าหมาย โดยผลลัพธ์จากการทำนายของงานวิจัยนั้น ได้เปอร์เซ็นต์ความแม่นยำถึง 67% แต่ด้วยผลทั้งหมดนั้น มาจากข้อมูลหลักทรัพย์จำนวนน้อยมากที่ผ่านเกณฑ์การเลือก แต่ผลที่ได้ก็เป็นไอเดียที่ดีในการนำไปวิจัยต่อไป

A. Srisawat (2010) [4] ได้ศึกษาความสัมพันธ์ของหลักทรัพย์ทั้งหมดในตลาดหลักทรัพย์ประเทศไทย Stock Exchange of Thailand (SET) โดยที่ผู้ศึกษาได้ใช้หลักการทำเหมืองข้อมูลแบบกฎความเชื่อมโยง(Association rule) ผู้ศึกษาได้ทำการแปลงข้อมูลหลักทรัพย์โดยใช้ราคาเปลี่ยนแปลงในแต่ละวันเป็นกลุ่มๆ ว่าหลักทรัพย์แต่ละหลักทรัพย์มีการเคลื่อนไหวในทิศทางขึ้นหรือลง โดยแบ่งเป็น แประเภท ขึ้น 1% ลง 1% ขึ้น 2% ลง 2% และ ขึ้น 3% ลง 3% โดยได้ใช้สมการกฎความเชื่อมโยง ผลที่ได้ออกมาเป็นกฎว่าหลักทรัพย์ใดบ้าง ที่ขึ้นลงตามกัน หรือ ขึ้นลงตรงข้ามกันด้วยอัตราเท่าไหร่ แต่จากผลลัพธ์นั้นกฎที่ได้มามักจะมีแต่การลงตามกันของหลักทรัพย์ใหญ่ๆที่ขึ้นตกตามตลาดหลักทรัพย์อยู่แล้ว อีกทั้งผลที่ได้นั่นเป็นผลจากการเคลื่อนไหวของราคาในวันเดียวกันกฎที่ได้มาจึงไม่สามารถนำไปใช้ประโยชน์ได้จริงเท่าใดนัก

W. Weiqing & Y. Lav [5] ได้ศึกษาการเคลื่อนไหวของราคาค่าเงินดอลล่าห์สหรัฐ (U.S. dollar index) และ ใช้สมการ Autoregressive integrated moving average (ARIMA) มาใช้ในการทำนายการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เคลื่อนไหวของราคาผลลัพธ์ที่ออกมาทำให้พบว่า โมเดล ARIMA จะให้ผลที่ดีสำหรับการทำนายระยะสั้น จะให้ผลที่ดีกว่า จากนั้นผู้ศึกษาได้ทำการเปรียบเทียบค่าเงินดอลล่าห์สหรัฐกับดัชนีตลาดหลักทรัพย์เซี่ยงไฮ้ (Shanghai index) ผลทดลองพบว่ามีค่าสหสัมพันธ์ในทางลบอย่างมาก จากนั้นผู้จัดทำได้สรุปว่าช่วงเวลาที่นำมาศึกษาเป็นช่วงเวลาที่มันเป็นดาวเทรนด์อย่างมากและได้อธิบายเชิงเศรษฐศาสตร์ว่า และราคาก็มีความผันผวนสูงมากในขาลงไม่สามารถนำมาอธิบายได้อย่างแม่นยำในโมเดลอนุกรมเวลาได้

ตารางที่ 2.2 สรุปข้อมูลงานวิจัยที่เกี่ยวข้อง

ผู้วิจัย	ตัวแบบ	ข้อมูล
C. Fonseka & L. Liyanage	Genetic Algorithm ด้วยวิธี Cross Correlation	หลักทรัพย์ทั้งหมดในตลาดหลักทรัพย์ประเทศออสเตเรีย Australian Stock Exchange(ASX)
A. Srisawat	Association rule	หลักทรัพย์ทั้งหมดในตลาดหลักทรัพย์ประเทศไทย Stock Exchange of Thailand (SET)
W. Weiqing & Y. Lav	Autoregressive integrated moving average (ARIMA)	ค่าเงินดอลล่าห์สหรัฐ (U.S. dollar index) และ ดัชนีตลาดหลักทรัพย์เซี่ยงไฮ้ (Shanghai index)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

Cross Correlation Autoregressive Integrated Moving Average (CARIMA)

ในการวิจัยครั้งนี้เป็นการวิจัยแบบทดลองการทำนายราคาหลักทรัพย์โดยวิธี CARIMA ซึ่งผู้วิจัยได้ดำเนินตามขั้นตอนต่อไปนี้

3.1. ตรวจสอบสมบัติความคงที่ของอนุกรมเวลาด้วย ADF Test

จากการทดสอบข้อมูลทั้ง 50 หลักทรัพย์ ด้วยสมการ Augmented Dickey-Fuller test (ADF Test)ในบทที่ 2 ผลที่ได้ คือข้อมูลทั้งหมดมีการกระจายปรกติและข้อมูลมีความคงที่ จึงไม่ต้องหาผลต่างลำดับต่อไป เราจะนำข้อมูลไปกำหนดพารามิเตอร์ในโมเดล ARIMA ในขั้นต่อไป รูปที่ 3.1 คือผลจากการใช่ ADF Test กับหลักทรัพย์ BAY ผลที่ได้แสดงให้เห็นว่าหลักทรัพย์ BAY มีคุณสมบัติคงที่ ค่า P-Value ต่ำกว่า 0.05 เราจึงสามารถสรุปได้ว่าหลักทรัพย์นี้มีคุณสมบัติคงที่ จึงสามารถนำหลักทรัพย์ไปใช้ทำนายในสมการเชิงเส้นได้ โดยที่หลักทรัพย์ทั้งหมดในดาต้าเซ็ทมีคุณสมบัติคงที่ทั้งหมด

```
        Augmented Dickey-Fuller Test

data:  na.omit(ROC(stock[, "CLOSE"]))
Dickey-Fuller = -6.0022, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary
```

รูปที่ 3.1 ผลการทดสอบหลักทรัพย์ BAY ด้วย ADF Test

3.2. กำหนดพารามิเตอร์ ARIMA(p,d,q) โดยพิจารณา จาก ACF และ PACF

พิจารณาจาก Correlogram เพื่อดูความสัมพันธ์ข้อมูลและค่า Akaike Information Criterion (AIC) และ Bayesian Information Criterion (BIC) เพื่อที่จะสามารถระบุได้ว่าแบบจำลองควรจะมี Auto regressive(p) เท่าใดและ Moving Average (q) เท่าใดโดยเลือกที่รูปแบบที่มีค่า AIC

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

และ BIC ต่ำที่สุดที่ผ่านกระบวนการ จากนั้นก็ทำการทำนายด้วย ARIMA และเก็บค่าจากการทำนายไว้เพื่อทดลองขั้นต่อไป จากรูป 3.1 เป็นตารางค่า AIC และ BIC ของโมเดลหลักทรัพย์ BAY โดยโมเดลที่เลือกใช้ในที่นี้คือ $(p = 0, q = 1, d = 0)$ เนื่องจากมีค่าข้างต้นต่ำที่สุด

ตารางที่ 3.1 ตารางค่า ค่า Akaike Information Criterion (AIC) และ Bayesian Information Criterion (BIC) ของโมเดลหลักทรัพย์ BAY

ARIMA	AIC	BIC
P=1,D=0,Q=0	-1307.22	-1296.71
P=0,D=1,Q=0	-1133.99	-1130.5
P=0,D=0,Q=1	-1307.92	-1296.71
P=1,D=1,Q=0	-1210.92	-1203.93
P=1,D=0,Q=1	-1305.24	-1291.24
P=0,D=1,Q=1	-1298.22	1291.22
P=1,D=1,Q=1	-1296.29	-1285.79

3.3. การหา Lead/Lag ด้วย Cross Correlation ระหว่างหลักทรัพย์แต่ละตัว

ผู้จัดทำได้นำหลักทรัพย์ทั้ง 50 หลักทรัพย์ มาหา Cross Correlation ไขว้กัน ในงานวิจัยนี้เราใช้ราคาปิดของหลักทรัพย์รายวันในการความสัมพันธ์ Cross Correlation โดยกำหนดเวลาการ Lead/Lag เป็นเวลา 50 วัน โดยได้ทำการหาโดยการนำหลักทรัพย์เป้าหมายที่เราจะทำการทำนายมาหาความสัมพันธ์กับ หลักทรัพย์ทั้งหมดในดาต้าเซ็ท และทำอย่างนี้กับหลักทรัพยทั้งหมดที่ใช้ในการทำการทดลอง เราจะใช้สมการในการหาความสัมพันธ์ด้วยสมการที่ 3.1

$$\rho X_t Y_t(i) = \sum_{t=1}^{796}[(X_t - \mu_X)(Y_{t-i} - \mu_Y)]/\delta_X \delta_Y \qquad (3.1)$$

โดยที่

$\rho X_t Y_t(i)$ คือ ค่าสัมประสิทธิ์สหสัมพันธ์ของหลักทรัพย์ X และ Y

X_t คือ ราคาหลักทรัพย์เป้าหมายเวลา ณ วันที่ t

Y_t คือ ราคาหลักทรัพย์อื่นในดาต้าเซตเวลา ณ วันที่ t

μ_X คือ ค่าเฉลี่ยของหลักทรัพย์เป้าหมาย X

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

μ_Y คือ ค่าเฉลี่ยของหลักทรัพย์อื่นในดาต้าเซต Y

δ_X คือ ส่วนเบี่ยงเบนมาตรฐานของหลักทรัพย์เป้าหมาย X

δ_Y คือ ส่วนเบี่ยงเบนมาตรฐานของหลักทรัพย์อื่นในดาต้าเซต Y

i คือ เวลาที่มีการ lead/lag (วัน)

3.4. รวมผล Cross Correlation กับ ARIMA

ในกระบวนการนี้เราจะนำผลจากการทำนายโดยวิธี ARIMA ของแต่ละหลักทรัพย์ ที่ทำนายเป็นเวลา 10 วัน เราจะทดลองเพิ่มประสิทธิภาพโดยใช้วิธีการแตกต่างกันดังนี้

3.4.1. การรวมผลตัวแปรเดียว CARIMA-1

ในการรวมผลแบบตัวแปรเดียวเราจะใช้หลักทรัพย์อื่นที่มีค่า $\rho X_t Y_t(i)$ สูงที่สุด และต้องมากกว่า 0.8หรือน้อยกว่า -0.8 และ ตัวแปรวันที่ lead ต้องมีค่ามากกว่า 10 หรือ $i > 10$ จากนั้นเราจะทำการรวมมันโดยย้อนกลับไปยังวันที่หลักทรัพย์อื่นที่ถูกเลือกมามีการ lead สูงสุด แล้วใช้อัตราการเปลี่ยนแปลง(Rate of Change)ของหลักทรัพย์นั้น 10 วันถัดมารวมเข้ากับผลการทำนายด้วยวิธี ARIMA ด้วยสมการ 3.2

$$\acute{X}_t = \hat{X}_t * \frac{\rho_{X_t Y_t(i)} * \left(1 + \frac{Y_{t-i} - Y_{t-i-1}}{Y_{t-i-1}}\right)}{5} \quad (3.2)$$

โดยที่

$\rho_{X_t Y_t(i)}$ คือ ค่าสัมประสิทธิ์สหสัมพันธ์ของหลักทรัพย์ X และ Y

X_t คือ ราคาหลักทรัพย์เป้าหมายเวลา ณ วันที่ t

Y_t คือ ราคาหลักทรัพย์อื่นในดาต้าเซตเวลา ณ วันที่ t

i คือ วันที่มีการ lead/lag (วัน)

3.4.2. การรวมผลหลายตัวแปร CARIMA-n

ในการรวมแบบหลายตัวแปรเราจะใช้หลักทรัพย์อื่นที่มีค่า $\rho X_t Y_t(i)$ มากกว่า 0.8 หรือน้อยกว่า -0.8 และ ตัวแปรวันที่ lead ต้องมีค่ามากกว่า 10 หรือ $i > 10$ จากนั้นเราจะทำการรวมผลโดยย้อนกลับไปยังวันที่หลักทรัพย์อื่นที่ถูกเลือกมามีการ lead สูงสุดของทุกหลักทรัพย์ แล้วใช้อัตราการเปลี่ยนแปลง(Rate of Change)ทั้งหมดของหลักทรัพย์นั้นๆเป็น 10 วันถัดมารวมเข้ากับผลการทำนายด้วยวิธี ARIMA ด้วยสมการ 3.3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\acute{X}_t = \hat{X}_t * \sum_{j=1}^{n} \frac{\frac{\rho_{X_t Y_{j,t}(i)} * \left(1 + \frac{Y_{j,t-i} - Y_{j,t-i-1}}{Y_{j,t-i-1}}\right)}{n}}{5} \quad (3.3)$$

โดยที่

$\rho_{X_t Y_{j,t}(i)}$ คือ ค่าสัมประสิทธิ์สหสัมพันธ์ของหลักทรัพย์ X และ Y

X_t คือ ราคาหลักทรัพย์เป้าหมายเวลา ณ วันที่ t

Y_t คือ ราคาหลักทรัพย์อื่นในดาต้าเซตเวลา ณ วันที่ t

i คือ จำนวนวันที่มีการ lead

n คือ จำนวนหลักทรัพย์ที่ผ่านเกณฑ์ทั้งหมด

3.5 อัลกอริทึม CARIMA-NF และ CARIMA-F

3.5.1 เส้นค่าเฉลี่ยเคลื่อนที่แบบปรกติของสหสัมพันธ์ (Simple Moving Average of Correlation)

เส้นค่าเฉลี่ยเคลื่อนที่แบบปรกติเป็นการคำนวณหาค่าเฉลี่ยราคาหลักทรัพย์หรือดัชนีในช่วงเวลาใดเวลาหนึ่ง โดยค่าเฉลี่ยเคลื่อนที่นี้อาจจะใช้ข้อมูลจากราคาปิด ราคาเปิด ราคาสูงสุด ต่ำสุด ก็ได้ แต่ในงานวิจัยนี้เราใช้มันมาหาค่าเฉลี่ยของสัมประสิทธิ์สหสัมพันธ์ระหว่างราคาปิดของหลักทรัพย์ X และ Y โดยใช้สมการดังนี้

$$Z_t = \rho_{X_t, Y_t(i)} \quad (3.4)$$

$$SMA(Z_t)_{10} = \frac{\sum Z_{t+1} + Z_{t+2} .. + Z_{t+n}}{10} \quad (3.5)$$

โดยที่

Z_t คือ สัมประสิทธิ์สหสัมพันธ์ระหว่างราคาปิดของหลักทรัพย์ X และ Y

$SMA(Z_t)_{10}$ คือ ค่าเฉลี่ยของสัมประสิทธิ์สหสัมพันธ์ระหว่างราคาปิดของหลักทรัพย์ X และ Y จำนวน 10 วัน ณ เวลา t

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.5.2 อัลกอริที่มการเทรดด้วยโมเดล CARIMA (Trading Algorithm with CARIMA)

ในส่วนนี้เราจะนำอัลกอริที่มซีอารีมา มาทำการซิมูเลชั่นการเทรดโดยแบ่งออกเป็นสองโมเดลดังนี้

3.5.2.1 โมเดลเทรดด้วยโมเดลแบบไม่มีการกรองทิศทาง CARIMA-NF

โมเดลเป็นการใช้ซีอารีมาแบบไม่มีการกรอง โดยโมเดลตั้งอยู่บนพื้นฐานที่ว่า การเหมือนกันของค่าสหสัมพันธ์เหมือนกันระหว่างหลักทรัพย์ X และ Y นั้นมีช่วงเวลาที่เหมือนกันและไม่เหมือนกัน เราจึงพัฒนาอัลกอริที่มในการเทรดเมื่อค่าสหสัมพันธ์มันมีทิศทางไปในเทรนด์ที่สูงขึ้นด้วยสมการดังต่อไปนี้

$$
\begin{aligned}
&If\ Z_t > SMA(Z_t)_{10}\ Then \\
&\acute{X}_t = \hat{X}_t * \left(1 + \frac{\rho_{X_t Y_t(i)} * \left(\frac{Y_{t-i} - Y_{t-i-1}}{Y_{t-i-1}}\right)}{10}\right) \\
&Else \\
&\acute{X}_t = \hat{X}_t
\end{aligned}
\tag{3.6}
$$

โดยที่

Z_t คือ สัมประสิทธิ์สหสัมพันธ์ระหว่างราคาปิดของหลักทรัพย์ X และ Y

$SMA(Z_t)_{10}$ คือ เส้นค่าเฉลี่ยแบบปรกติของสัมประสิทธิ์สหสัมพันธ์ระหว่างราคาปิดของหลักทรัพย์ X และ Y ณเวลา t เฉลี่ยเป็นจำนวน 10 วัน

$\hat{X}_t$ คือ ราคาการทำนายจากวิธีอารีมา

X_t คือ หลักทรัพย์ X ณ เวลา t

Y_t คือ หลักทรัพย์ Y ณ เวลา t

3.5.2.2 โมเดลเทรดด้วยโมเดลแบบมีการกรองทิศทาง CARIMA-F

โมเดลเป็นการใช้ซีอารีมาแบบไม่มีการกรอง โดยโมเดลตั้งอยู่บนพื้นฐานที่ว่า การเหมือนกันของค่าสหสัมพันธ์เหมือนกันระหว่างหลักทรัพย์ X และ Y นั้นมีช่วงเวลาที่เหมือนกันและไม่เหมือนกัน เราจึงพัฒนาอัลกอริที่มในการเทรดเมื่อค่าสหสัมพันธ์มันมีทิศทางไปในเทรนด์ที่สูงขึ้น จากนั้นเราจะสนใจทิศทางของการ Lead ในที่นี้เราจะนำเฉพาะทิศทางราคาทางด้านลบของหลักทรัพย์ตัว Lead เท่านั้นที่จะนำมาใช้ช่วยในการทำนายด้วยสมการดังต่อไปนี้

$$
If\ Z_t > SMA(Z_t)_{10}\ \&\&\ \left(\frac{Y_{t-i} - Y_{t-i-1}}{Y_{t-i-1}}\right) < 0\ Then
$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\acute{X}_t = \hat{X}_t * (1 + \frac{\rho_{X_t Y_t(i)} * \left(\frac{Y_{t-i} - Y_{t-i-1}}{Y_{t-i-1}}\right)}{10}) \quad (3.7)$$

$$Else$$

$$\acute{X}_t = \hat{X}_t$$

โดยที่

Z_t คือ สัมประสิทธิ์สหสัมพันธ์ระหว่างราคาปิดของหลักทรัพย์ X และ Y

$SMA(Z_t)_{10}$ คือ เส้นค่าเฉลี่ยแบบปรกติของสัมประสิทธิ์สหสัมพันธ์ระหว่างราคาปิดของหลักทรัพย์ Xและ Y ณเวลา t เฉลี่ยเป็นจำนวน 10 วัน

$\hat{X}_t$ คือ ราคาการทำนายจากวิธีอารีมา

X_t คือ หลักทรัพย์ X ณ เวลา t

Y_t คือ หลักทรัพย์ Y ณ เวลา t

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

ผลการทดลองและอภิปรายผลการทดลอง

ในบทนี้ จะนำเสนอที่มาของข้อมูล และผลการวิจัยของโมเดลการทำนายทั้ง 3 แบบ โดยในการพัฒนาโปรแกรมผู้เขียนได้ใช้โปรแกรมภาษา R ในการทำงานวิจัยครั้งนี้ โดยจะแบ่งเป็น 3 ส่วนการวัดผล และจบด้วยข้อสรุปของการวิจัยครั้งนี้ โดยที่จะมีโครงสร้างดังนี้

4.1. ข้อมูลและแหล่งข้อมูล

4.2. ผลการพยากรณ์โดย ARIMA และ CARIMA ทั้งสองโมเดล

4.3. อภิปรายผลการเทรดด้วยโมเดล CARIMA-NF และ CARIMA-F

4.4. ข้อสรุปงานวิจัย

4.1. ข้อมูลและแหล่งข้อมูล

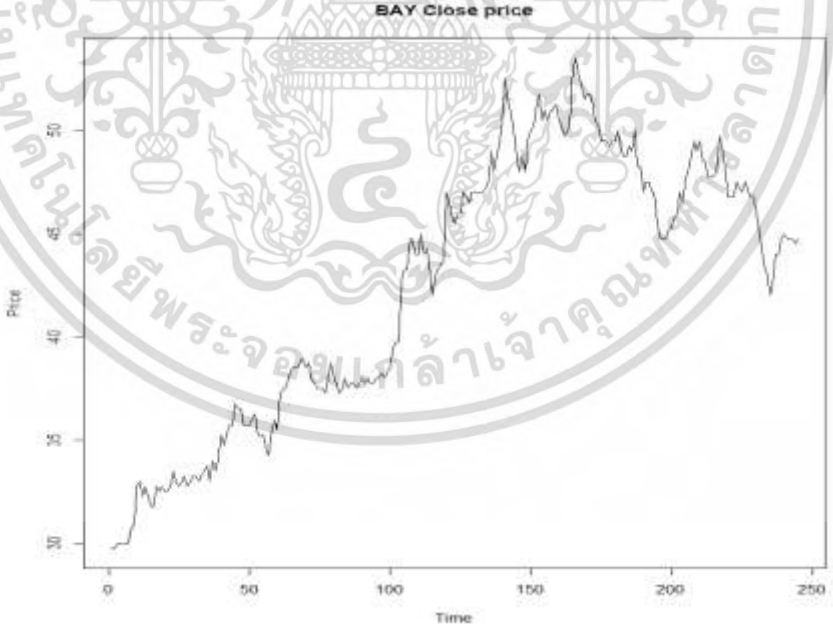
ผู้จัดทำได้ทำการเก็บรวมรวบข้อมูลที่นำมาใช้ในการวิจัยได้แก่ข้อมูลราคาของหลักทรัพย์ตระกูล SET50 จำนวน 50 หลักทรัพย์ SET50 คือ บริษัทหลักทรัพย์ในบริษัทหลักทรัพย์ที่มีสภาพคล่องสูงสุด 50 อันดับแรกของประเทศไทย กล่าวคือ มีปริมาณการซื้อขายต่อวันสูงมาก และ ปริมาณการซื้อขายต่อวัน โดยได้ทำการเก็บรวบรวมข้อมูลมาจาก http://www.set.or.th/

ข้อมูลหลักทรัพย์ในตระกูล SET50 ประกอบด้วยหลักทรัพย์ ADVANC, AOT, BANPU, BAY, BBL, BCP, BDMS, BEC, BH, BIGC, BJC, BLA, BTS, CENTEL, CPALL, CPF, CPN, DELTA, DTAC, EGCO, GLOBAL, GLOW, HMPRO, INTUCH, IRPC, IVL, JAS, KBANK, KKP, KTB, LH, M, MINT, PS, PTT, PTTEP, PTTGC, RATCH, ROBINS, SCB, SCC, SCCC, TCAP, THCOM, TMB, TRUE, TOP, TTW, TUF, VGI โดยข้อมูลทั้งหมดประกอบด้วย ราคาเปิด ราคาสูงสุด ราคาต่ำสุด ราคาปิด และ จำนวนการซื้อขาย ตั้งแต่วันที่ 1 มกราคม 2014 จนถึงวันที่ 31 ธันวาคม 2015 จำนวนรวม 250 วัน โดยในงานวิจัยนี้ได้เลือกเฉพาะราคาปิดมาใช้ในการทดลองทั้งการ ทำนาย และ การคำนวนสหสัมพันธ์แบบข้าม(Cross Correlation) เพราะมองว่า **ราคาปิด** ได้แสดงถึงลักษณะนิสัยของหลักทรัพย์นั้นๆได้ดีที่สุด ตัวอย่างของข้อมูลดังตารางที่ 4.1 และข้อมูลเฉพาะราคาปิดในช่วงเวลาที่ใช้ในการทำวิจัย ดังรูปที่ 4.1 ราคาจริงของหลักทรัพย์ BAY และ รูปที่ 4.2 การเคลื่อนไหวของดัชนีหลักทรัพย์ SET50

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.1 ตัวอย่างข้อมูลหลักทรัพย์ BAY จากวันที่ 1 มกราคม 2014 จึงวันที่ 14 มกราคม 2014

Y-M-D วัน-เดือน-ปี	OPEN ราคาเปิด	HIGH ราคาสูงสุด	LOW ราคาต่ำสุด	CLOSE ราคาปิด	VOLUME ปริมาณ
2014-01-01	29.75	30.25	29.75	29.75	1,894,333
2014-01-02	30.00	30.25	29.75	29.75	1,533,635
2014-01-03	30.00	30.00	29.75	30	1,522,725
2014-01-04	30.00	30.5	30.00	30.00	3,043,550
2014-01-07	30.00	30.25	30.00	30.00	2,235,900
2014-01-08	30.25	30.5	30.00	30.00	575,702
2014-01-09	30.00	30.25	30.00	30.00	452,100
2014-01-10	30.00	30.75	30	30.75	1,267,900
2014-01-11	30.75	31.00	30.25	31.00	1,099,058
2014-01-14	30.00	30.25	30.00	30.00	2,235,900



รูปที่ 4.1. กราฟราคาปิดของหลักทรัพย์ในช่วงเวลาที่ใช้ในการทดลองของหลักทรัพย์ BAY โดยเริ่มตั้งแต่วันที่ 1 มกราคม 2014 จนถึง วันที่ 31 ธันวาคม 2014

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 4.2 กราฟราคาปิดของหลักดัชนี SET50 ในช่วงเวลาที่ใช้ในการทดลองโดยเริ่มตั้งแต่ วันที่ 1 มกราคม 2014 จนถึง วันที่ 31 ธันวาคม 2014

4.2. อภิปรายผลการพยากรณ์โดยโมเดล ARIMA และ โมเดล CARIMA

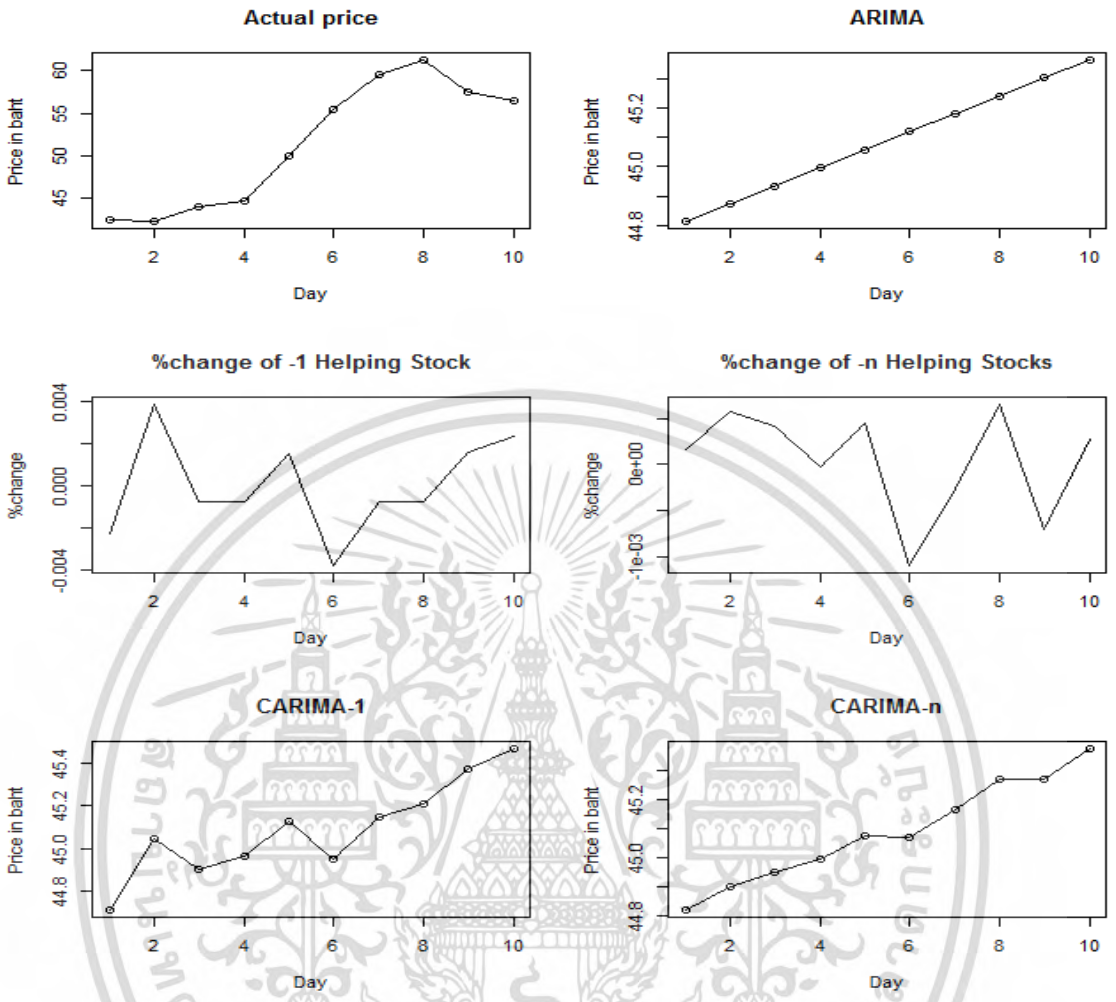
ในวิทยานิพนธ์เล่มนี้ได้ผู้จัดทำได้ทำการทดลองการใช้หลักทรัพย์อื่นมาช่วยในการทำนายหลักทรัพย์ที่เราต้องการทำนาย โดยเปรียบเทียบ 3 โมเดล ด้วย 3 ตัวชี้วัด คือ

1. ARIMA การทำนายด้วยโมเดล ARIMA แบบธรรมดา

2. CARIMA-1 คือการแบบใช้โมเดล ARIMA โดยนำหลักทรัพย์อื่นมาช่วยในการทดลองที่มีค่าสหสัมพันธ์สูงที่สุดกับหลักทรัพย์ที่ต้องการทำนายเพียงตัวเดียว

3. CARIMA-n คือการแบบใช้โมเดล ARIMA โดยนำหลักทรัพย์อื่นมาช่วยในการทดลองที่มีค่าสหสัมพันธ์สูงกว่า 0.8 กับหลักทรัพย์ที่ต้องการทำนายทั้งหมดแบบถ่วงน้ำหนักด้วยค่าสัมประสิทธิ์ของค่าสหสัมพันธ์ของแต่ละตัว

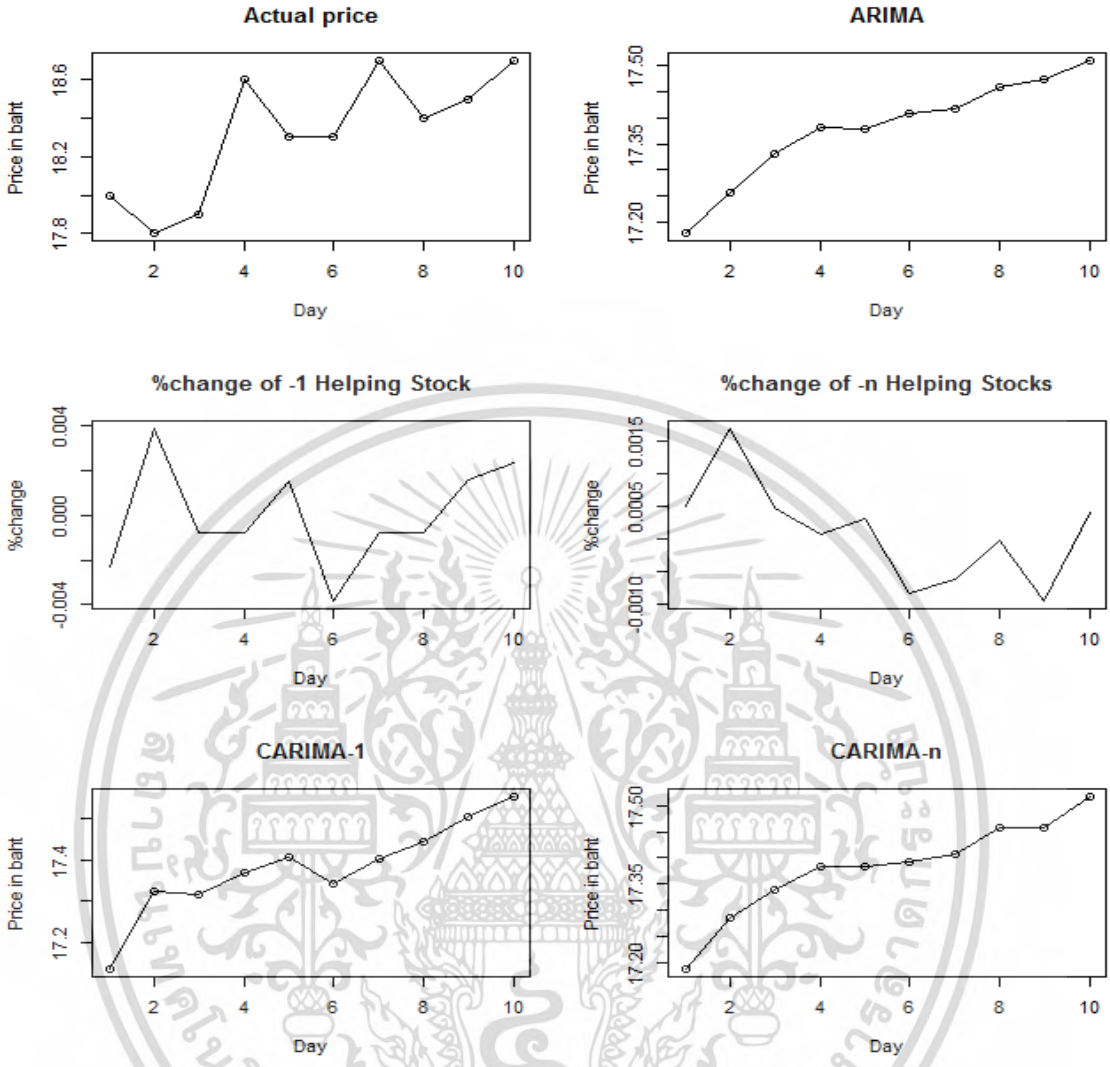
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.3 กราฟเปรียบเทียบผลการทำนายหลักทรัพย์ BAY เป็นเวลา 10 วันจากวันที่ 1 มกราคม 2015 จนถึงวันที่ 13 มกราคม 2015

ผลการทดลอง ในรูปที่ 4.3 กราฟบนซ้ายคือกราฟราคาหลักทรัพย์ BAY ในระยะเวลาการทำนาย กราฟบนขวาคือกราฟราคาหลักทรัพย์ BAY ทำนายโดย ARIMA โดยมี MAPE 14.62% Correlation 0.91 ตามลำดับ กราฟซ้ายล่างคือกราฟราคาหลักทรัพย์ BAY ทำนายโดย CARIMA-1 ที่ใช้หลักทรัพย์ PS ตั้งแต่วันที่ 17-30 ธันวาคม 2014 ดังรูป %change of -1 helping stock มาช่วยในการทำนาย โดยมี MAPE 14.62% Correlation 0.70 และ กราฟขวาล่างคือกราฟราคา BAY ทำนายโดย CARIMA-n ที่ใช้หลักทรัพย์ BDMS CENTEL HMPRO KTB M PS TTW ตั้งแต่วันที่ 17-30 ธันวาคม 2014 และ หลักทรัพย์ BLA ตั้งแต่วันที่ 12-25 ธันวาคม 2014 ถ่วงน้ำหนัก ดังรูป %change of -n helping stock โดยมี MAPE 14.63% Correlation 0.88 แม้จะมีค่า MAPE ที่สูงแต่มันมี Correlation ที่สูงมากในการเทรดจริงๆจึงมีประโยชน์มากกว่า

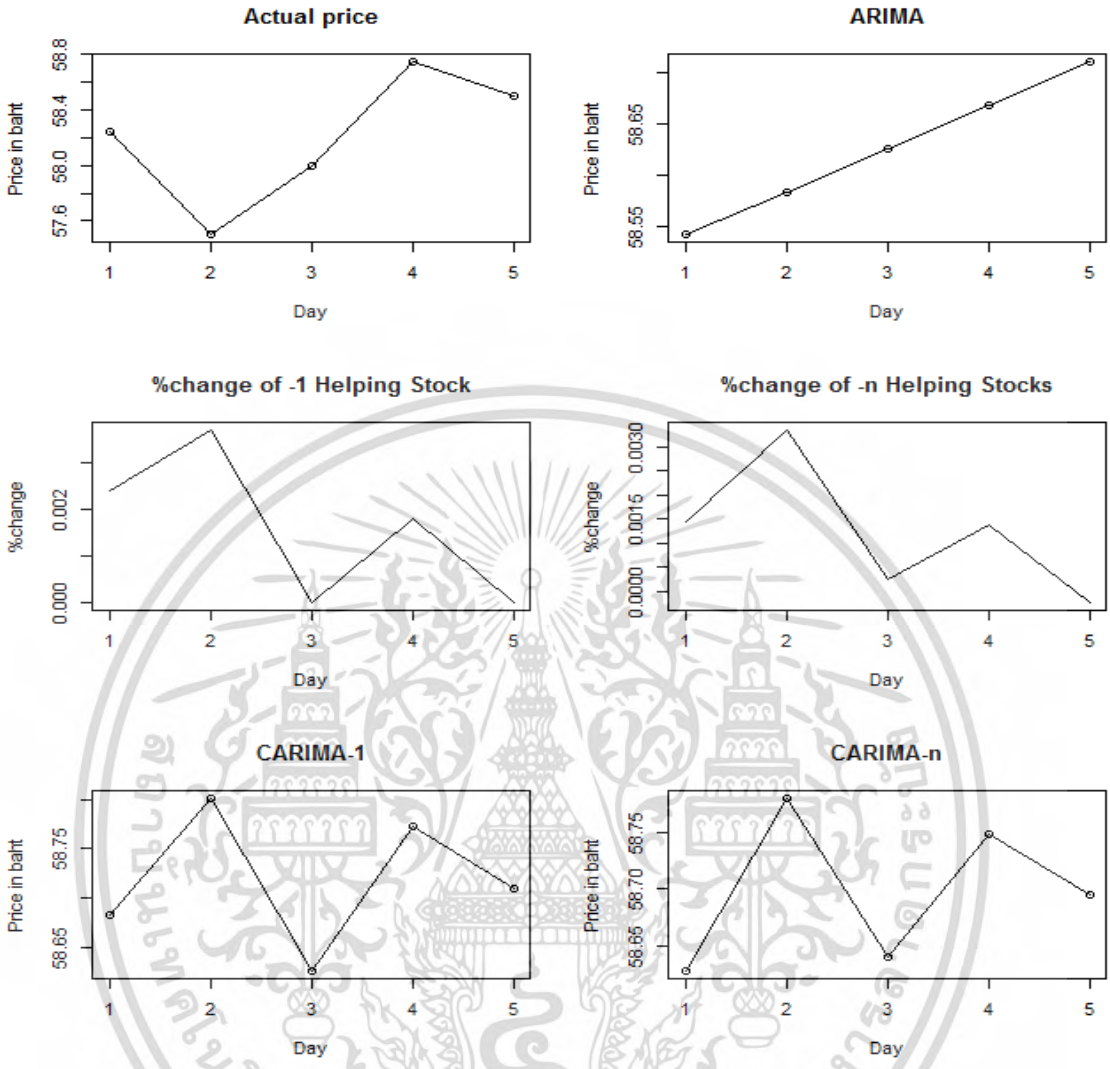
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.4 กราฟเปรียบเทียบผลการทำนายหลักทรัพย์ BDMS เป็นเวลา 10 วันจากวันที่ 1 มกราคม 2015 จนถึงวันที่ 13 มกราคม 2015

ผลการทดลอง ในรูปที่ 4.4 กราฟบนซ้ายคือกราฟราคาหลักทรัพย์ BDMS ในระยะเวลาการทำนาย กราฟบนขวาคือกราฟราคาหลักทรัพย์ BDMS ทำนายโดย ARIMA โดยมี MAPE 5.13 % Correlation 0.68 ตามลำดับ กราฟซ้ายล่างคือกราฟราคาหลักทรัพย์ BDMS ทำนายโดย CARIMA-1 ที่ใช้หลักทรัพย์ HMPRO ตั้งแต่วันที่ 17-30 ธันวาคม 2014 ดังรูป %change of -1 helping stock มาช่วยในการทำนาย โดยมี MAPE 5.13% Correlation 0.46 และ กราฟขวาล่างคือกราฟราคา BDMS ทำนายโดย CARIMA-n ที่ใช้หลักทรัพย์ BDMS CENTEL HMPRO KTB M PS TTW ตั้งแต่วันที่ 17-30 ธันวาคม 2014 ดังรูป %change of -n helping stock โดยมี MAPE 5.12% Correlation 0.63

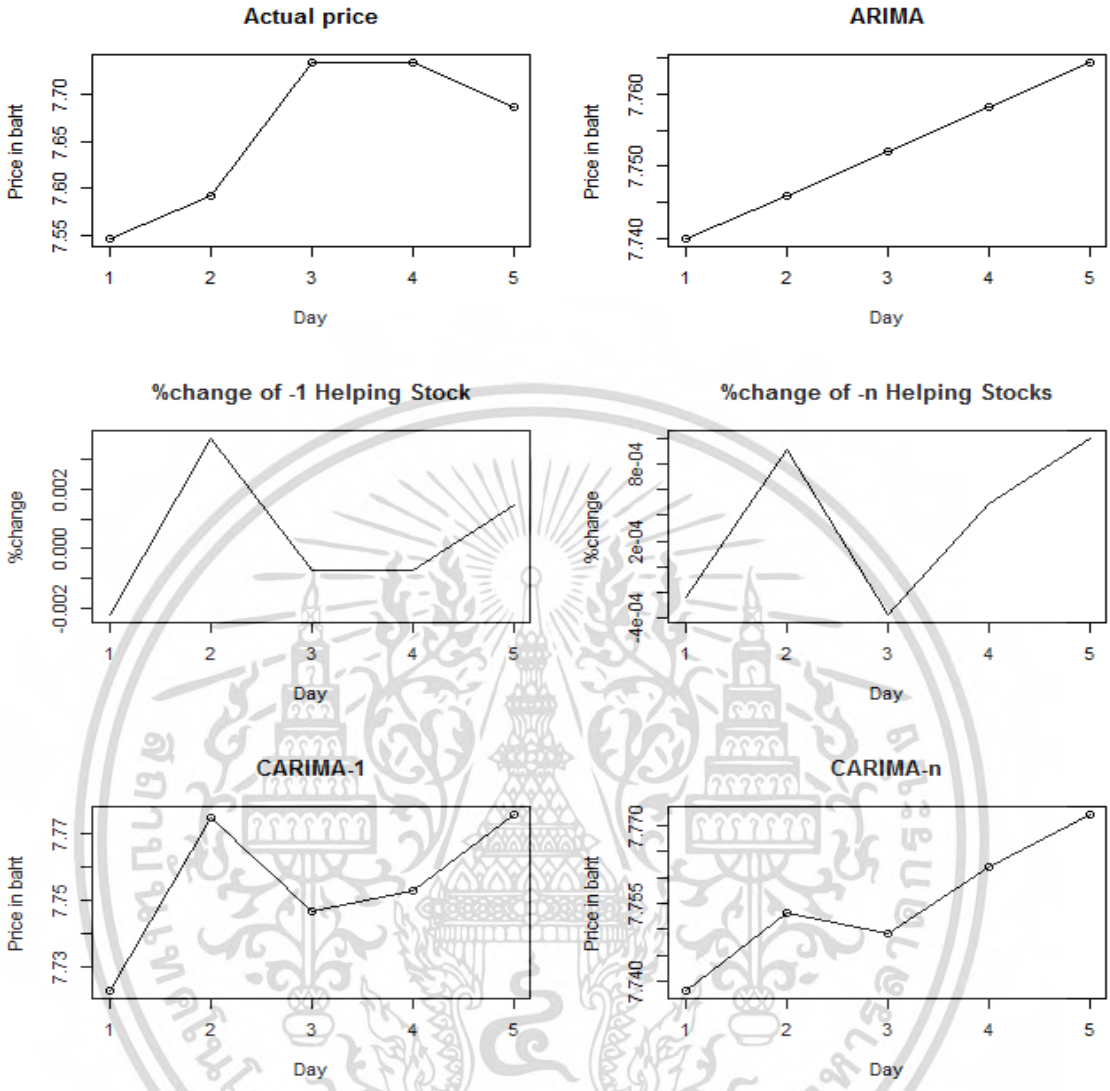
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.5 กราฟเปรียบเทียบผลการทำนายหลักทรัพย์ M เป็นเวลา 5 วันจากวันที่ 1 มกราคม 2015 จนถึงวันที่ 7 มกราคม 2015

ผลการทดลอง ในรูปที่ 4.5 กราฟบนซ้ายคือกราฟราคาหลักทรัพย์ M ในระยะเวลาการทำนาย 5 วัน กราฟบนขวาคือกราฟราคาหลักทรัพย์ M ทำนายโดย ARIMA โดยมี MAPE 0.79% Correlation 0.69 ตามลำดับ กราฟซ้ายล่างคือกราฟราคาหลักทรัพย์ M ทำนายโดย CARIMA-1 ที่ใช้หลักทรัพย์ CPN ดังรูป %change of -1 helping stock มาช่วยในการทำนาย ตั้งแต่วันที่ 24-30 ธันวาคม 2014 โดยมี MAPE 0.89% Correlation -0.61 และ กราฟขวาล่างคือกราฟราคา M ทำนายโดย CARIMA-n ที่ใช้หลักทรัพย์ CPALL CPN ตั้งแต่วันที่ 24-30 ธันวาคม 2014 ดังรูป %change of -n helping stock มาช่วยในการทำนาย โดยมี MAPE 0.86% Correlation -0.69 เพราะหลักทรัพย์ตัวนำที่นำมาใช้ในเวลานั้นมีค่าสัมประสิทธิ์ของสหสัมพันธ์สูงสุดที่ Lag 0 จึงทำให้ไม่เหมาะสมนักที่จะนำมาใช้ในการทำนาย

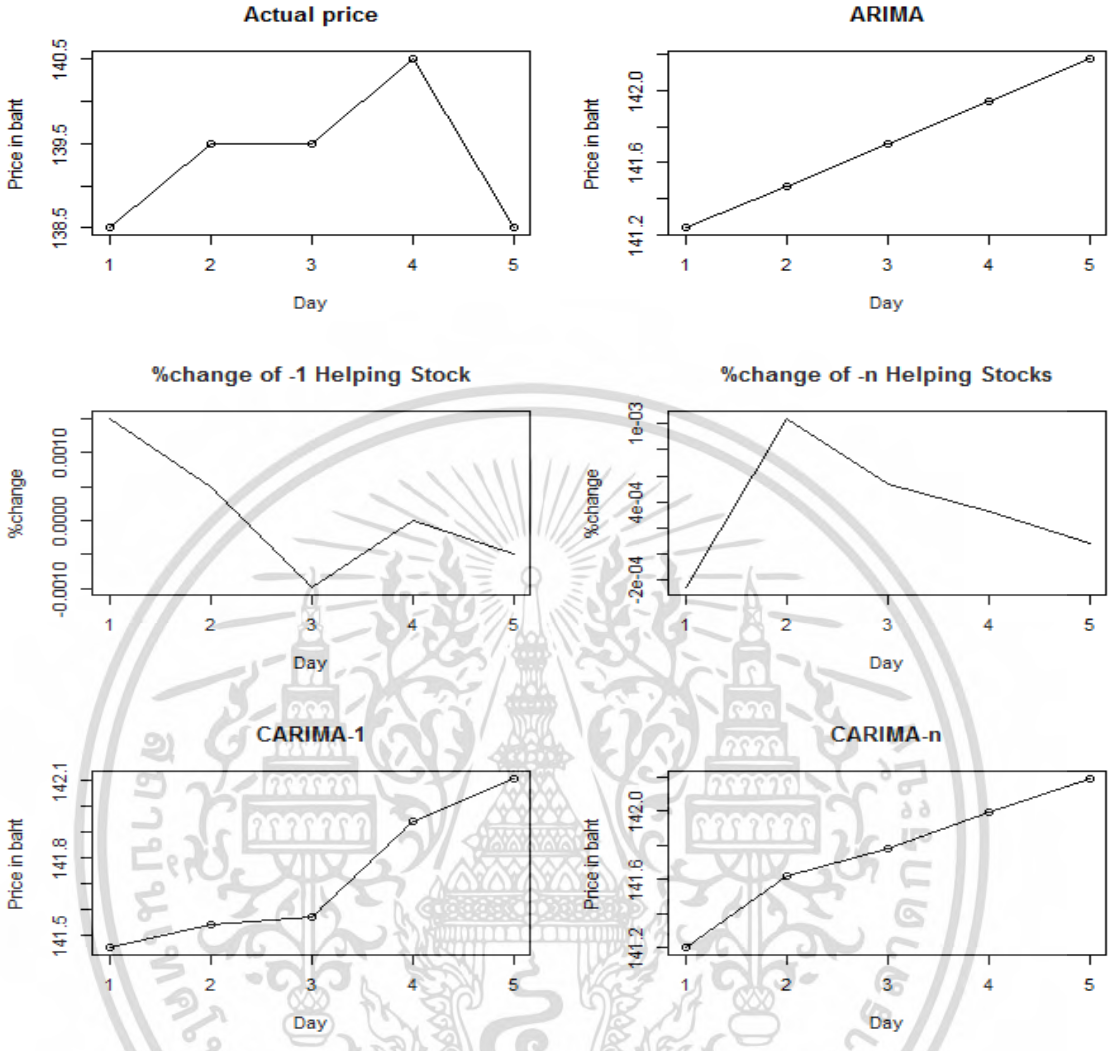
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.6 กราฟเปรียบเทียบผลการทำนายหลักทรัพย์ HMPRO เป็นเวลา 5 วันจากวันที่ 1 มกราคม 2015 จนถึง วันที่ 7 มกราคม 2015

ผลการทดลอง ในรูปที่ 4.6 กราฟบนซ้ายคือกราฟราคาหลักทรัพย์ HMPRO ในระยะเวลาการทำนาย 5 วัน กราฟบนขวาคือกราฟราคาหลักทรัพย์ HMPRO ทำนายโดย ARIMA โดยมี MAPE 1.22% Correlation 0.53 ตามลำดับ กราฟซ้ายล่างคือกราฟราคาหลักทรัพย์ HMPRO ทำนายโดย CARIMA-1 ที่ใช้หลักทรัพย์ PS ตั้งแต่วันที่ 17-30 ธันวาคม 2014 ดังรูป %change of -1 helping stock มาช่วยในการทำนาย โดยมี MAPE 1.25% Correlation 0.03 และ กราฟขวาล่างคือกราฟราคา HMPRO ทำนายโดย CARIMA-n ใช้หลักทรัพย์ BAY CENTEL M PS ตั้งแต่วันที่ 24-30 ธันวาคม 2014 และ หลักทรัพย์ BLA ตั้งแต่วันที่ 12-18 ธันวาคม 2014 ดังรูป %change of -n helping stock โดยมี MAPE 1.26% Correlation 0.56

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.7 กราฟเปรียบเทียบผลการทำนายหลักทรัพย์ BH เป็นเวลา 5 วันจากวันที่ 1 มกราคม 2015 จนถึง วันที่ 7 มกราคม 2015

ผลการทดลอง ในรูปที่ 4.7 กราฟบนซ้ายคือกราฟราคาหลักทรัพย์ BH ในระยะเวลาการทำนาย 5 วัน กราฟบนขวาคือกราฟราคาหลักทรัพย์ BH ทำนายโดย ARIMA โดยมี MAPE 4.95% Correlation 0.21 ตามลำดับ กราฟซ้ายล่างคือกราฟราคาหลักทรัพย์ BH ทำนายโดย CARIMA-1ใช้หลักทรัพย์ BDMS โดยมีดังรูป %change of -1 helping stock มาช่วยในการทำนาย MAPE 3.52% Correlation 0.15 และกราฟขวาล่างคือกราฟราคา BH ทำนายโดย CARIMA-n ใช้หลักทรัพย์ TTW KBANK GLOW BDMS ตั้งแต่วันที่ 24-30 ธันวาคม 2014 และ หลักทรัพย์ HMPRO PTT ตั้งแต่วันที่ 15-19 ธันวาคม 2014 ดังรูป %change of -n helping stock มาช่วยในการทำนาย โดยมี MAPE 4.96% Correlation 0.37

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากกราฟจะเห็นได้ว่า ผลการทำนายของ ARIMA นั่นมีผลออกมาเป็นเส้นตรงในหลายๆครั้ง ทำให้ขาดรายละเอียดในการทำนายไปบ้าง ในงานวิจัยนี้ผู้จัดทำได้ใช้แนวคิดสหสัมพันธ์แบบข้ามมาเพื่อช่วยในการทำนาย จากผลในกราฟแสดงให้เห็นว่า CARIMA-1 และ CARIMA-n นั้นจะมีผลการทำนายที่มีรายละเอียดมากกว่าการทำนายของ ARIMA แต่ในด้านของสหสัมพันธ์ที่ตั้งเป้าว่าจะนำมาช่วยในการทำนายนั้นกลับพบว่า ARIMA ให้ค่าที่ดีกว่าทั้งสหสัมพันธ์และสหสัมพันธ์แบบระยะทาง (Distance Correlation) ขณะที่ ค่าร้อยละความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Percentage Error) ไม่ต่างกันมากนัก ดังเราจะกล่าวถึงในหัวข้อต่อไป

4.2.1 ค่าร้อยละความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Percentage Error)

ในส่วนนี้จะแสดงให้ถึงค่าร้อยละความคลาดเคลื่อนสัมบูรณ์ของการทำนายจากโมเดลการทำนายด้วยวิธีการทำนายแบบ ARIMA ตามสมการที่ (2.17), CARIMA-1 โดยที่ 1 คือ หลักทรัพย์ที่ค่าสหสัมพันธ์ในการ lead สูงสุดจำนวน 1 ตัว ตามสมการที่ (3.2), CARIM-n โดยที่ n คือจำนวนหลักทรัพย์ที่ผ่านเกณฑ์ตามสมการที่ (3.3) โดยวัดผลจากการทำนายหลักทรัพย์นั้นๆจากวันที่เป็นเวลา 10 วันมาเปรียบเทียบกัน โดยหลักทรัพย์ในคอลั่ม ที่มีค่าเป็น NA คือไม่มีหลักทรัพย์ที่เข้าข่ายมาช่วยทำนายในโมเดลนั้น

ตารางที่ 4.2. ผลการทดลองเปรียบเทียบค่าร้อยละความคลาดเคลื่อนสัมบูรณ์ของ ARIMA, CARIMA-1, CARIMA-n จากการทำนายเป็นเวลา 10 วัน

Name	ARIMA	CARIMA-1	CARIMA-n
ADVANC	1.75%	NA	NA
AOT	1.68%	1.69%	1.72%
BANPU	1.66%	NA	NA
BAY	14.62%	14.62%	14.63%
BBL	3.07%	3.12%	3.08%
BCP	1.68%	NA	NA
BDMS	5.13%	5.13%	5.12%
BEC	2.81%	NA	NA
BH	2.15%	2.13%	2.16%
BIGC	1.47%	1.54%	1.44%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Name	ARIMA	CARIMA-1	CARIMA-n
BJC	2.41%	NA	NA
BLA	1.03%	1.03%	1.07%
BTS	4.13%	4.08%	4.14%
CENTEL	2.7%	NA	NA
CPALL	1.43%	NA	NA
CPF	1.92%	NA	NA
CPN	1.34%	NA	NA
DELTA	2.66%	NA	NA
DTAC	2.58%	NA	NA
EGCO	1.46%	1.5%	1.45%
GLOBAL	2.20%	NA	NA
GLOW	2.08%	2.12%	2.09%
HMPRO	2.11%	2.11%	2.12%
INTUCH	2.43%	NA	NA
IRPC	4.45%	NA	NA
IVL	3.66%	3.69%	3.69%
JAS	1.92%	NA	NA
KBANK	5.12%	5.13%	5.13%
KKP	0.97%	NA	NA
KTB	2.94%	2.94%	2.94%
LH	1.02%	NA	NA
M	0.65%	0.67%	0.66%
MINT	2.81%	2.87%	2.84%
PS	5.93%	5.96%	5.97%
PTT	2.23%	2.35%	2.25%
PTTEP	5.19%	NA	NA
PTTGC	5.87%	5.89%	5.88%
RATCH	0.48%	0.52%	0.47%

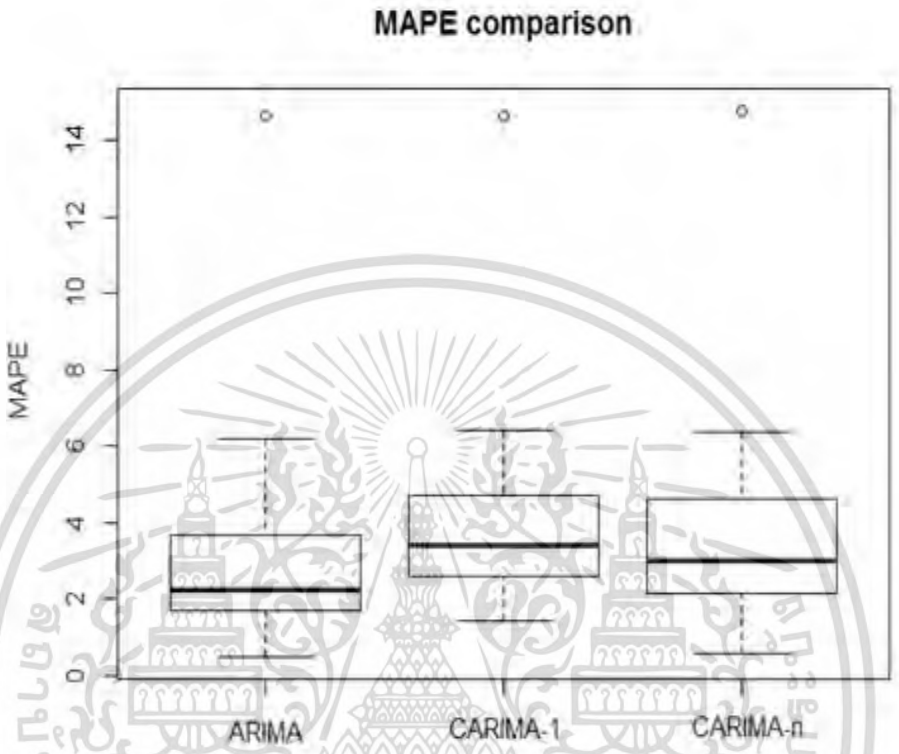
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Name	ARIMA	CARIMA-1	CARIMA-n
ROBINS	3.44%	NA	NA
SCB	1.22%	1.26%	1.23%
SCC	1.02%	NA	NA
SCCC	2.25%	2.25%	2.25%
TCAP	2.01%	NA	NA
THCOM	5.72%	NA	NA
TMB	1.85%	1.88%	1.91%
TOP	6.17%	6.19%	6.19%
TRUE	6.00%	5.97%	6.00%
TTW	3.33%	3.33%	3.32%
TUF	4.28%	4.26%	4.26%
VGI	1.67%	NA	NA

จากตารางที่ 4.2 จะเห็นได้ว่าค่าร้อยละความคลาดเคลื่อนสัมบูรณ์ของทั้ง 3 โมเดลการทำนาย ค่อนข้างใกล้เคียงกัน โดยไม่ต่างกันไม่มากนักแต่โดยภาพรวม โมเดลการทำนาย ARIMA จะให้ค่าร้อยละความคลาดเคลื่อนสัมบูรณ์ต่ำที่สุด และ CARIMA-1 จะให้ค่าร้อยละความคลาดเคลื่อนสัมบูรณ์มากที่สุด และ CARIMA-n จะให้ค่าคลาดเคลื่อนสัมบูรณ์รองลงมา แต่ก็ไม่ได้ต่ำกว่า CARIMA-1 มากนัก โดยที่ผลออกมาในรูปแบบนี้เพราะว่า ผลการทดลองที่ได้ สามารถอธิบายได้ ดังนี้ CARIMA-1 และ CARIMA-n เป็นการใช้โมเดล ARIMA ทำนายผลโดยอาศัยค่าจากหลักทรัพย์อื่นเข้ามาช่วยในการทำนาย โดยที่หลักทรัพย์ส่วนใหญ่ที่จะนำมาใช้ในการทำนายจะเป็นหลักทรัพย์ที่มีค่า Coefficients สูงสุดที่ lag = 0 ซึ่งถ้าเราใช้ค่าที่ lag = 0 นี้ ก็จะเปรียบเสมือนกับการใช้ของตนเองในวันก่อนหน้า มาตัดสินค่าในวันที่ตามมา ทำให้ค่าที่ได้มีลักษณะของการสุ่ม เมื่อทำการวิเคราะห์ผลลัพธ์ที่ได้จากโมเดล CARIMA-1 และ CARIMA-n จะพบว่า CARIMA-1 ให้ค่าร้อยละความคลาดเคลื่อนสัมบูรณ์มากกว่า เนื่องจาก เป็นการเลือกเฉพาะหลักทรัพย์ที่มีค่าสหสัมพันธ์สูงสุดเท่านั้นมาช่วยในการทำนาย ซึ่งมักจะเป็นค่าของตนเองในวันก่อนหน้าดังกล่าวมาแล้วข้างต้น ส่วนโมเดล CARIMA-n เป็นโมเดลที่เลือกกลุ่มของหลักทรัพย์ที่ใช้ค่าสหสัมพันธ์สูงเกินที่หลักเกณฑ์กำหนดไว้มากกว่าหนึ่งตัวมาช่วยในการทำงานผล ซึ่งทำให้มีการใช้ก็มูลของหลักทรัพย์อื่นๆ มาใช้ในการถ่วงน้ำหนัก ทำให้ค่าร้อยละความคลาดเคลื่อนสัมบูรณ์ลดลงกว่า CARIMA-1 เพราะการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ถ่วงน้ำหนักโดยหลักทรัพย์มากกว่า 1 หลักทรัพย์นั้นทำให้ค่าการเปลี่ยนแปลงที่เรานำมาช่วยในการทำนายนั้นไม่มากนัก เมื่อมีการเหวี่ยงไม่มากนัก ค่าคลาดเคลื่อนสัมบูรณ์จึงไม่มากตามไปด้วย



รูปที่ 4.8 ผลการทดลองเปรียบเทียบค่าร้อยละความคลาดเคลื่อนสัมบูรณ์ของ ARIMA, CARIMA-1 และ CARIMA-n

จากข้อมูลในรูปที่ 4.8 เราจะพบว่าจะมีค่าการคลาดเคลื่อนเป็น Outlier อยู่ 1 หลักทรัพย์ คือหลักทรัพย์ BAY เนื่องจากราคาหลักทรัพย์ BAY นั้น ได้มีการเคลื่อนไหวของราคารุนแรงที่สุดจึงทำให้การทำนายเนื่องจากช่วงปีใหม่ปี 2015 หลักทรัพย์ BAY เป็นหลักทรัพย์ตัวนำตลาดการเคลื่อนไหวจึงผันผวนสูงมาก สมการเชิงเส้นของ ARIMA ไม่สามารถนำมาอธิบายได้ทันจึงทำให้มีค่าคลาดเคลื่อนสัมบูรณ์โดดที่สุด แต่ถึงแม้ค่าร้อยละความคลาดเคลื่อนสัมบูรณ์ของ BAY จะสูงสุด แต่ในเชิงสหสัมพันธ์กลับมีค่าสูงที่สุด ดังจะกล่าวถึงต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.2 ค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เปรียบเทียบโดยระยะเวลาการทำนาย

ในส่วนนี้ จะเป็นการทำนายของตัวแบบโมเดลทั้ง 3 โมเดล คือ ARIMA ตามสมการที่ (2.17), CARIMA-1 โดยที่ 1 คือ หลักทรัพย์ที่ค่าสหสัมพันธ์ในการ lead สูงสุดจำนวน 1 ตัว ตามสมการที่ (3.2), CARIM-n โดยที่ n คือจำนวนหลักทรัพย์ที่ผ่านเกณฑ์ตามสมการที่ (3.3) ด้วยค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เปรียบเทียบตามระยะเวลา โดยวัดผลตั้งแต่การทำนาย 1 วัน จนถึง 10 วัน

ตารางที่ 4.3 ผลการทดลองเปรียบเทียบค่าร้อยละความคลาดเคลื่อนเฉลี่ยของการทำนายหลักทรัพย์ทั้งหมดระหว่าง ARIMA, CARIMA-1 และ CARIMA–2 เปรียบเทียบโดยระยะเวลาการทำนายจาก 10 วัน จนถึง 1 วัน

Day	ARIMA	CARIMA-1	CARIMA-n
10	2.97	3.49	3.46
9	2.90	3.43	3.42
8	2.82	3.38	3.36
7	2.72	3.28	3.27
6	2.60	3.14	3.12
5	2.47	2.98	2.96
4	2.47	2.97	2.96
3	2.49	3.13	3.11
2	2.46	3.18	3.15
1	2.08	2.62	2.62

จากผลการทดลองในตารางที่ 4.3 จะเห็นได้ว่า โมเดลทั้ง 3 โมเดลมีแนวโน้มที่จะทำนายได้ดีในวันที่ใกล้เคียงมากกว่าวันที่ไกลออกไป ซึ่งสามารถอธิบายได้ว่า ค่าร้อยละความคลาดเคลื่อนในการทำนายจะแปรผกผันกับค่า **สหสัมพันธ์**ของข้อมูล ในที่นี้เนื่องจากโดยส่วนใหญ่แล้ว ค่า**สหสัมพันธ์**จะสูงที่สุดที่วันที่ตรงกัน หรือ lag = 0 และทำลดลงเรื่อยๆ เมื่อจำนวนวันที่ทำนายห่างออกไปมากขึ้น ซึ่งค่าร้อยละความคลาดเคลื่อนในการทำนายเพิ่มขึ้นเรื่อยๆ ในขณะเดียวกัน ค่าร้อยละความคลาดเคลื่อนของ CARIMA-n ยังคงให้ผลดีกว่า CARIMA-1 ด้วยเหตุผลเดียวกับข้างต้น แต่ปัญหาของการวัดด้วยค่าร้อยละความคลาดเคลื่อนนั้นเมื่อนำมาใช้ทำนายเพื่อการลงทุนจริงๆกลับไม่ได้ให้ผลดีนัก หลักทรัพย์ BAY ที่ให้ค่าความคลาดเคลื่อนสูงสุดกลับให้ผลลัพธ์ความถูกต้องมากที่สุด ดังจะกล่าวถึงต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.3 ค่าสหสัมพันธ์ (Correlation)

ในส่วนนี้จะแสดงให้ถึงค่าความสหสัมพันธ์ของการทำนายจากโมเดลการทำนายด้วยวิธีการ ARIMA ตามสมการที่ (2.17), CARIMA-1 โดยที่ 1 คือ หลักทรัพย์ที่ค่าสหสัมพันธ์ในการ lead สูงสุดจำนวน 1 ตัว ตามสมการที่ (3.2), CARIM-n โดยที่ n คือจำนวนหลักทรัพย์ที่ผ่านเกณฑ์ตามสมการที่ (3.3) โดยวัดผลจากการทำนายหลักทรัพย์นั้นๆจากวันที่เป็นเวลา 10 วัน โดยนำมาเปรียบเทียบกัน

ตารางที่ 4.4 ผลการทดลองเปรียบเทียบค่าสหสัมพันธ์เฉลี่ยของการทำนายหลักทรัพย์ทั้งหมดระหว่าง ARIMA, CARIMA-1 และ CARIMA-n จากการทำนายเป็นเวลา 10 วัน

Name	ARIMA	CARIMA-1	CARIMA-n
ADVANC	0.56	NA	NA
AOT	0.69	0.55	0.65
BANPU	0.01	NA	NA
BAY	0.90	0.70	0.88
BBL	0.18	0.4	0.25
BCP	0.22	NA	NA
BDMS	0.79	0.69	0.76
BEC	-0.75	NA	NA
BH	0.08	-0.02	0.13
BIGC	-0.47	-0.66	-0.50
BJC	0.71	NA	NA
BLA	-0.59	-0.34	-0.52
BTS	0.14	0.14	-0.44
CENTEL	0.07	NA	NA
CPALL	-0.26	NA	NA
CPF	-0.13	NA	NA
CPN	0.44	NA	NA
DELTA	0.08	NA	NA
DTAC	-0.38	NA	NA
EGCO	0.14	0.05	0.10

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

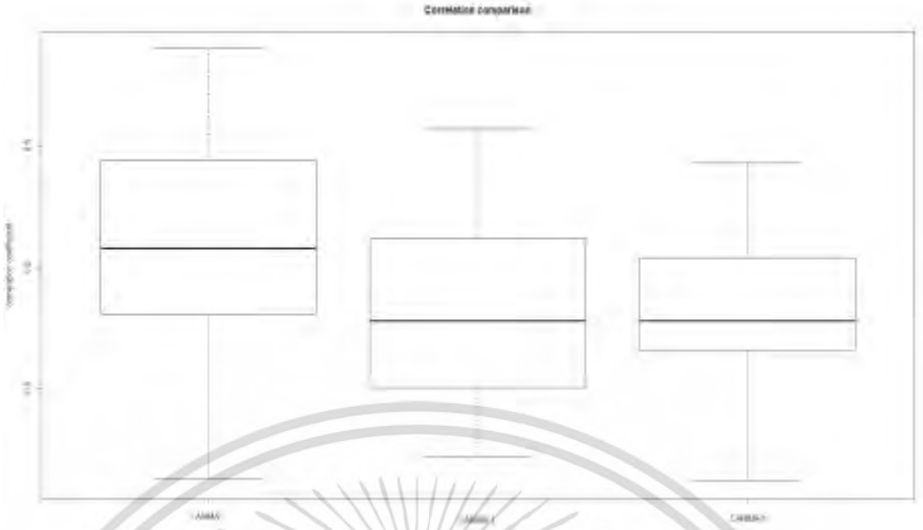
Name	ARIMA	CARIMA-1	CARIMA-n
GLOBAL	-0.01	NA	NA
GLOW	0.28	-0.03	0.21
HMPRO	-0.50	-0.46	-0.50
INTUCH	-0.64	NA	NA
IRPC	-0.74	NA	NA
IVL	0.34	0.20	0.20
JAS	0.42	NA	NA
KBANK	0.61	0.60	0.58
KKP	0.81	NA	NA
KTB	0.21	-0.10	0.08
LH	-0.05	NA	NA
M	0.66	0.57	0.56
MINT	0.03	-0.15	-0.10
PS	-0.87	-0.88	-0.88
PTT	-0.14	-0.37	-0.15
PTTEP	0.43	NA	NA
PTTGC	-0.13	-0.10	-0.14
RATCH	-0.12	-0.43	-0.17
ROBINS	-0.86	NA	NA
SCB	-0.01	-0.56	-0.65
SCC	0.75	NA	NA
SCCC	-0.19	-0.28	-0.38
TCAP	0.9	NA	NA
THCOM	0.21	NA	NA
TMB	0.8	0.6	0.65
TOP	-0.43	-0.5	-0.07
TRUE	0.84	0.85	0.82
TTW	NA	-0.12	-0.23

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Name	ARIMA	CARIMA-1	CARIMA-n
TUF	0.43	0.02	0.02
VGI	-0.43	NA	NA

จากผลการทดลองในตารางที่ 4.4 จะเห็นได้ว่า ARIMA ให้ผลลัพธ์ที่ดีสำหรับข้อมูลส่วนใหญ่ ตามด้วย CARIMA-n และ CARIMA-1 ตามลำดับ ซึ่งเมื่อทำการเปรียบเทียบผลลัพธ์จากการใช้ CARIMA-1 และ CARIMA-n ในการทำนาย ผลลัพธ์ที่ได้คือ ค่า**สหสัมพันธ์**ระหว่างค่าจริง กับค่าที่ได้จากการทำนาย ของ CARIMA-1 มีค่าต่ำกว่า ค่าที่ได้จาก CARIMA-n ซึ่งจะเห็นได้ว่า ถึงแม้ว่า CARIMA-1 ซึ่งเป็นการเลือกเฉพาะหลักทรัพย์ที่มีค่า**สหสัมพันธ์**สูงที่สุดมาช่วยในการทำนายน่าจะให้ผลลัพธ์ที่ดีที่สุด แต่ในความเป็นจริง CARIMA-n ซึ่งนำกลุ่มของทรัพย์ทรัพย์ที่มีค่าสัมประสิทธิ**สหสัมพันธ์**เกินระดับที่กำหนดไว้ให้ผลลัพธ์ที่ดีกว่า เนื่องจาก CARIMA-1 มีแนวโน้มสูงมากที่จะนำเฉพาะค่าของตนเองในวันก่อนหน้ามาทำนายผล ซึ่งจะได้ผลลัพธ์ที่ลู่เข้าสู่การสุ่มมากกว่า CARIMA-n ในส่วนของโมเดล CARIMA-n การจะได้ผลการทำนายที่ดีนั้นเพราะว่า การที่เราให้น้ำหนักกับหลักทรัพย์มากกว่า 1 ตัวนั้น มันจจะมีการ จะขึ้นอยู่กับข้อมูลที่มีเป็นสำคัญ และ ยังขึ้นอยู่กับการตั้งค่า Threshold อีกด้วย ถ้ามีการตั้งค่าที่เหมาะสมก็จะทำให้ CARIMA-n ทำงานได้อย่างมีประสิทธิภาพมากขึ้น เนื่องจากจะได้กลุ่มของหลักทรัพย์ที่เหมาสมในการทำนายผลมากขึ้น รูปที่ 4.9 คือการแสดงผลของตาราง 4.4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.9 ผลการทดลองเปรียบเทียบค่าสหสัมพันธ์เฉลี่ยของการทำนายหลักทรัพย์ทั้งหมดระหว่าง ARIMA, CARIMA-1 และ CARIMA-n จากการทำนายเป็นเวลา 10 วัน

4.2.4 ค่าสหสัมพันธ์เปรียบเทียบโดยระยะเวลาการทำนาย

การทดลองนี้เป็นการทดลองเปรียบเทียบค่าสหสัมพันธ์โดยระยะเวลาการทำนายของโมเดลทั้ง 3 โมเดล คือ วิธีการ ARIMA ตามสมการที่ (2.17), CARIMA-1 โดยที่ 1 คือ หลักทรัพย์ที่ค่าสหสัมพันธ์ในการ lead สูงสุดจำนวน 1 ตัว ตามสมการที่ (3.2), CARIM-n โดยที่ n คือจำนวนหลักทรัพย์ที่ผ่านเกณฑ์ตามสมการที่ (3.3) ในที่นี้เราเปรียบเทียบตั้งแต่ระยะเวลาการทำนาย 10 วัน ลดลงไปจนเหลือ 6 วัน เนื่องจากถ้าจำนวนวันน้อยเกินไปการวัดด้วยค่าสหสัมพันธ์จะผันผวนมาก จนไม่เหมาะสมที่จะใช้วัด ในที่นี้เราจึงกำหนดไว้ที่ 6 วัน

ตารางที่ 4.5. ผลการทดลองเปรียบเทียบค่าสหสัมพันธ์ระหว่าง ARIMA CARIMA-1 และ CARIMA–n เปรียบเทียบโดยระยะเวลาการทำนายจาก 10 วัน จนถึง 6 วัน

Day	ARIMA	CARIMA-1	CARIMA-n
10	0.103	0.014	0.043
9	0.120	-0.004	0.056
8	0.130	0.001	0.069
7	0.143	-0.011	0.103
6	0.106	-0.040	0.099

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากผลการทดลองในตารางที่ 4.5 จะเห็นได้ว่า โมเดลทั้งสามมีแนวโน้มที่จะทำนายได้ดีในวันที่ใกล้เคียงมากกว่าวันที่ไกลออกไป ซึ่งสามารถอธิบายได้ว่า ค่าร้อยละความคลาดเคลื่อนในการทำนายจะแปรผกผันกับค่า **สหสัมพันธ์**ของข้อมูล ในที่นี้เนื่องจากโดยส่วนใหญ่แล้ว ค่า**สหสัมพันธ์**จะสูงที่สุดที่วันที่ตรงกัน หรือ lag = 0 และทำลดลงเรื่อยๆ เมื่อจำนวนวันที่ทำนายห่างออกไปมากขึ้น ซึ่งค่าร้อยละความคลาดเคลื่อนในการทำนายเพิ่มขึ้นเรื่อยๆ ในขณะเดียวกัน ค่าร้อยละความคลาดเคลื่อนของ CARIMA-n ยังคงให้ผลดีกว่า CARIMA-1 ด้วยเหตุผลเดียวกับข้างต้นคือ การที่โมเดล CARIMA-n ใช้การถ่วงน้ำหนักหลักทรัพย์มากกว่า 1 ตัว ค่าจึงไม่ผันผวนมากเท่า CARIMA-1 ที่ใช้หลักทรัพย์อื่นเพียงตัวเดียว หลักทรัพย์ที่ให้ค่าสหสัมพันธ์สูงเป็นตัวชี้วัดที่บอกได้ว่า การทำนายนั้นๆสามารถช่วยเหลือในการซื้อขายหลักทรัพย์ได้มากกว่าค่าร้อยละความคลาดเคลื่อน

4.2.5 ค่าสหสัมพันธ์แบบระยะทาง

ในส่วนนี้จะแสดงให้ถึงค่าความสหสัมพันธ์แบบระยะทางของการทำนายจากโมเดลการทำนายด้วยวิธีการ ARIMA ตามสมการที่ (2.17), CARIMA-1 โดยที่ 1 คือ หลักทรัพย์ที่ค่าสหสัมพันธ์ในการ lead สูงสุดจำนวน 1 ตัว ตามสมการที่ (3.2), CARIM-n โดยที่ n คือจำนวนหลักทรัพย์ที่ผ่านเกณฑ์ตามสมการที่ (3.3) โดยวัดผลจากการทำนายหลักทรัพย์นั้นๆจากวันที่เป็นเวลา 10 วัน โดยนำมาเปรียบเทียบกัน

ตารางที่ 4.6 ผลการทดลองเปรียบเทียบค่าสหสัมพันธ์แบบระยะทางระหว่าง ARIMA CARIMA-1และ CARIMA-n จากการทำนายเป็นเวลา 10 วัน

Name	ARIMA	CARIMA-1	CARIMA-n
ADVANC	0.67	NA	NA
AOT	0.75	0.62	0.7
BANPU	0.39	NA	NA
BAY	0.95	0.74	0.91
BBL	0.44	0.59	0.59
BCP	0.64	NA	NA
BDMS	0.84	0.76	0.82
BEC	0.74	NA	NA
BH	0.34	0.29	0.38
BIGC	0.6	0.76	0.64

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Name	ARIMA	CARIMA-1	CARIMA-n
BJC	0.79	NA	NA
BLA	0.65	0.52	0.6
BTS	0.38	0.38	0.51
CENTEL	0.34	NA	NA
CPALL	0.58	NA	NA
CPF	0.45	NA	NA
CPN	0.65	NA	NA
DELTA	0.64	NA	NA
DTAC	0.46	NA	NA
EGCO	0.57	0.56	0.55
GLOBAL	0.44	NA	NA
GLOW	0.57	0.46	0.53
HMPRO	0.66	0.6	0.65
INTUCH	0.79	NA	NA
IRPC	0.88	NA	NA
IVL	0.55	0.42	0.42
JAS	0.47	NA	NA
KBANK	0.79	0.78	0.78
KKP	0.87	NA	NA
KTB	0.55	0.43	0.5
LH	0.47	NA	NA
M	0.71	0.67	0.67
MINT	0.57	0.6	0.56
PS	0.91	0.9	0.91
PTT	0.58	0.65	0.56
PTTEP	0.51	NA	NA
PTTGC	0.59	0.6	0.59
RATCH	0.42	0.62	0.42

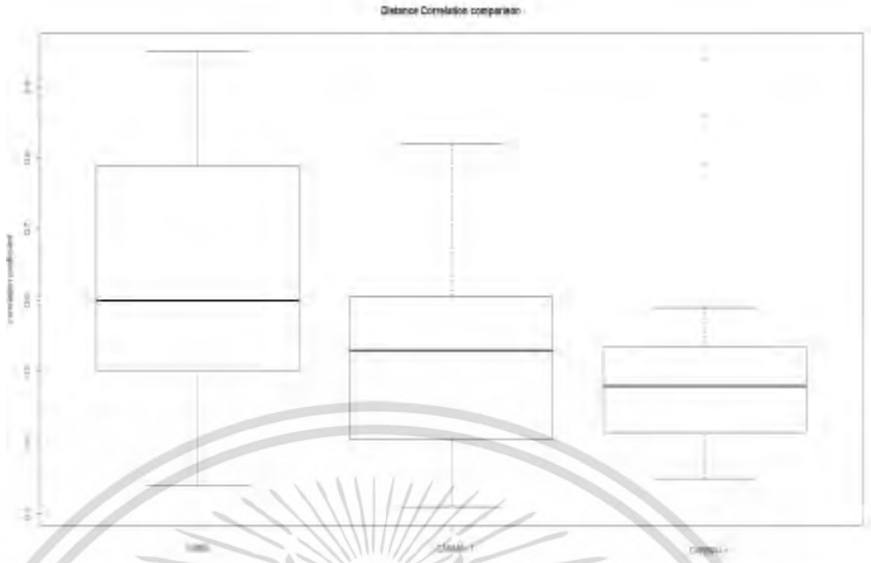
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Name	ARIMA	CARIMA-1	CARIMA-n
ROBINS	0.88	NA	NA
SCB	0.62	0.63	0.84
SCC	0.82	NA	NA
SCCC	0.5	0.43	0.55
TCAP	0.87	NA	NA
THCOM	0.51	NA	NA
TMB	0.88	0.73	0.76
TOP	0.68	0.75	0.47
TRUE	0.89	0.9	0.88
TTW	NA	0.37	0.52
TUF	0.56	0.35	0.35
VGI	0.49	NA	NA
SUMMARY	0.63	0.59	0.62

จากตารางที่ 4.6 จะเห็น ว่า ค่าเฉลี่ยของสหสัมพันธ์แบบระยะทางของโมเดล CARIMA-1 เป็นตัวที่เคลื่อนที่เข้าใกล้การเป็นอิสระต่อกันมากที่สุด ในขณะที่ ARIMA และ CARIMA-n จะให้ผลที่มีทิศทางที่มีผลกับทิศทางของราคาจริงของหลักทรัพย์มากกว่า

ARIMA ใช้โมเดลสมการเชิงเส้นมาอธิบายอย่างเดียวก็จริง แต่ผลการทำนายมาอธิบายกลับให้ค่าดีสุด ต่างกับ CARIMA-1 ได้มีการใช้หลักทรัพย์อื่นมาช่วยทำนายเพียงตัวเดียวนั้นให้ผลผิดพลาดซึ่งจะนำไปสู่การสุ่มหรือการผลักดันการทำนายไปในทางที่เป็นอิสระต่อกันมากกว่า ขณะที่ CARIMA-n แนวโน้มไปในทางเดียวกันกับ CARIMA-1 แต่สิ่งที่ช่วยมันก็คือ การให้น้ำหนักของหลักทรัพย์หลายๆหลักทรัพย์มาช่วยลดการสุ่มออกไปนั่นเอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.10 ผลการทดลองเปรียบเทียบค่าสหสัมพันธ์แบบระยะทางระหว่าง ARIMA CARIMA-1และ CARIMA-n จากการทำนายเป็นเวลา 10 วัน

4.2.6 ค่าสหสัมพันธ์แบบระยะทางเปรียบเทียบโดยระยะเวลาการทำนาย

การทดลองนี้เป็นการทดลองเปรียบเทียบค่าสหสัมพันธ์แบบระยะทางโดยระยะเวลาการทำนายโดยโมเดลทั้ง 3 โมลเดล คือ ARIMA ตามสมการที่ (2.17), CARIMA-1 โดยที่ 1 คือ หลักทรัพย์ที่ค่าสหสัมพันธ์ในการ lead สูงสุดจำนวน 1 ตัว ตามสมการที่ (3.2), CARIM-n โดยที่ n คือจำนวนหลักทรัพย์ที่ผ่านเกณฑ์ตามสมการที่ (3.3) ในที่นี้เราเปรียบเทียบตั้งแต่ระยะเวลาการทำนาย 10 วัน ลดลงไปจนเหลือ 6 วัน เนื่องจากถ้าจำนวนวันน้อยเกินไปการวัดด้วยค่าสหสัมพันธ์จะผันผวนมาก จนไม่เหมาะสมที่จะใช้วัดในที่นี้เราจึงกำหนดไว้ที่ 6 วัน

ตารางที่ 4.7. ผลการทดลองเปรียบเทียบค่าสหสัมพันธ์แบบระยะทางระหว่าง ARIMA CARIMA-1 และ CARIMA–n เปรียบเทียบโดยระยะเวลาการทำนายจาก 10 วัน จนถึง 6 วัน

Day	ARIMA	CARIMA-1	CARIMA-n
10	0.63	0.60	0.62
9	0.64	0.61	0.63
8	0.65	0.62	0.65
7	0.67	0.63	0.66
6	0.70	0.64	0.68

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.7 จะเห็นว่ายิ่งระยะการทำนายใกล้วันปัจจุบันเท่าไหร่ผลการทำนายยิ่งมีแรงให้มันมีแรงผลักมันออกจากการเป็นอิสระต่อกัน ซึ่งในอนาคตต่อไปเราอาจจะจัดกลุ่มหลักทรัพย์ที่สามารถทำนายได้แปรผันตาม และ กลุ่มของหลักทรัพย์ที่แปรผันตรงข้าม และ ทำการตรงกันข้ามกับมัน ขณะที่ระยะเวลาการทำนายก็ควรใช้ในระยะสั้นเพราะมันจะให้ค่าที่มีพลังมากที่สุด โดยสรุปถ้าเราต้องการการทำนายที่ให้ผลลัพธ์ที่มีทิศทางไม่ว่าจะนำหรือตามการใช้ตัวชี้วัดสหสัมพันธ์แบบระยะทางก็เหมาะสม

4.2.7 อภิปรายผลการเทรดด้วยโมเดล ARIMA, CARIMA-1 และ CARIMA-n

ในส่วนนี้แสดงให้เห็นจำนวนเปอร์เซ็นต์ครั้งที่ทำนายทิศทางไปทิศทางเดียวกันของหลักทรัพย์ต่างๆเป็นเวลา 10 วัน ARIMA ตามสมการที่ (2.17), CARIMA-1 โดยที่ 1 คือ หลักทรัพย์ที่ค่าสหสัมพันธ์ในการ lead สูงสุดจำนวน 1 ตัว ตามสมการที่ (3.2), CARIM-n โดยที่ n คือจำนวนหลักทรัพย์ที่ผ่านเกณฑ์ตามสมการที่ (3.3)

ตารางที่ 4.8 ตารางเปอร์ความถูกต้องของการทำนายจากการซื้อขายตามการทำนาย ด้วย ARIMA, CARIMA-1 และ CARIMA-n จากผลการทำนายเป็นเวลา 10 วัน

Name	ARIMA	CARIMA-1	CARIMA-n
ADVANC	40%	NA	NA
AOT	70%	60%	70%
BANPU	60%	NA	NA
BAY	70%	40%	70%
BBL	60%	60%	40%
BCP	70%	NA	NA
BDMS	80%	60%	70%
BEC	40%	NA	NA
BH	80%	80%	80%
BIGC	40%	30%	40%
BJC	60%	NA	NA
BLA	40%	40%	40%
BTS	70%	40%	50%
CENTEL	50%	NA	NA
CPALL	60%	NA	NA

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

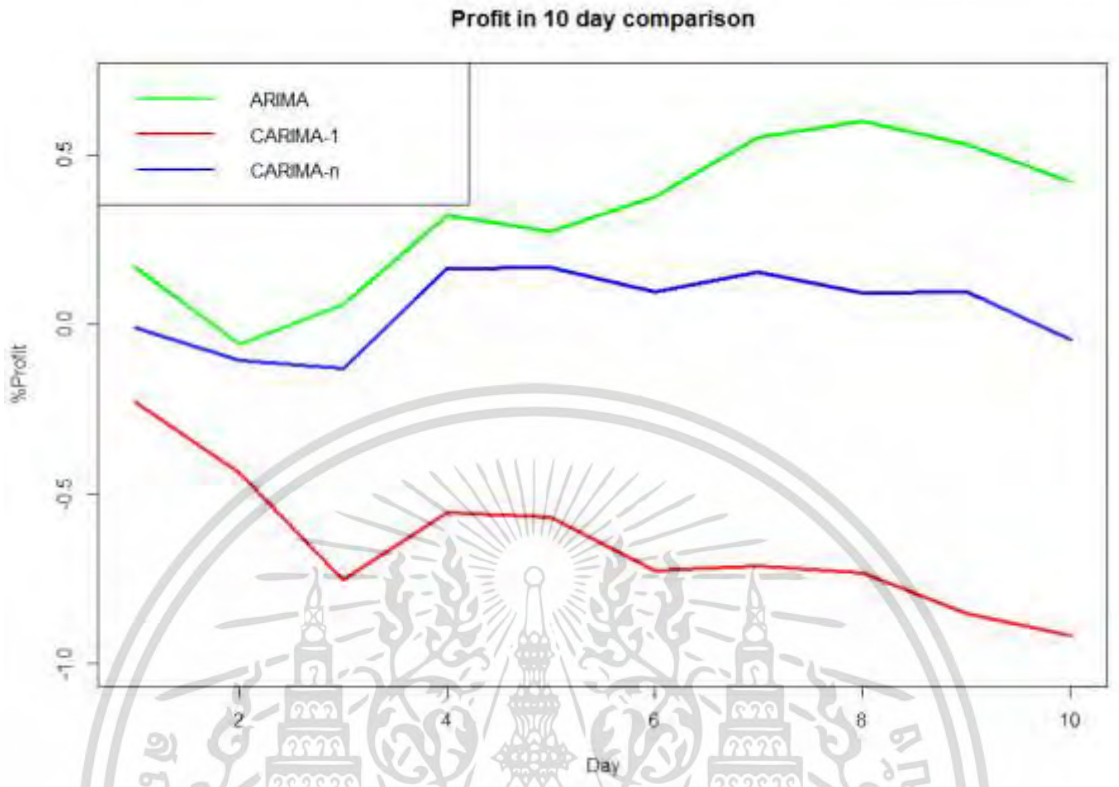
Name	ARIMA	CARIMA-1	CARIMA-n
CPF	80%	NA	NA
CPN	80%	NA	NA
DELTA	40%	NA	NA
DTAC	40%	NA	NA
EGCO	40%	30%	40%
GLOBAL	30%	NA	NA
GLOW	70%	50%	60%
HMPRO	50%	40%	40%
INTUCH	50%	NA	NA
IRPC	60%	NA	NA
IVL	50%	50%	50%
JAS	40%	NA	NA
KBANK	50%	50%	40%
KKP	80%	NA	NA
KTB	40%	30%	30%
LH	50%	NA	NA
M	60%	60%	80%
MINT	50%	50%	50%
PS	30%	20%	30%
PTT	50%	30%	50%
PTTEP	70%	NA	NA
PTTGC	50%	60%	60%
RATCH	60%	30%	50%
ROBINS	80%	NA	NA
SCB	30%	40%	50%
SCC	70%	NA	NA
SCCC	40%	50%	60%
TCAP	90%	NA	NA

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Name	ARIMA	CARIMA-1	CARIMA-n
THCOM	70%	NA	NA
TMB	70%	60%	60%
TOP	50%	50%	50%
TRUE	60%	60%	60%
TTW	50%	70%	70%
TUF	70%	30%	30%
VGI	60%	NA	NA
SUMMARY	57.05%	46.07%	52.32%

จากผลการทดลองในตารางที่ 4.8 เป็นผลการทดลองการทำนายผลลัพธ์ว่าโมเดล ARIMA, CARIMA-1 และ CARIMA-n มาเปรียบเทียบกันทางด้านการเปลี่ยนแปลงขึ้นลง ตามกันมากแค่ไหน โดยผลรวมแล้วแปรพันธ์ตามค่าสหสัมพันธ์เพราะสหสัมพันธ์เป็นการวัดการเคลื่อนไหวตามกันของหลักทรัพย์ ขณะที่ค่าร้อยละความคลาดเคลื่อนบอกได้เพียงแค่ว่าค่าที่ทำนายออกมาห่างกับราคาจริงแค่ไหนโดยไม่ได้บอกทิศทางใดหลักทรัพย์ที่มีค่าร้อยละความคลาดเคลื่อนต่ำอาจจะได้ผลตอบแทนที่ต่ำกว่าหลักทรัพย์ที่ให้ค่าร้อยละความคลาดเคลื่อนสูงได้ โดยโมเดลที่ให้ผลความถูกมากที่สุดก็เป็นโมเดลต้นแบบ ARIMA โดยทำนายถูกไป 57.05% ตามมาด้วยโมเดล CARIMA-n ที่ให้ผลลัพธ์ตามมาที่ 52.32% และ CARIMA-1 ที่ให้ผลความถูกต้องเพียง 46.07% เท่านั้น ต่อจากนี้เราจะลองทำการนำผลการทำนายไปซิมมูเลชั่นคิดผลกำไรจากการซื้อขายต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.11 กราฟเปอร์เซ็นต์กำไรจากการซื้อขายตามการทำนาย ด้วย ARIMA, CARIMA-1 และ CARIMA-n จากการทำนายเป็นเวลา 10 วัน

ในรูปที่ 4.11 เป็นกราฟเปอร์เซ็นต์กำไรจากการซื้อขายตามการทำนายถ้าทำนายว่าหลักทรัพย์นั้นๆจะขึ้นเราจะทำการซื้อ และถ้าเราทำนายว่ามันจะลงเราก็จะขาย ผลที่ได้แสดงให้เห็นว่าผลลัพธ์จากโมเดล ARIMA ให้ผลลัพธ์ดีที่สุดตามค่าสหสัมพันธ์เช่นกัน โดย ARIMA ให้ผลดีที่สุด ตามมาด้วย CARIMA-1 และ CARIMA-n ตามลำดับ จะเห็นได้ว่าแม้โมเดล CARIMA-1 จะทำนายถูกเกิน 50% แต่จำนวนครั้งที่ทำนายถูกกลับให้ผลกำไรน้อยกว่าจำนวนครั้งที่ทำนายผิด และในโมเดล CARIMA-1 ที่หลักทรัพย์นำของมันเป็น Lag 0 บ่อยมาก มันจึงทำให้เกิดการสุ่มสูง ค่ายิ่งทำนายห่างออกไปผลลัพธ์กำไรยิ่งลดลง ขณะที่ CARIMA-n ยังมีการให้น้ำหนักลดการเหวี่ยงจากหลักทรัพย์อื่นที่มาช่วยทำนายตัวเดียว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3. อภิปรายผลการเทรดด้วยโมเดล ARIMA, CARIMA-NF และ CARIMA-F

ในส่วนนี้จะนำเสนอผลการเทรดด้วยโมเดล CARIMA-NF และ CARIMA-F โดยทำการเทรดบนดัชนีหลักทรัพย์ SET50 โดยใช้การช่วยเหลือการทำนายโดยการครอสโอเวอร์ของดัชนีหลักทรัพย์ MSCI USA

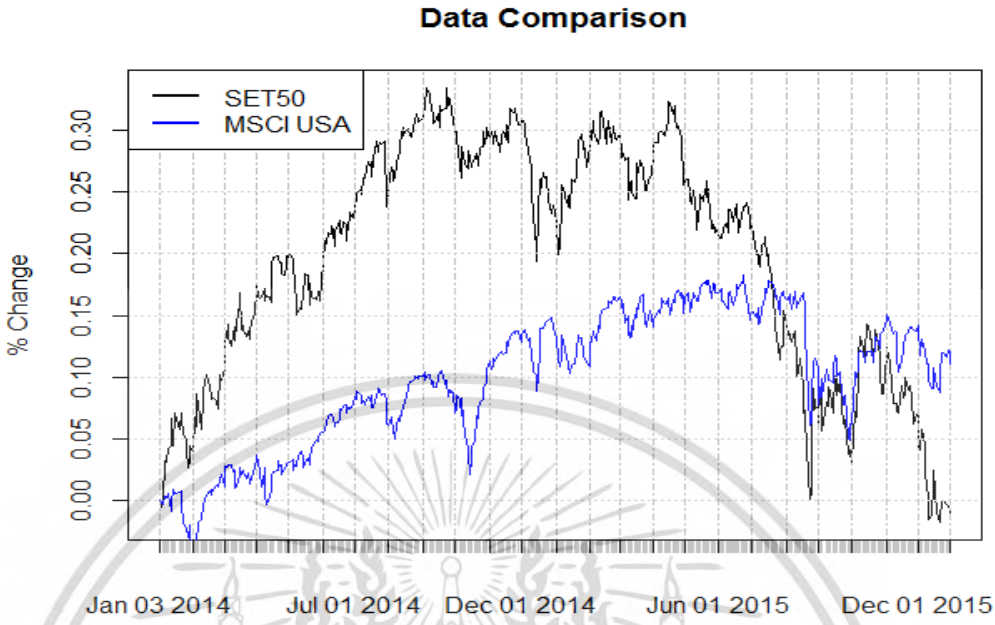
4.3.1. ข้อมูลที่ใช้ในการซิมมูเลชั่นการเทรด

ข้อมูลที่ใช้คือ ข้อมูลดัชนีเป้าหมายในการทำนายคือ ดัชนีหลักทรัพย์ SET50 และดัชนีหลักทรัพย์ MSCI USA ตั้งแต่ วันที่ 1 มกราคม 2014 จนถึง วันที่ 2 ธันวาคม 2015 โดยแบ่งเป็นข้อมูลที่ใช้ในการหาสหสัมพันธ์เริ่มต้น 250 วัน และเป็นข้อมูลทำนาย 250 วัน จากตารางที่ 4.9 คือตัวอย่างข้อมูลดัชนีหลักทรัพย์ SET50 และ ตัวอย่างข้อมูลราคาปิดของ SET50 และ MSCI USA ตั้งแต่วันที่ 1 มกราคม 2014 จนถึง 14 มกราคม 2014 และตัวอย่างรูปกราฟ ROC ของราคาปิดของ SET50 และ MSCI USA เป็นเวลา 2 ปี ดังรูปที่ 4.12 สีดำคือ SET50 และ สีน้ำเงินคือ MSCI USA และรูปค่าสัมประสิทธิ์สหสัมพันธ์ของ SET50 และ MSCI USA เป็นเวลาสองปีในรูปที่ 4.13 โดยที่สีดำคือ ค่าสัมประสิทธิ์สหสัมพันธ์ และ สีเขียวคือเฉลี่ยของค่าสัมประสิทธิ์ของค่าสหสัมพันธ์ 10 วัน

ตารางที่ 4.9 ตัวอย่างข้อมูลดัชนีหลักทรัพย์ SET50

Y-M-D วัน-เดือน-ปี	OPEN ราคาเปิด	HIGH ราคาสูงสุด	LOW ราคา	CLOSE ราคาปิด	VOLUME ปริมาณ
2014-01-01	878.09	878.80	833.29	833.29	76,750,592
2014-01-02	822.27	837.38	816.50	829.89	250,170,624
2014-01-03	829.46	838.87	815.56	837.68	140,455,296
2014-01-04	842.15	861.71	840.12	861.63	56,614,912
2014-01-07	861.40	871.47	854.50	859.35	79,738,624
2014-01-08	864.58	875.56	856.62	860.00	153,884,608
2014-01-09	850.59	861.79	845.18	858.25	355,128,192
2014-01-10	849.36	880.70	849.21	880.70	87,176,512
2014-01-11	874.54	892.46	872.68	887.93	299,030,016
2014-01-14	885.12	889.09	870.00	871.02	45,212,160

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



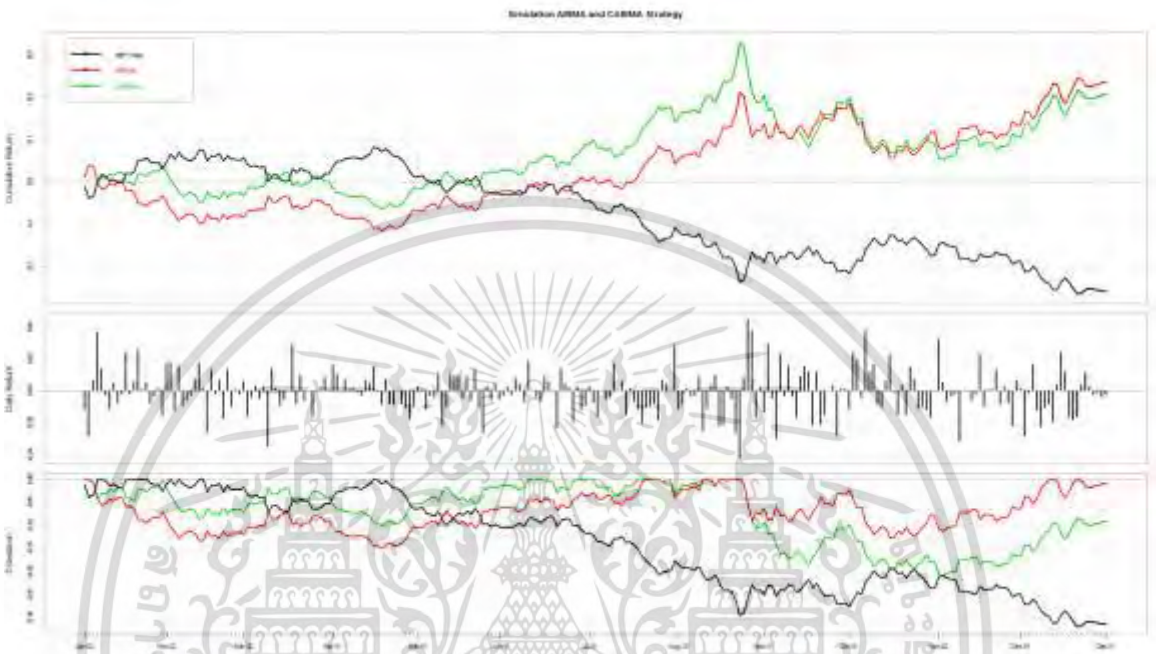
รูปที่ 4.12 ตัวอย่างข้อมูลราคาปิดของ SET50 และ MSCI USA ตั้งแต่ วันที่ 1 มกราคม 2014 จนถึง วันที่ 31 ธันวาคม 2015

รูปที่ 4.13 ตัวอย่างข้อมูลค่าสหสัมพันธ์ของ SET50 และ MSCI USA ตั้งแต่ วันที่ 1 มกราคม 2014 จนถึง วันที่ 31 ธันวาคม 2015

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3.2. อภิปรายผลซิมมูเลชั่นของโมเดล CARIMA-NF

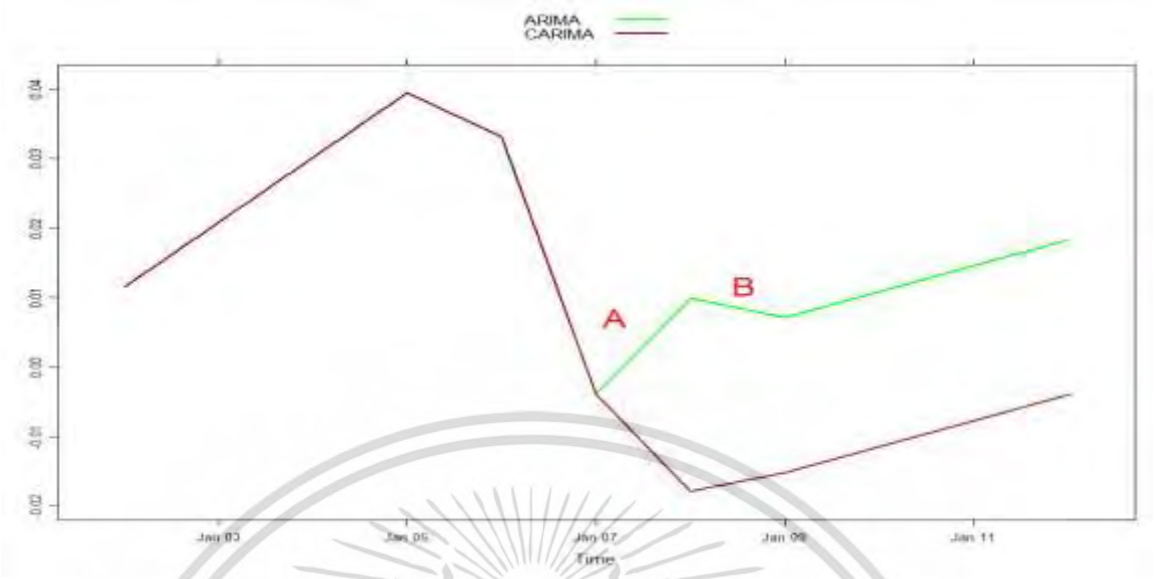
ในส่วนนี้เป็นการอภิปรายผลการทดลองซิมมูเลชั่นการเทรดด้วยโมเดล CARIMA-NF ตามสมการที่ (3.6) โดยทดลองตั้งแต่วันที่ 1 มกราคม จนถึง 31 ธันวาคม 2015 เป็นเวลา 1 ปี (250 วัน)



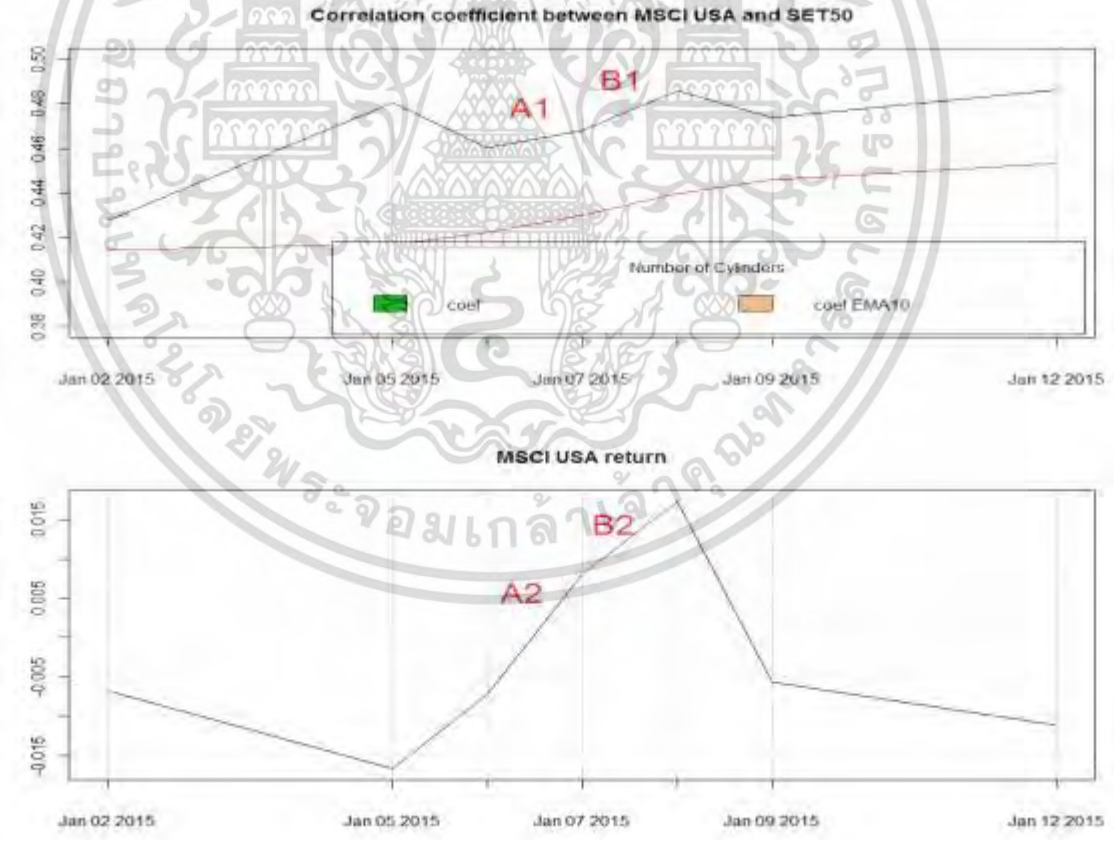
รูปที่ 4.14 ผลเปรียบเทียบซิมมูเลชั่นโมเดล ARIMA และ CARIMA-NF ตั้งแต่ วันที่ 1 มกราคม 2015 จนถึง วันที่ 31 ธันวาคม 2015

จากรูปที่ 4.14 ผลการทดลองจากการซิมมูเลชั่นทำนายหลักทรัพย์ SET50 เป็นเวลา 1 ปี 250 วัน โดยทำการ Forward ทีละ 1 วัน จากนั้นเราจึงทำการซื้อขายหลักทรัพย์ตามราคาที่ทำนาย ผลการทำนายในรูปที่ 4.14 ผลการซิมมูเลชั่นของโมเดล ARIMA สีแดง โมเดล CARIMA-NF สีเขียว และ SET50 สีดำ โดยที่ SET50 มีการปรับตัวลดลง 27.83% ขณะที่โมเดล ARIMA ได้กำไร 23.56% และ โมเดล CARIMA-NF ทำกำไรได้ 21.12% ตามลำดับ จากผลการทำนายเราจะเห็นว่า CARIMA-NF มีช่วงเวลาที่ดีกว่า ARIMA อยู่ช่วง 9 เดือนแรก เราจะดูกันให้ละเอียดตามรูปที่ 4.15 ณ วันที่ 8 มกราคม 2015 ณ จุด A เป็นช่วงเวลาที่การนำของ MSCI USAมีผลทำให้การทำนายของ CARIMA ดีขึ้นเทียบกับ ARIMA แบบธรรมดา เพราะโมเดล CARIMA-NF สามารถใช้ประโยชน์จากการ lead ของ MSCI USA ดังรูปที่ 4.16 ด้านบน A1 คือค่าสหสัมพันธ์ณวันที่ 7 มกราคม 2015 จะเห็นว่า ค่าสหสัมพันธ์อยู่เหนือ ค่าเฉลี่ย ค่า ROC ของ MSCI USA ณ วันที่ 7 มกราคม 2015 จึงถูกนำไปช่วยโมเดล CARIMA ในการทำนายราคาหลักทรัพย์ SET50 ณ วันที่ 8 มกราคม 2015 ให้มีค่าทำนายเปลี่ยนไปเป็นตรงกันข้ามกับ ARIMA ณ วันนี้โมเดล CARIMA จึงสามารถทำกำไรได้ขณะที่โมเดล ARIMA ทำนายค่าผิดจึงขาดทุนไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.15 ผลเปรียบเทียบซิมมูเลชั่นโมเดล ARIMA และ CARIMA-NF วันที่ 1 มกราคม 2015 จนถึง วันที่ 12 มกราคม 2015

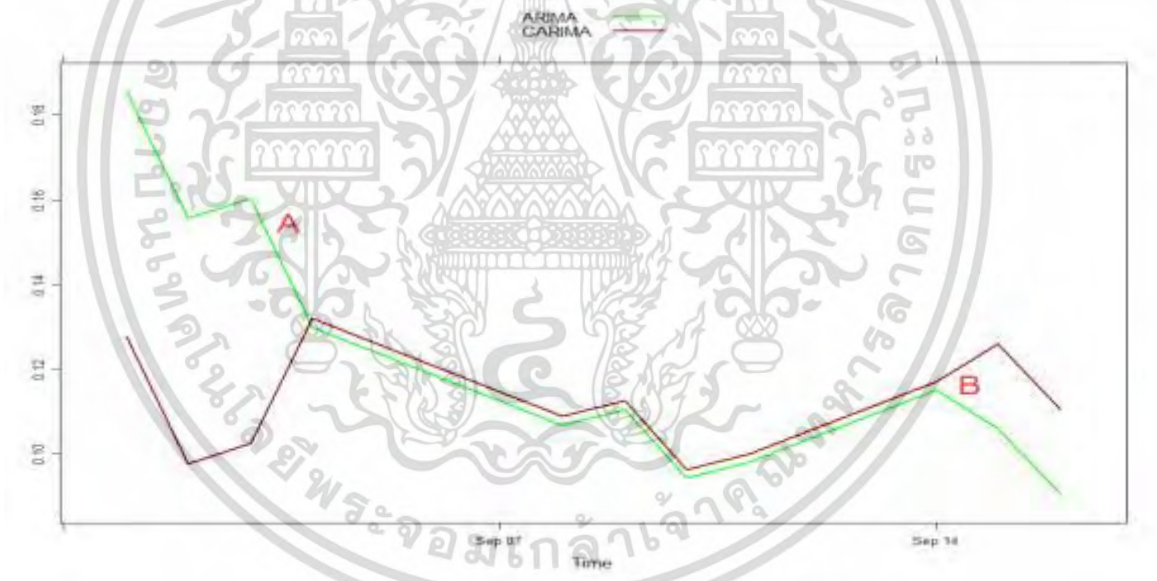


รูปที่ 4.16 ค่าสหสัมพันธ์ของ SET50 และ MSCI USA (บน) และ ราคาเปลี่ยนแปลงของ MSCI USA (ล่าง) ตั้งแต่ วันที่ 1 มกราคม 2015 จนถึง วันที่ 12 มกราคม 2015

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

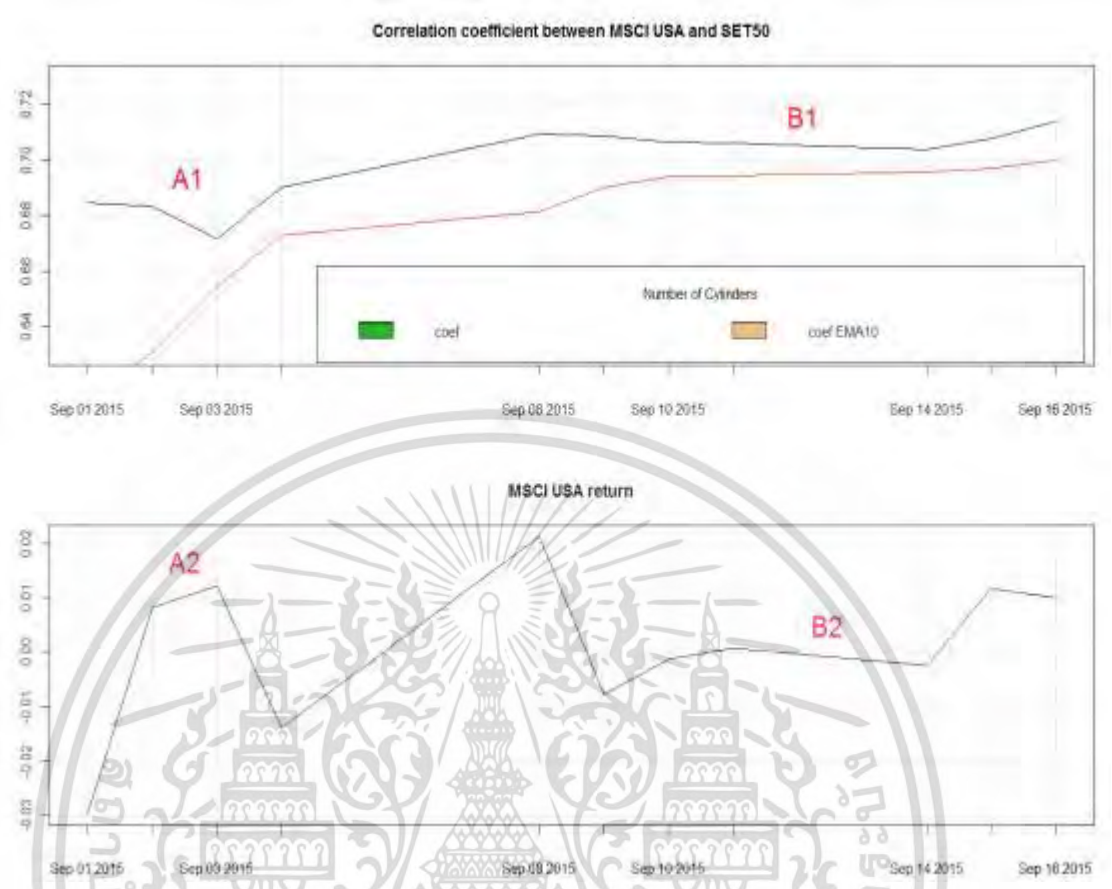
ขณะที่ จุด B ในรูป 4.15 แสดงผลการเทรด ณ วันที่ 9 มกราคม 2015 โมเดล CARIMA เกิดการขนาดทุน ขณะที่ ARIMA ทำกำไรได้ เป็นเหตุเนื่องมาจากค่าสหสัมพันธ์ดังรูป 4.16 จุด B1 คือ วันที่ค่าสหสัมพันธ์ถูกเลือกใช้ เป็นเหตุให้ ค่า ROC ของ MSCI USA ณ วันที่ 8 จึงถึงถูกนำมาทำนายราคาวันที่ 9 มกราคม 2015 แต่เนื่องจากค่าสหสัมพันธ์ของวันนั้นๆวิ่งห่างออกจากกัน จึงทำให้ผลการทำนายผิดพลาดขึ้นมานั่นเอง

ขณะที่เมื่อมาถึงวันที่ 4 กันยายน 2015 และ 15 กันยายน 2015 ณจุด A และ B ตามลำดับ ดังรูป 4.18 ค่าสหสัมพันธ์ ณ วันที่ 3 กันยายน 2015 จุด A1 และ วันที่ 14 กันยายน 2015 จุด B1 เกิดการเคลื่อนตัวออกห่างกันของสหสัมพันธ์อย่างรวดเร็วทำให้ค่า ROC ของ MSCI USA ณ จุด A2 และ B2 ถูกนำมาใช่ช่วยในการทำนายราคาในวันถัดไปในขณะที่ค่าสหสัมพันธ์มันกำลังลดลงนั้นทำให้ผลการทำนายของ CARIMA-NF ไปในทิศทางตรงกันข้ามกับ ARIMA จึงทำให้เกิดความผิดพลาดขึ้นมาก ทำให้โมเดลเกิดการขาดทุนและแพ้ ARIMA ไปในที่สุดเมื่อจบปี



รูปที่ 4.17 ผลเปรียบเทียบซิมมูเลชั่นโมเดล ARIMA และ CARIMA-NF ตั้งแต่วันที่ 1 กันยายน ถึง 16 กันยายน 2015

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.18 ค่าสหสัมพันธ์ของ SET50 และ MSCI USA (บน) และ ราคาเปลี่ยนแปลงของ MSCI USA (ล่าง) ตั้งแต่ วันที่ 1 กันยายน 2015 จนถึง วันที่ 16 กันยายน 2015

จากนั้นเราจะใช้ตัววัดอีกตัวทางโลกของการเงินคืออัตราส่วนชาร์ป โดยได้ผลดังนี้ 1. SET50 ที่ -1.21 ค่าที่ต่ำกว่า 1 คือการลงทุนที่ไม่น่าพอใจ เพราะถึงจะได้กำไรแล้ว ก็ไม่คุ้มค่ากับความเสี่ยงที่วัดด้วยส่วนเบี่ยงเบนมาตราฐานของกำไรในช่วงเวลานั้น 2. ARIMA ที่ 1.09 ค่าเกิน 1 มาแต่ก็ไม่มากนัก เราสรุปได้ว่าถ้าเราลงทุนด้วยวิธีนี้ในระยะเวลา 1 ปีที่ผ่านมาก็ได้ผลเป็นที่พอใช้ 3. CARIMA-NF ที่ 0.96 ผลที่ได้มาไม่ต่างกับ ARIMA มากนัก แต่อย่างไรก็ดี ต่อมาในส่วน MAPE ในส่วนค่าทำนายที่เพิ่มความเหวี่ยงของ CARIMA-NF จะสูงกว่า ARIMA อยู่เล็กน้อยมากๆที่ 1.03% ต่อ 1.02% เนื่องจากเราทำนายเพียงแค่วันเดียวทั้ง 2 โมเดลผลที่ได้มาไม่ต่างกันมากนัก อันดับสุดท้าย Maximum Drawdown คือช่วงที่ ดังสรุปในตารางที่ 4.10

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.10 ตารางสรุปประสิทธภาพของโมเดล ARIMA CARIMA-NF

MODEL	SET50	ARIMA	CARIMA-NF
Profit	-25.83%	23.56%	21.12%
Sharpe Ratio	-1.10	1.09	0.96
Maximum	32%	15%	20%
MAPE	-	1.02%	1.03%
Accuracy	-	53.57%	54.00%

4.3.2. อภิปรายผลซิมมูเลชั่นของโมเดล CARIMA-F

ในส่วนนี้เป็นการอภิปรายผลการทดลองซิมมูเลชั่นการเทรดด้วยโมเดล CARIMA-F ตามสมการที่ (3.7) โดยทดลองตั้งแต่วันที่ 1 มกราคม จนถึง 31 ธันวาคม 2015 เป็นเวลา 1 ปี (250 วัน)

รูปที่ 4.19 ผลเปรียบเทียบซิมมูเลชั่นโมเดล ARIMA และ CARIMA-F ตั้งแต่ วันที่ 1 มกราคม 2015 จนถึง วันที่ 31 ธันวาคม 2015

จากรูปที่ 4.19 ผลการทดลองจากการซิมมูเลชั่นทำนายหลักทรัพย์ SET50 เป็นเวลา 1 ปี 250 วัน โดยทำการ Forward ทีละ 1 วัน จากนั้นเราจึงทำการซื้อขายหลักทรัพย์ตามราคาที่ทำนาย ผลการทำนายในรูปที่ 4.19 ผลการซิมมูเลชั่นของโมเดล ARIMA สีแดง โมเดล CARIMA-F สีเขียว และ SET50 สีดำ โดยที่ SET50 มีการปรับตัวลดลง 27.83% ขณะที่โมเดล ARIMA ได้กำไร 23.56% และ โมเดล CARIMA-F ทำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

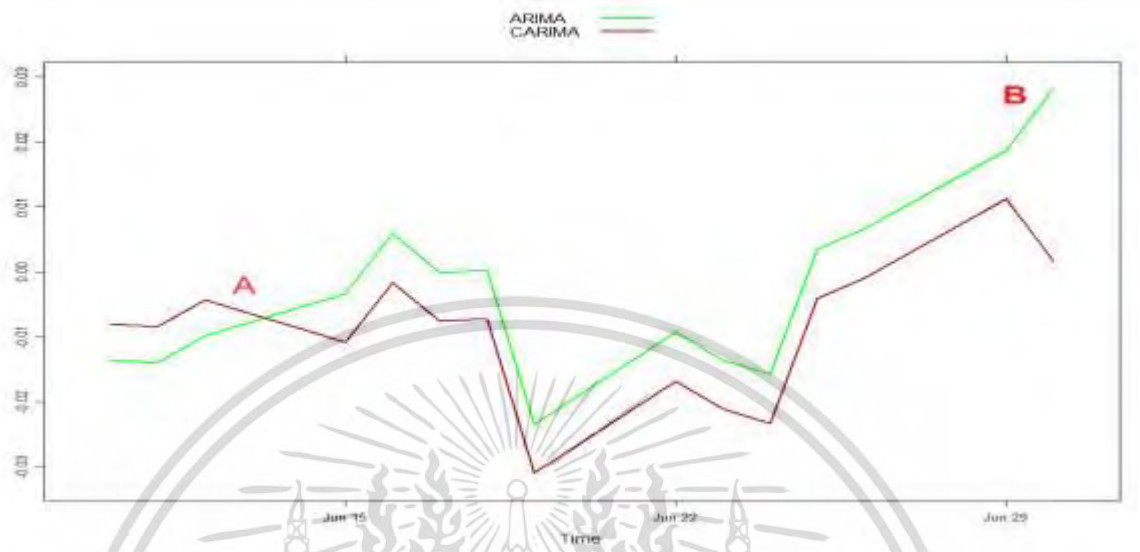
กำไรได้ 25.07% ตามลำดับ จากผลการทำนายเราจะเห็นว่า CARIMA-F เมื่อเราทำการกรองเฉพาะทิศทางการเปลี่ยนแปลงของดัชนี MSCI USA เฉพาะด้านลงแล้วในช่วงแรกโมเดลทั้งสองไม่มีความแตกต่างกันทางด้านการทำนายเลย กล่าวคือการทำนายของ ARIMA ในเวลาที่ค่าเฉลี่ยของสัมประสิทธิ์สหสัมพันธ์ในวันนั้นมากกว่าค่าเฉลี่ยนั้น การเปลี่ยนแปลงของดัชนีนำ MSCI USA มีการเคลื่อนไหวที่สอดคล้องกับการทำนายของ ARIMA หรือ ไม่ก็การเปลี่ยนแปลงของดัชนีตัวนำในวันนั้นไม่มากพอที่จะทำให้โมเดล ARIMA เปลี่ยนทิศทาง จนกระทั่งถึงเดือน กรกฎาคม ไปจนถึง ธันวาคม 2015 ที่ดัชนีตลาดหลักทรัพย์ MSI USA เกิดความผันผวนอย่างหนักในครึ่งปีหลังเกิดการลดลงลงอย่างมากในช่วงครึ่งปีหลัง เนื่องด้วยมีข่าวไม่ดีเกิดขึ้นมากมายในโลกเรา ทั้งสงครามยูเครน ทั้งนี้เป็นเหตุให้การลดลงของดัชนี MSCI USA ในช่วงเวลาครึ่งปีหลัง มีส่วนช่วยในการทำนายของ CARIMA ให้ดียิ่งขึ้น ดังรูปที่ 4.20

รูปที่ 4.20 ผลเปรียบเทียบซิมมูเลชั่นโมเดล ARIMA และ CARIMA-F เดือน 1 กรกฎาคม จนถึง 30 ธันวาคม 2015

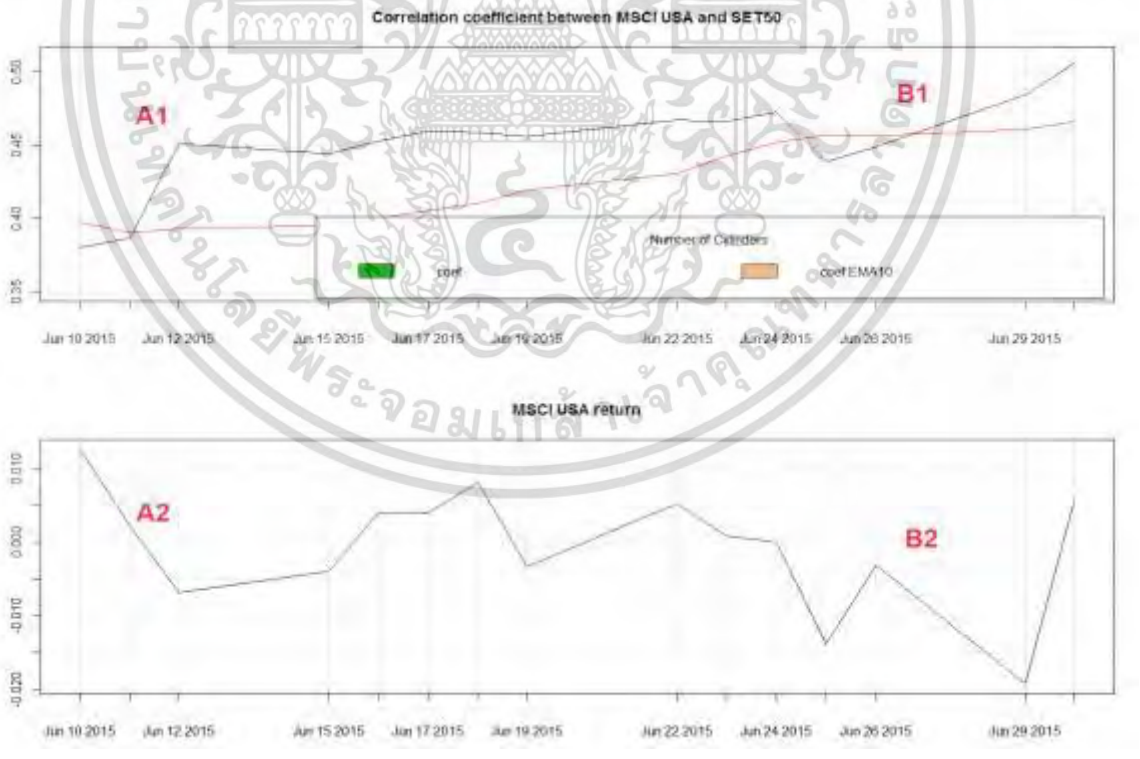
เราจะดูกันให้ละเอียดตามรูปที่ 4.21 ณ วันที่ 15 มิถุนายน 2015 ณ จุด A และ ณ วันที่ 30 มิถุนายน 2015 ณ จุด B เป็นช่วงเวลาที่การนำของ MSCI USA มีผลทำให้การทำนายของ CARIMA ดีขึ้นเทียบกับ ARIMA แบบธรรมดา เพราะ โมเดล CARIMA-NF สามารถใช้ประโยชน์จากการ lead ของ MSCI USA ดังรูปที่ 4.22 ด้านบน จุด A1 คือค่าสหสัมพันธ์ ณ วันที่ 12 มิถุนายน 2015 และ จุด B2 คือค่าสหสัมพันธ์ ณ วันที่ 29 มิถุนายน 2015 จะเห็นว่า ค่าสหสัมพันธ์อยู่เหนือ ค่าเฉลี่ย ค่า ROC ของ MSCI USA ณ วันที่ 12 มิถุนายน 2015 A2 และ 29 มิถุนายน 2015 จึงถูกนำไปช่วยโมเดล CARIMA ในการทำนายราคาหลักทรัพย์ SET50 ณ วันที่ 15 มิถุนายน 2015 (เนื่องจากติดวันหยุดเสาร์อาทิตย์) และ วันที่ 30 มิถุนายน 2015 ตามลำดับในโมเดลนี้เราจะใช้เฉพาะวันที่ดัชนี MACI USA มีค่าเป็นลบเท่านั้น จากนั้นใช้มันไปช่วยในการทำนายให้มีค่าทำนายเปลี่ยนไปเป็นตรงกันข้ามกับ ARIMA ณ วันนี้โมเดล CARIMA จึง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ทำนายได้ถูกต้องกว่า ARIMA ใน 2 วันที่กล่าวมานี้ และวันอื่นๆจนจบปีก็จะมีทั้งผิดและถูกสลับกันไปแต่โดยภาพรวมแล้วการกรองเฉพาะการเปลี่ยนแปลงทางด้านลบของดัชนีให้ผลที่ดีกว่า



รูปที่ 4.21 ผลเปรียบเทียบซิมมูเลชั่นโมเดล ARIMA และ CARIMA-NF ตั้งแต่วันที่ 10 มิถุนายน ถึง 30 มิถุนายน 2015



รูปที่ 4.22 ค่าสหสัมพันธ์ของ SET50 และ MSCI USA (บน) และ ราคาเปลี่ยนแปลงของ MSCI USA (ล่าง) ตั้งแต่วันที่ 10 มิถุนายน ถึง 30 มิถุนายน 2015

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากนั้นเราจะมาวัดอัตราส่วนชาร์ป โมเดล CARIMA-F เพิ่มจากโมเดล CARIMA-NF ที่ 0.95 มาเป็น 1.32 สามารถเอาชนะโมเดลทั้งหมดไปได้ โดยมีอัตราความเบี่ยงเบนมาตราฐานของกำไรลดลง ขณะที่กำไรณ วันสิ้นปีก็มากสุดในทุกโมเดล ต่อมาในส่วน MAPE ในส่วนค่าทำนายที่เพิ่มความเหวี่ยงของ CARIMA-F จะสูงกว่า ARIMA อยู่เล็กน้อยมากที่ 1.03% ต่อ 1.02% และเท่ากับ CARIMA-NF เนื่องจากเราทำนายเพียงแค่วันเดียวทั้ง 2 โมเดลผลที่ได้มาไม่ต่างกันมากนัก อันดับสุดท้าย Maximum Drawdown คือช่วงที่อัตรากำไรลดลงมามากที่สุดดีกว่า CARIMA-NF พอสมควรโดยลดลงมาเหลือ 15% และในช่วงเวลาดังที่เหลือค่า Drawdown ในภาพรวมครึ่งปีหลังของโมเดล CARIMA-F ในรูปที่ 4.11 ทำได้ดีกว่าโมเดลอื่นอีกด้วย ดังตารางสรุปประสิทธิภาพของระบบในตารางที่ 4.11

ตารางที่ 4.11 ตารางสรุปประสิทธภาพของโมเดล ARIMA CARIMA-F

MODEL	SET50	ARIMA	CARIMA-F
Profit	-25.83%	23.56%	25.07%
Sharpe Ratio	-1.10	1.09	1.32
Maximum	32%	15%	15%
MAPE	-	1.02%	1.03%
Accuracy	-	53.57%	54.00%

4.4. สรุปงานวิจัย

ข้อจำกัดหนึ่งของงานวิจัยนี้คือ เนื่องจากข้อมูลที่นำมาใช้ในการทำนายและช่วยในการทำนายนั้นมีจำนวนไม่มากพอและไม่หลากหลายพอ ข้อมูลทั้งหมดมาจาก ตลาดหลักทรัพย์เดียวกันและเป็น หลักทรัพย์ที่มีขนาดใหญ่(มีมูลค่าสูงสุดในประเทศ 50 อันดับแรก) จึงมีการแกว่งของราคาน้อยกว่าหลักทรัพย์ที่มีขนาดเล็กกว่า จึงทำให้เป็นการยากที่จะหลักทรัพย์ที่มีค่าสหสัมพัธสูงพอที่จะนำมาช่วยในการทำนาย อีกทั้งหลักทรัพย์ส่วนใหญ่จะมีการค่าสหสัมพันธ์แบบข้ามได้ค่าสูงสุดหรือต่ำสุดที่ lag 0 นั่นแปลว่าหลักทรัพย์นั้นๆได้มีการเปลี่ยนแปลงทั้งขึ้นและลงในวันเดียวกันเป็นส่วนใหญ่ จึงทำให้ไม่สมเหตุผลนักที่จะใช้มันช่วยในการทำนาย ถ้าต้องการข้อมูลที่สัมบูรณ์ขึ้น ก็จำเป็นจะต้องเสียค่าใช้จ่ายที่ค่อนข้างสูง เนื่องข้อมูลในระดับนี้จะมีการซื้อขายเชิงพาณิชย์เข้ามาเกี่ยวข้อง

การใช้หลักเกณฑ์การเลือกแบบกำหนดว่าค่าสัมประสิทธิ์ของค่าสหสัมพันธ์มากกว่าค่าใดค่าหนึ่งนั้นยังมีจุดด้อยตรงที่ค่าสัมประสิทธิ์ของค่าสหสัมพันธ์นั้นถึงแม้จะสูงแต่มันอาจจะอยู่ในเทรนด์ที่กำลังลดลงมาก็ได้ฉะนั้นมันจึงไม่ได้ผลลัพธ์ที่ดีมากนักในการทำนาย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เนื่องจากข้อมูลหลักทรัพย์นั้นเป็นข้อมูลที่มีความผันผวนสูง ขณะที่ทางด้านโมเดล ARIMA นั้นเป็นโมเดลอนุกรมเวลาที่เหมาะสมกับการใช้อธิบายและทำนายอนุกรมเวลาที่มีความผันผวนน้อยกว่า ผลการทำนายจึงออกมาไม่ดีมากนัก ในเรื่องระยะเวลาการทำนายนั้น ยิ่งทำนายห่างออกไปหลายวันราคาหลักทรัพย์ยิ่งมีความผันผวนมากยิ่งขึ้น การเปลี่ยนแปลงเคลื่อนไหวของราคาหลักทรัพย์มีความไม่คงที่กล่าวคือราคาหลักทรัพย์มีโอกาสที่จะเปลี่ยนแปลงไปในทางทิศทางตรงข้ามกับ 1 วันก่อนหน้าสูง โมเดล ARIMA ไม่สามารถอธิบายได้ทันนัก จึงเป็นผลให้ยิ่งทำนายจำนวนวันห่างออกไปน้อยผลจึงยิ่งดีขึ้น โมเดล ARIMA จึงเหมาะสมกับการทำนายข้อมูลหลักทรัพย์ในระยะสั้นมากกว่าระยะยาว

ส่วนในโมเดล CARIMA แบบทำนายเพียงวันเดียว และที่ทำการซิมมูเลชั่นซื้อขายนั้น ในส่วนของโมเดล CARIMA-NF ที่ใช้วิธีการคำนึงเทรนด์ของค่าสหสัมพันธ์ของดัชนีหลักทรัพย์ นั้นทำงานได้ดีในช่วงเวลาบางช่วงเวลาเท่านั้น เนื่องด้วยการเคลื่อนไหวของทิศทางราคาของสองดัชนีมีการเคลื่อนตัวออกห่างกันในช่วงเวลาอันสั้น ทำให้เทรนด์ของมันแยกห่างออกจากกันอย่างรวดเร็วอีกทั้งการแจกแจงของทิศทางราคาในทิศทางลงมีการเคลื่อนไหวรุนแรงกว่า ดังนั้นการแยกตัวออกจากกันอย่างรวดเร็วจึงทำให้โมเดล CARIMA-NFแพ้ ARIMA ไปในสุด ขณะที่โมเดล CARIMA-F ของเราที่มีกรองทิศทางของสหสัมพันธ์เฉพาะทิศทางลงเท่านั้น ทำให้ชนะโมเดลทุกโมเดลไปได้ในที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
สรุปผลการวิจัยและข้อเสนอแนะ

5.1. สรุปผลการวิจัย

ในงานวิจัยนี้ได้ทำการศึกษาโมเดลการทำนายชื่อ Autoregressive integrated moving average (ARIMA) ในการทำนายหลักทรัพย์ในตลาดหลักทรัพย์ไทย โดยเลือกใช้ข้อมูล SET50 ในการทำนายราคาปิดรายวันเป็นระยะเวลา 10 วันเพื่อทำการเปรียบเทียบผลลัพธ์กับการใช้สหสัมพันธ์แบบข้ามในการช่วยการทำนาย 3 แบบเปรียบเทียบกัน โดยพบว่า การทำนายแบบใช้ตัวช่วยเป็นหลักทรัพย์อื่นหลายตัวเป็นตัวแบบที่ลดความผันผวนได้มากที่สุด ในด้านสหสัมพันธ์ก็ให้ผลดีที่ดีกว่าแบบโมเดลที่ใช้การรวมผลหลักทรัพย์อื่นมาช่วยในการทำนายเพียงตัวเดียว แต่หลักๆของปัญหาการวิจัยนี้คือการทำนายโดยการนำหลักทรัพย์อื่นที่เป็น Lag 0 มาช่วยทำนายนั้นไม่สมเหตุผล ในที่สุดแล้วด้วยข้อกำจัดของปริมาณข้อมูลและหลักเกณฑ์การเลือกจึงทำให้โมเดล ARIMA แบบทั่วไปยังคงให้ผลลัพธ์ที่ดีกว่าอยู่ ซึ่งก็เป็นประเด็นที่จะพัฒนากันต่อไป

ขณะที่โมเดลการทำนายที่เราซิมมูเลชั้น CARIMA-NF และ CARIMA-F ทั้งสองโมเดลนั้นสามารถปิดปัญหาเรื่องการไม่มีตัวนำที่ไม่มีนัยยะสำคัญที่แท้จริงและปัญหาเรื่องเทรนด์ของค่าสหสัมพันธ์ก็หายไปเช่นกัน ขณะที่การแจกแจงทิศทางราคาในทางลงที่เรากรองใน CARIMA-F นั้นให้ผลดีกว่าทุกโมเดลในการทดลองนี้

5.2. ข้อเสนอแนะ

1. การทำนายข้อมูลข้อมูลอนุกรมเวลานั้น ส่งสำคัญอันดับแรกคือในการทำนายข้อมลเชิงอนุกรมเวลา
   นั้นการทำนายให้ได้ค่าร้อยละความคลาดเคลื่อนสัมบูรณ์น้อยนั้นไม่จำเป็นที่มีประโยชน์ สำหรับใน โลกแห่งความจริงมากนัก จากการทดลอง การทำนายระยะเวลายาว 10 วันนั้นแสดงให้เห็นว่าหลาย หลักทรัพย์ที่ให้ค่าค่าร้อยละความคลาดเคลื่อนสัมบูรณ์น้อยสูงเมื่อเรานำมาซิมมูเลชั่นซื้อขายกลได้ผลที่ดีกว่า ดังนั้นการพยายามสร้างระบบเพื่อช่วยในการตัดสินใจซื้อขายควรให้ความสนใจกับทิศทางของราคามากกว่าค่าร้อยละความคลาดเคลื่อนสมบูรณ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. การทำนายด้วยสมการเชิงเส้นอย่างตัวแบบ ARIMA นั้นเหมาะกับการทำนายระยะสั้นมากกว่าระยะในการทำนายหลักทรัพย์เพื่อการซื้อขายควรทำให้ตัดสินใจในระยะเวลาที่สั้นกว่า
3. ควรหาหลักทรัพย์ที่การนำหรือการตามที่มีค่าสูงสุดไม่ใช่ Lag 0 ซึ่งจะสมเหตุสมผลมากกว่าในช่วงเวลาทำนาย

5.3. แนวทางการพัฒนาต่อ

สำหรับแนวทางการพัฒนาต่อในเมื่อเรารู้แล้วว่าโมเดลการทำนายนั้นอาจแบ่งได้เป็นหัวข้อดังนี้

1. ARIMA นั้นเหมาะสมกับการทำนายระยะสั้นจึงสมควรใช้โมเดลนี้ทำนายผลวันต่อวันแบบทดสอบไปข้างหน้า
2. เรื่องหลักทรัพย์อื่นที่นำมาช่วยในการทำนายจะใช้แฟกเตอร์ที่หลากหลายกว่านี้โดยไม่ใช่ข้อมูลที่มจากตลาดแห่งเดียวกัน ทำให้มีการนำและตามที่หลากหลายยิ่งขึ้น
3. การใช้หลักเกณฑ์ในการเลือกใช้หลักทรัพย์อื่นมาช่วยทำนายนั้นควรใช้วิธีที่สามารถสลับได้ตามช่วงเวลา
4. การใช้หลักเกณฑ์การเลือกแบบกำหนดว่าค่าสัมประสิทธิ์ของค่าสหสัมพันธ์มากกว่าค่าใดค่าหนึ่งนั้นยังมีจุดด้อยตรงที่ค่าสัมประสิทธิ์ของค่าสหสัมพันธ์นั้นถึงแม้จะสูงแต่มันอาจจะอยู่ในเทรนด์ที่กำลังลดลงมาก็ได้ ในอนาคตนอกจากจะหาความสัมพันธ์ของหลักทรัพย์มาช่วยในการทำนาย หลักเกณฑ์ในการเลือกควรจะเป็นการกำหนดเทรนด์มากกว่า แม้ค่าสัมประสิทธิ์จะไม่มากแต่ถ้าเราสามารถกำหนดได้ว่ามันกำลังขึ้นก็ควรใช้มากกว่า
5. ในโมเดล CARIMA-NF และ CARIMA-F ต้องพัฒนาหาวิธีสกัดเอาค่าสหสัมพันธ์มาใช้โดยวิธีที่ดีกว่านี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

[1] P. S.P. Cowpertwait and A. W. Metcalfe, **"Introductory Time Series with R"**, Springer, 2009, pp.137 – 155

[2] G.J. Székely, M. L. Rizzo, and N. K. Bakirov, **"Measuring and testing dependence by correlation of distances"**, Annals of Statistics, 2007, pp.2769 – 2794

[3] C. Fonseka, and L. Liyanage, **"A Data mining algorithm to analyse stock market data using lagged correlation"**, 4th International Conference on Information and Automation for Sustainability, 2008, pp.163 – 166

[4] W. Weiqing and Y. Lav, **"A Study of the USDX Based on ARIMA Model A Correlation analysis between the USDX and the Shanghai index".** Consumer Electronics, Communications and Networks (CECNet)

[5] A. Srisawat, **"Discovery Stock Trading Patterns: A Case Study of Thai Stock Market"**, International Journal of Intelligent Information Processing, 2012, Vol. 3 Issue 1, pp.1-9

[6] ช. คันธรส, **"การพยากรณ์ราคาหลักทรัพย์ของบริษัท เจริญโภคภัณฑ์อาหาร จากัด (มหาชน) โดยวิธี ARIMA-GARCH"**, คณะเศรษฐศาสตร์ มหาวิทยาลัยเชียงใหม, 2554

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก

ก.1 งานวิจัยที่ตีพิมพ์

มีงานวิจัยที่ตีพิมพ์ในงานประชุมวิชาการ 1 งาน คือ

Predicting SET50 Stock Prices Using CARIMA(Cross Correlation ARIMA)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

19th International Computer Science and Engineering Conference (ICSEC)
Chiang Mai, Thailand, 23-26 November, 2015

Predicting SET50 Stock Prices Using CARIMA (Cross Correlation ARIMA)

Sornpon Wichaidit
Dept. of Computer Engineering
Faculty of Engineering
King Mongkut's Institute
of Technology Ladkrabang, Thailand
Email: s6601133@kmitl.ac.th

Surin Kittitornkun
Dept. of Computer Engineering
Faculty of Engineering
King Mongkut's Institute
of Technology Ladkrabang, Thailand
Email: kksurin@staff.kmitl.ac.th

Abstract—**Investing in stocks is one of the most popular approaches for money investment. This paper aims to predict short-term stock prices of SET50 of Stock Exchange of Thailand (SET). The proposed method is called CARIMA (Cross Correlation Autoregressive Integrated Moving Average. The basic idea of CARIMA is to find the most highly correlated stock to predict the target one in addition to ARIMA predicted price. The results of CARIMA model yield better price trends (measured by 10-day correlation coefficient) while % MAEs (Mean Absolute Errors) are quite similar with those of ARIMA.**

Keywords - Stock, Prediction, ARIMA, Correlation, Time Series

I. Introduction

Today, Stock Trading is a very popular method of short-term investment. It is known that the time series of stock prices fluctuate greatly but not totally random. Therefore, the investigation of factors that the financial data has long attracted researchers from various different areas, such as: mathematicians, economists and more recently, computer scientists. With the advantages of information technique, a huge amount of stock trading data can be collected easily. The current focus is to design a method of extracting useful information from the collected data. Therefore, data mining has drawn considerable attentions from the young generations of investor in order to predict the changes or discover the patterns of the stock prices.

There are two analytical methods for stock trading: 1.Fundamental [3]; 2.Technical Analysis [4]. Fundamental analysis focuses on the central factors of a company. Such as financial statements, Earnings per Share (EPS), Return on Assets (ROA), Price-Earnings Ratio (P/E Ratio), Return on Equity (ROE), price and book value (P/BV). On the other hand, Technical analyses use the recent historical data of an individual stock, including opening, highest, lowest and closing prices and the volume of stock, to predict future stock price movements. There are many indicators that can be created from the data. Such as Moving Average Convergence Divergence (MACD), Average Directional Index (ADX), Exponential Moving Average (EMA), Double Exponential Moving Average (DEMA), and Relative Strength Index (RSI).

This paper proposes a simple yet effective method for predicting short-term movements of a target stock based on ARIMA Model with another stock with highest lead/lag correlation with the target one. The correlation coefficient can be effectively incorporated with the original ARIMA thus called Cross Correlation ARIMA (CARIMA). The performance of CARIMA is evaluated based on SET (Stock Exchange of Thailand) 50 dataset in terms of trend similarity and %Mean Absolute Error(MAE).

The rest of this paper is organized as follows. Section 2 presents the existing related work. Section 3 explains the theory of ARIMA and the proposed CARIMA method. Section 4 describes dataset and results. Eventually, Section 5 provides the conclusion including the future work.

II. Related Work

There is much work based on time series technique for analyzing and predicting trends of the stock market around the world. All of these researches have been using various factors that have impacts on the stock market as input to predict the trends.

C. Fonseka and L. Liyanage [5] developed an algorithm for predicting an individual stock from the Australian Stock Exchange(ASX) with correlation in 2008. Meanwhile, S. Chaigusin, et al [6], used a feedforward backpropagation neural network to predict the movement of SET index. The inputs of the models consist of seven nodes including the Dow Jones index, Nikkei index, Hang Seng index, Gold prices, Minimum Loan Rate (MLR), and the exchange rates of the Thai Baht and the US dollar. In 2012, A. Srisawat [7] applied an association rule mining technique for discovering relationships between individual stocks from SET. Recently, W. Weiqing and Y. Lav [8] used ARIMA Model to study volatile characteristics in the US dollar index itself in 2013. They used measurement statistical model to fit the ARIMA Model to predict US Dollar index movement for one month.

As mentioned above, all of them used various data mining techniques to predict the stock or index. Very few of them have used relationships between stocks to help in the stock price prediction process. As a result, we are going to take advantages of this gap to propose our method.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

19th International Computer Science and Engineering Conference (ICSEC)
Chiang Mai, Thailand, 23-26 November, 2015


```
Augmented Dickey-Fuller Test

data:  stock[, "CLOSE"]
Dickey-Fuller = -2.2875, Lag order = 9, p-value = 0.4566
alternative hypothesis: stationary
```

Fig. 1. Unit Root Test of ADVANC

III. Our Proposed Method

The method used to develop ARIMA model for stock price prediction with combining correlation of the other stock (the stock that has lead/lag correlation to the focused stock) can be explained in subsections below

A. Unit Root Test

We make use of the Augmented DickeyFuller test(ADF) as the unit root test for the dataset whether it is stationary or not. So we test it on the closing price of an individual stock. The result of this test shown in Fig. 1 that the data is stationary.

B. ARIMA model

Autoregressive integrated moving average (ARIMA) model is generalisation of an autoregressive moving average (ARMA) model. ARIMA include differencing I to ARMA AR(autoregressive)+I(integrated)+MA(moving average). The equation of ARIMA Model is.

$$X_t = m + \sum_{t=1}^{p} \Theta_{i-p} \Delta_d X'_{t-i} - \sum_{t=1}^{q} \ominus_{i-q} \epsilon_{t-i} \qquad (1)$$

Where m is a constant, Θ and $\ominus$ are parameter of autoregressive and moving average parts, ϵ is error(white noise), i is lead/lag time variable (days), Δ is the difference, X' is a closing price of an individual stock, t is a day variable. p, d and q are orders of autoregressive, difference and moving average parts, respectively.

The ARIMA predicted stock price X_t is parameterized as

$$X_t = ARIMA(p, d, q) \qquad (2)$$

where AR: p is order of the autoregressive part, I: d is the degree of first differencing involved, and MA: q is order of the moving average part.

To select the best model of ARIMA(p,q,d), Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) can be evaluated to estimate the quality as listed in Table 1. In this case, ARIMA(0,1,0) model is selected to predict ADVANC because of the smallest values of AIC and BIC. The model returns the smallest AIC of 4582.58 and relatively small BIC of 4587.26. in Fig. 2 show that there is no correlation between predicted ADVANC and the actual price movement of ADVANC.

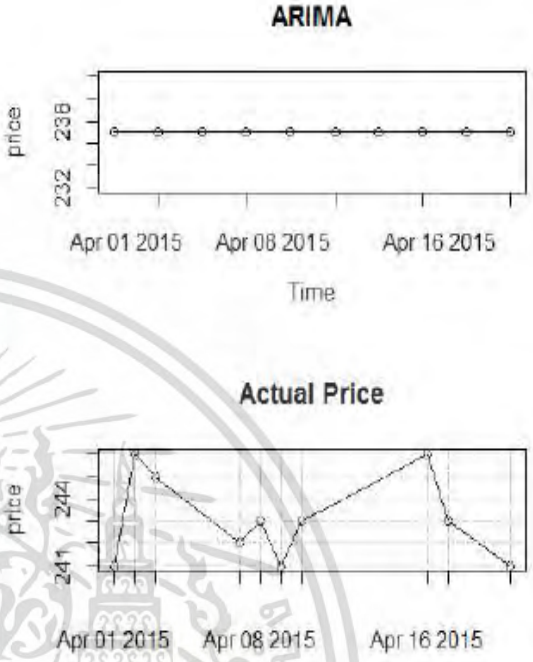


Fig. 2. ADVANC: ARIMA Predicted Price, X_t (Top) vs Actual Price, X'_t (Bottom)

TABLE I
ARIMA Model Selection of ADVANC based on Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC)

ARIMA	AIC	SC
p=1,d=0,q=0	4594.27	4608.31
p=0,d=1,q=0	4582.58	4587.26
p=0,d=0,q=1	6846.23	6860.27
p=1,d=1,q=0	4584.58	4593.94
p=1,d=0,q=1	4596.27	4614.99
p=0,d=1,q=1	4584.58	4593.94
p=1,d=1,q=1	4586.58	4600.62

C. Cross Correlation with Lead/Lag i days

We assume the closing price is the one that shows the realest price movement behaviour rather than High Low Open price. Cross Correlation analysis of closing prices is a measure of the interrelationship between two stocks prices with a function of lead/lag time of one company relative to another. The correlation coefficient can be calculated with respect to lead/lag time i days as defined in Eq.(3).

$$\rho_{x_t y_t}(i) = \sum_{t=1}^{796} [(X_t - u_X)(Y_{t-i} - u_Y)]/\delta_X \delta_Y \qquad (3)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

19th International Computer Science and Engineering Conference (ICSEC)
Chiang Mai, Thailand, 23-26 November, 2015

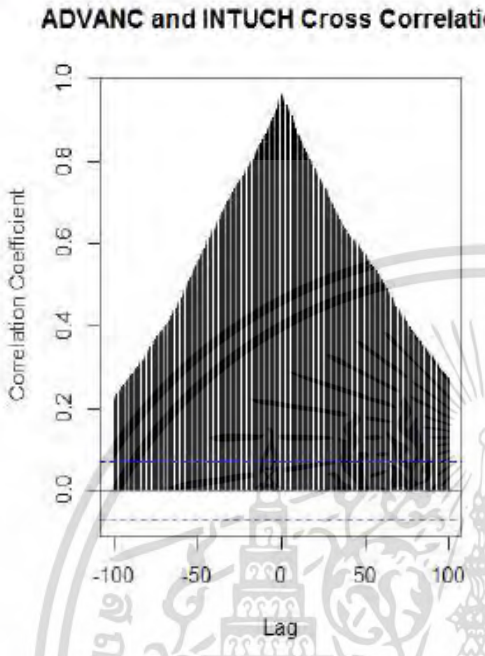


Fig. 3. Cross Correlation Coefficient $\rho_{X_t Y_t}(i)$ with Lead/Lag i between BAY and BH

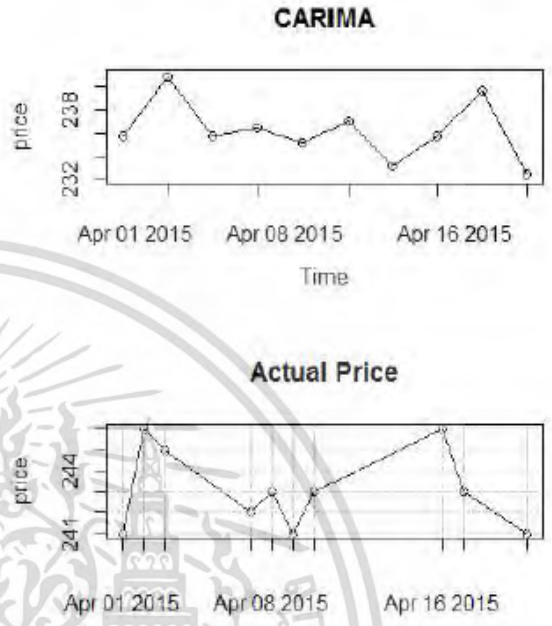


Fig. 4. ADVANC CARIMA Predicted Price $\hat{X}_t$ with Lead, $i = 10$ from INTUCH (Top) vs Actual Price, X'_t (Bottom)

where t is the day variable that starts from January 1st, 2012 to March 31st, 2015, i is the lead/lag time variable in days, δ_X is a standard deviation of stock X, δ_Y is a standard deviation of stock Y, u_X is a mean of stock X and u_Y is a mean of stock Y.

We choose the highly correlated coefficient whose $|\rho_{X_t Y_t}(i)| \geq 0.8$ at time lag $t - i$. It means that stock Y_t is leading stock X_t by i days. As a result, stock Y_{t-i} can improve ARIMA Model to predict stock X_t price movement. The example of cross correlation between ADVANC and INTUCH shown in Fig. 3.

D. CARIMA: Cross Correlation ARIMA

After we get the results of ARIMA price X_t predicted up to 10 days from April 1-18, 2015. The highly correlated stock Y whose $|\rho_{X_t Y_t}(i)| \geq 0.8$ that leads stock X for i days is picked. By shifting back stock Y to day $t - i$, the Rate of Change (ROC) from Y_{t-i} is applied with respect to X_t. Finally, the CARIMA predicted $\hat{X}_t$ is a product of $\rho_{X_t Y_t}(i)$, X_t, and ROC of Y_{t-i} as shown in Eq.(4).

$$\hat{X}_t = \rho_{X_t Y_t}(i) \times X_t \times (1 + \frac{Y_{t-i+1} - Y_{t-i}}{Y_{t-i}}) \quad (4)$$

where $\rho_{X_t Y_t}(i)$ is the correlation coefficient between stock x and y, X is the ARIMA predicted price, Y is the most correlated stock, t is day variable and i is the lead time variable in days.

IV. Experiment and Results

In this section, there are 2 main activities. Firstly, we explain the details of SET50 dataset. Secondly, we present and discuss the results.

A. SET50 Dataset

The SET50 dataset is collected from Stock Exchange Thailand (SET). It consists of the 50 most valuable companies (stocks) in Thailand. We use their closing prices to find the correlation coefficient of each individual stock to the others with the lead and lag time. The dataset starts from January 1st, 2012 ($t = 1$) to March 31, 2015 ($t = 796$).

B. Results and Discussion

The stock names, 10-day correlation coefficients, and %MAEs between CARIMA $\hat{X}_t$ and actual prices X'_t vs. ARIMA X_t and actual prices X'_t are listed in Table 2. NA in $\rho_{X_t X'_t}$ column means Not Available because the ARIMA result is a straight line resulting divided by zero. The price trends of CARIMA are closer to the actual prices than those of ARIMA alone (measured by 10-day correlation coefficients), although the %MAEs (Mean Absolute Errors) are quite similar. %MAEs of CARIMA is not better than %MAEs of ARIMA.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

TABLE II
10-DAY CORRELATION COEFFICIENTS AND %MAE OF CARIMA, $\hat{X}_t$ AND ACTUAL STOCK PRICES, X'_t VS. ARIMA, X_t AND ACTUAL STOCK PRICES, X'_t. (NA : NOT AVAILABLE)

Stock	$\rho_{\hat{X}_t X'_t}$	$\rho_{X_t X'_t}$	$\%MAE_{\hat{X}_t X'_t}$	$\%MAE_{X_t X'_t}$
JAS	0.58	NA	2.29	2.13
ADVANC	0.40	NA	2.85	2.51
RATCH	0.35	NA	0.84	0.51
GLOBAL	0.33	-0.39	1.53	1.32
IVL	0.31	NA	2.07	2.01
THCOM	0.30	NA	3.96	2.23
BANPU	0.27	NA	3.65	3.20
TCAP	0.15	NA	1.86	1.46
BCP	0.14	NA	4.61	4.78
EGCO	0.10	-0.87	1.53	1.51
KKP	0.08	-0.15	0.89	0.74
PTTEP	0.07	-0.56	6.82	6.36
CENTEL	0.04	NA	3.94	4.10
TTW	0.03	-0.88	1.38	1.41
BAY	-0.09	0.32	4.62	5.17
BH	-0.11	-0.75	3.67	1.74
GLOW	-0.17	-0.74	2.64	2.46
TRUE	-0.20	0.80	3.84	4.07
BDMS	-0.22	-0.39	2.62	2.42
CPN	-0.33	NA	3.20	3.61
SCB	-0.38	-0.51	1.77	1.33
PS	-0.48	-0.71	2.62	2.31

Because the main purpose of ARIMA is to predict price movement as close to the actual price as possible. But CARIMA incorporates the cross correlation coefficient with ARIMA in order to improve the correlation of the price movements. For example, CARIMA predicted prices of ADVANC predicted by INTUCH can be plotted in Fig. 4 and compared with those of ARIMA in Fig. 2. We can see that those of CARIMA are more correlated while the predicted price from ARIMA is a straight line thus not correlated to the actual price. Although %MAE are slightly higher than ARIMA. CARIMA can yield better price trends as $\rho_{\hat{X}_t X'_t}$ are better than $\rho_{X_t X'_t}$.

Since the dataset is not big enough, only 50 individual stocks, it is hard to find the leading/lagging stocks with high correlation coefficients to make the better prediction. Moreover, most of the stocks have the highest/ least correlation coefficients in lag i = 0 that can be predicable that stock always move up and down together in the same day. Its not reasonable to use that to help ARIMA to predict individual stock price movement.

V. CONCLUSION

This paper aims to incorporate cross correlation coefficient with ARIMA to predict short-term (daily) SET50 stock price movement. The proposed method is called CARIMA to predict stock A using the most highly correlated stock B within 10-day lead. The empirical results obtained and compared CARIMA predicted prices $\hat{X}_t$ and ARIMA predicted X_t against the actual prices X'_t. In terms of performance evaluation, CARIMA prices are more correlated to the actual prices and their %MAEs are similar to those of ARIMA. For future work, we intend to apply two or more stocks as predictors to CARIMA for each individual SET50 stock. It is expected that they can improve the prediction performance even more.

REFERENCES

[1] P. S.P. Cowpertwait and A. W. Metcalfe, *Introductory Time Series with R*, Springer, 2009, pp.137-155
[2] C. M.Conover and D. R. Peterson, *The Lead-Lag Relationship between the Option and Stock Markets prior to Substantial Earnings Surprises and the Effect of Securities Regulation*, Journal of Financial and Strategic Decisions, Spring 1999, Vol. 12 No. 1
[3] B. Graham , J. Zweig and W. Premchaiswadi, *The Intelligent Investor: The Definitive Book on Value Investing. A Book of Practical Counsel*, Collins Business Essentials, 2006
[4] S. Nison, *Beyond Candlesticks: New Japanese Charting Techniques Revealed*, Wiley Finance, 2009
[5] C. Fonseka, and L. Liyanage, *A Data mining algorithm to analyse stock market data using lagged correlation*, 4th International Conference on Information and Automation for Sustainability, 2008, pp.163 - 166
[6] S. Chaigusin, C. Chirathamjaree, and J. Clayden, *A Data mining algorithm to analyse stock market data using lagged correlation*, Proceedings of the International Conference on Computational Intelligence for Modelling Control and Automation, 2008, pp. 670 - 673
[7] A. Srisawat, *Discovery Stock Trading Patterns: A Case Study of Thai Stock Market*, International Journal of Intelligent Information Processing, 2012, Vol. 3 Issue 1, pp.1-9
[8] W. Weiqing and Y. Lav, *A Study of the USDX Based on ARIMA Model A Correlation analysis between the USDX and the Shanghai index*. Consumer Electronics, Communications and Networks (CECNet), 3rd, 2013
[9] J. A. Ryan, J. M. Ulrich, and W. Thielen, *quantmod: Quantitative Financial Modelling Framework*, Version: 0.4-5, https://cran.r-project.org/web/packages/quantmod/index.html, 2015
[10] R.J. Hyndman, *forecast: Forecasting Functions for Time Series and Linear Models*, Version: 6.1, https://cran.r-project.org/web/packages/forecast/index.html, 2015

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อ-นามสกุล นายศรพล วิชัยดิษฐ

วัน เดือน ปีเกิด 23 เมษายน 2533

ที่อยู่ 99/635 หมู่บ้านชดลดาสุวรรณภูมิ ตำบล ศรีษะจรเข้น้อย
อำเภอบางเสาธง จังหวัดสมุทรปราการ 10540 โทร.080-5354927

ประวัติการศึกษา 2556 วิทยาการบัณฑิต สาขาวิทยาการเทคโนโลยีสารสนเทศ (เกียรตินิยมอันดับ1)มหาวิทยาลัยเทคโนโลยีมหานคร

ความชำนาญเฉพาะด้าน 1.) ระบบซื้อขายหลักทรัพย์เชิงปริมาณ(Quantitative Trading System)
2.) พัฒนาโปรแกรมภาษา R, Python, Java, C++, C#, Matlab, PHP

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้