

การพยากรณ์ยอดขายรถยนต์ที่มีฤดูกาลและมีสถานการณ์พิเศษ

CAR SALES FORECASTING WITH SEASONS AND EVENTS

วิชญา รัตนเมธาวิ

WITCHAYA RATTANAMETAWEE

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร  
ปริญญาปรัชญาดุษฎีบัณฑิต สาขาวิชาคณิตศาสตร์ประยุกต์  
ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์  
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2562

KMITL-2019-SC-D-001-004

การพยากรณ์ยอดขายรถยนต์ที่มีฤดูกาลและมีสถานการณ์พิเศษ

CAR SALES FORECASTING WITH SEASONS AND EVENTS

วิชญา รัตนเมธาวิ

WITCHAYA RATTANAMETAWE

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร  
ปริญญาปรัชญาดุษฎีบัณฑิต สาขาวิชาคณิตศาสตร์ประยุกต์  
ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์  
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. ๒๕๖๒

KMITL-2019-SC-D-001-004

# CAR SALES FORECASTING WITH SEASONS AND EVENTS

WITCHAYA RATTANAMETAWEE

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE  
DEGREE OF DOCTOR OF PHILOSOPHY IN APPLIED MATHEMATICS  
DEPARTMENT OF MATHEMATICS FACULTY OF SCIENCE  
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG  
2019

KMITL-2019-SC-D-001-004

COPYRIGHT 2019

FACULTY OF SCIENCE

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

|                             |  |
|-----------------------------|--|
| หัวข้อวิทยานิพนธ์           | การพยากรณ์ยอดขายรถยนต์ที่มีฤดูกาลและมีสถานการณ์พิเศษ |
| นักศึกษา                    | นางสาววิชญา รัตนเมธาวี                               |
| รหัสประจำตัว                | 60605092   |
| ปริญญา                      | ปรัชญาดุษฎีบัณฑิต                                    |
| สาขาวิชา                    | คณิตศาสตร์ประยุกต์                                   |
| ภาควิชา                     | คณิตศาสตร์   |
| พ.ศ.                        | 2562   |
| อาจารย์ที่ปรึกษาวิทยานิพนธ์ | รศ.ดร. ฉัฐไชย์ ลีนาวงศ์                              |

### บทคัดย่อ

ในงานวิจัยนี้นำเสนอตัวแบบเพื่อพยากรณ์ยอดขายรถยนต์ในประเทศไทย โดยใช้ตัวแบบการพยากรณ์สองแบบ คือ แบบที่ 1 ตัวแบบการพยากรณ์เชิงอนุกรมเวลาที่มีการแยกส่วนประกอบและแบบที่ 2 ตัวแบบการพยากรณ์การถดถอยเชิงเส้นแบบพหุ ซึ่งตัวแบบทั้งสอง จะพิจารณาเพิ่มเติมในส่วนของผลกระทบจากฤดูกาลและสถานการณ์พิเศษ โดยใช้ข้อมูลยอดขายรถยนต์นั่งขนาดเล็กในประเทศไทย ตั้งแต่ปี พ.ศ. 2548-2561 ซึ่งเป็นช่วงเวลาที่รวมสถานการณ์พิเศษที่ประเทศไทยประสบถึง 2 สถานการณ์ หนึ่งคือ อุทกภัยครั้งใหญ่ในปี 2554 และสอง โครงการคืนภาษีรถยนต์คันแรกของรัฐบาลในปี 2554-2555 เพื่อกระตุ้นอุตสาหกรรมยานยนต์ ทั้งนี้ สำหรับตัวแบบการพยากรณ์การถดถอยเชิงเส้นแบบพหุ จะใช้ผลิตภัณฑ์มวลรวมในประเทศ (GDP) ของประเทศไทย และอัตราดอกเบี้ยเงินกู้ระยะยาว 12 เดือน เป็นตัวแปรอธิบาย โดยผลลัพธ์ทางสถิติที่ได้จากตัวแบบทั้งสองแบบที่พิจารณาผลกระทบจากฤดูกาลและสถานการณ์พิเศษนี้ เมื่อนำไปเปรียบเทียบกับตัวแบบที่ไม่มีฤดูกาลและ/หรือสถานการณ์พิเศษ พบว่า ตัวแบบที่นำเสนอ ให้ค่าเฉลี่ยของค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน (MAPE) ต่ำกว่าอย่างมาก ยิ่งไปกว่านั้น ตัวแบบการถดถอยเชิงเส้นที่เพิ่มปัจจัยฤดูกาลและสถานการณ์พิเศษนี้ ยังให้ค่าสัมประสิทธิ์การกำหนดค่าที่ปรับปรุง (Adjusted R<sup>2</sup>) ที่สูงกว่าเช่นกัน

**คำสำคัญ :** การพยากรณ์เชิงอนุกรมเวลา, วิธีการแยกส่วนประกอบ, การพยากรณ์สถานการณ์พิเศษ, การถดถอยเชิงพหุ, ยอดขายรถยนต์ในประเทศไทย

|                |   |
|----------------|---|
| Thesis Title   | Car Sales Forecasting with Seasons and Events |
| Student        | Witchaya Rattanametawee                       |
| Student ID     | 60605092                                      |
| Degree         | Doctor of Philosophy                          |
| Program        | Applied Mathematics                           |
| Department     | Mathematics                                   |
| Year           | 2019  |
| Thesis Advisor | Assoc. Prof. Dr. Chartchai Leenawong          |

## ABSTRACT

In this dissertation, two types of forecasting models are proposed for Thailand's car sales data. One is a time series decomposition and the other is the causal multiple linear regression. Both types of the proposed models incorporate the effects of seasonality as well as special events. The data used is the compact car sales in Thailand from the years 2005-2018 covering two huge-impact incidents, namely, the 2011 nationwide big flood and the 2011-2012 government's initiative tax-incentive program for boosting the automobile industry. For the case of causal regression, Thailand's Gross Domestic Products (GDP) and the 12-month Loan's Interest rate are used as explanatory variables. Statistical results obtained from the proposed models with seasons and special events, compared to those from the models with neither seasons nor both yield better accuracy according to Mean Absolute Percentage Error (MAPE). Furthermore, the causal regression models also give higher the Adjusted Coefficient of Determination (Adjusted  $R^2$ ).

**Keywords:** Time Series Forecasting, Decomposition Method, Event Forecasting, Multiple Regression, Thailand's car sales

## กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จได้อย่างดีด้วยคำแนะนำและคำปรึกษาจาก รศ.ดร.ฉัฐไชย์ สีนาวงศ์ อาจารย์ผู้ควบคุมวิทยานิพนธ์ และ รศ.ดร.พรฤดี เนติโสภาคกุล ที่ให้คำชี้แนะ ให้กำลังใจตลอดการเรียนและการทำวิทยานิพนธ์ ข้าพเจ้ารู้สึกซาบซึ้งในความอนุเคราะห์จากท่านอาจารย์ และขอกราบขอบพระคุณเป็นอย่างสูงยิ่ง

ขอขอบพระคุณ รศ.ดร.สุวรรณ ถังมณี ผศ.ดร.กาญจนา คำนึ่งกิจ ผศ.ดร.ใจปอง เกษมสุวรรณ และ ดร.วรรณพร สรรประเสริฐ คณะกรรมการที่กรุณาให้คำชี้แนะจุดบกพร่องและแก้ไขข้อผิดพลาดต่าง ๆ อย่างดียิ่ง

ขอขอบพระคุณ ศ.ดร.สุเทพ สนวนใต้ ที่กรุณาอนุเคราะห์ให้คำแนะนำต่าง ๆ ตลอดจนคณาจารย์ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ทุก ๆ ท่าน ที่เป็นกำลังใจและให้คำแนะนำตลอดมา

ขอขอบคุณพี่ ๆ เพื่อน ๆ และน้อง ๆ ในภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ที่เป็นกำลังใจให้เสมอ

สุดท้ายนี้ข้าพเจ้าขอกราบขอบพระคุณบิดามารดาที่เป็นกำลังใจ ให้การสนับสนุนในทุก ๆ ด้าน และขอขอบคุณคณะกรรมการการอุดมศึกษาที่สนับสนุนงบประมาณในการศึกษาและการวิจัย ทำให้ข้าพเจ้าสามารถทำวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงด้วยดี

วิชญา รัตนเมธาวิ

# สารบัญ

|   | หน้า     |
|---|----------|
| บทคัดย่อภาษาไทย .....                                       | ก        |
| บทคัดย่อภาษาอังกฤษ .....                                    | ข        |
| กิตติกรรมประกาศ .....                                       | ค        |
| สารบัญ .....  | ง        |
| สารบัญตาราง .....   | ฉ        |
| สารบัญรูป .....   | ช        |
| <br>  |          |
| <b>บทที่ 1 บทนำ .....</b>                                   | <b>1</b> |
| 1.1 อุตสาหกรรมยานยนต์ในประเทศไทย .....                      | 1        |
| 1.2 วัตถุประสงค์ของงานวิจัย .....                           | 2        |
| 1.3 ขอบเขตของงานวิจัย .....                                 | 2        |
| 1.4 ขั้นตอนการวิจัย .....                                   | 3        |
| 1.5 ผลที่คาดว่าจะได้รับ .....                               | 3        |
| <br>  |          |
| <b>บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง .....</b>          | <b>4</b> |
| 2.1 ทฤษฎีที่ใช้ในการวิเคราะห์ .....                         | 4        |
| 2.1.1 การพยากรณ์และวิธีการพยากรณ์ .....                     | 4        |
| 2.1.2 การเลือกเทคนิคในการพยากรณ์ .....                      | 4        |
| 2.1.3 กระบวนการพยากรณ์ .....                                | 6        |
| 2.1.4 การวิเคราะห์อนุกรมเวลา .....                          | 7        |
| 2.2 การวิเคราะห์แนวโน้มและความแปรผันตามฤดูกาล .....         | 10       |
| 2.2.1 การวิเคราะห์แนวโน้ม .....                             | 10       |
| 2.2.2 การวิเคราะห์ความแปรผันตามฤดูกาล .....                 | 11       |
| 2.3 ตัวแบบอนุกรมเวลา .....                                  | 12       |
| 2.3.1 ตัวแบบการบวก .....                                    | 12       |
| 2.3.2 ตัวแบบการคูณ .....                                    | 12       |
| 2.4 การวิเคราะห์การถดถอยเชิงเส้น .....                      | 13       |
| 2.4.1 การถดถอยเชิงเส้นอย่างง่าย .....                       | 13       |
| 2.4.2 การถดถอยเชิงเส้นแบบพหุ .....                          | 15       |
| 2.4.3 การวิเคราะห์การถดถอยสำหรับตัวแปรเชิงคุณภาพ .....      | 18       |
| 2.5 การพิจารณาความสัมพันธ์และการเปรียบเทียบค่าพยากรณ์ ..... | 22       |
| 2.5.1 การวิเคราะห์สหสัมพันธ์ .....                          | 22       |
| 2.5.2 การประเมินค่าความผิดพลาดจากการพยากรณ์ .....           | 24       |
| 2.5.3 สัมประสิทธิ์การตัดสินใจ .....                         | 24       |
| 2.6 ทบทวนวรรณกรรม .....                                     | 25       |

|  |    |
|--|----|
| บทที่ 3 วิธีการดำเนินงานวิจัย .....                          | 28 |
| บทที่ 4 ผลการวิจัย .....                                     | 36 |
| 4.1 ตัวแบบพยากรณ์จากการวิเคราะห์ข้อมูลอนุกรมเวลา .....       | 36 |
| 4.2 ตัวแบบพยากรณ์จากการวิเคราะห์การถดถอยเชิงเส้นแบบพหุ ..... | 41 |
| บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ .....                    | 56 |
| 5.1 สรุปผลการวิจัย .....                                     | 56 |
| 5.2 ข้อเสนอแนะ .....   | 56 |
| บรรณานุกรม .....   | 57 |
| ภาคผนวก .....  | 60 |
| ประวัติผู้เขียน .....  | 81 |

# สารบัญตาราง

| ตารางที่  | หน้า |
|---|------|
| 1.1 ปริมาณการผลิตรถยนต์ 5 ประเทศในอาเซียน ปี 2557-2558 .....                                  | 1    |
| 2.1 ข้อมูลอนุกรมเวลา .....  | 7    |
| 2.2 ตัวอย่างตัวแปรเชิงคุณภาพ การกำหนดตัวแปรหุ่นและค่าของตัวแปรหุ่น .....                      | 19   |
| 2.3 ระดับความสัมพันธ์ของข้อมูลจากค่าสัมประสิทธิ์สหสัมพันธ์ .....                              | 23   |
| 4.1 ตารางแสดงค่าดัชนีสถานการณ์ที่ได้จากขั้นตอนที่ 2.2 – 2.4 .....                             | 36   |
| 4.2 ตารางแสดงค่าดัชนีฤดูกาลที่ได้จากขั้นตอนที่ 2.6 – 2.8 .....                                | 37   |
| 4.3 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของยอดขายรถยนต์กับค่า GDP และอัตราดอกเบี้ย .....             | 42   |
| 4.4 แสดงค่าสัมประสิทธิ์การถดถอยของตัวแบบพยากรณ์ถดถอยเชิงเส้นแบบพหุ<br>ตัวแบบที่ 1 .....       | 43   |
| 4.5 ตารางกำหนดรหัสตัวเลขสำหรับตัวแปรปัจจัยฤดูกาล .....  | 46   |
| 4.6 แสดงข้อมูลยอดขายรถยนต์รายเดือน GDP อัตราดอกเบี้ยและปัจจัยฤดูกาล .....                     | 47   |
| 4.7 แสดงค่าสัมประสิทธิ์การถดถอยของตัวแบบพยากรณ์ถดถอยเชิงเส้นแบบพหุ<br>ตัวแบบที่ 2 .....       | 48   |
| 4.8 ตารางกำหนดรหัสตัวเลขสำหรับตัวแปรปัจจัยสถานการณ์พิเศษ .....                                | 50   |
| 4.9 แสดงยอดขายรถยนต์รายเดือน GDP อัตราดอกเบี้ย เมื่อมีปัจจัยฤดูกาล<br>และสถานการณ์พิเศษ ..... | 51   |
| 4.10 แสดงค่าสัมประสิทธิ์การถดถอยของตัวแบบพยากรณ์ถดถอยเชิงเส้นแบบพหุ<br>ตัวแบบที่ 3 .....      | 52   |
| 4.11 แสดงค่าทางสถิติของตัวแบบทั้ง 3 ตัวแบบ .....  | 54   |

# สารบัญรูป

| รูปที่ | หน้า   |
|--------|--|
| 1.1    | กราฟแสดงยอดขายรถยนต์นั่งขนาดเล็กมากรายเดือนของประเทศไทย<br>ระหว่างเดือนมกราคม 2550 ถึงเดือนธันวาคม 2559 ..... 2                                    |
| 2.1    | แนวโน้มยอดขายของสินค้าชนิดหนึ่ง ..... 8  |
| 2.2    | ข้อมูลอนุกรมเวลาที่มีการแปรผันตามฤดูกาลเชิงบวก ..... 9   |
| 2.3    | ข้อมูลอนุกรมเวลาที่มีการแปรผันตามฤดูกาลเชิงลบ ..... 9  |
| 2.4    | ข้อมูลอนุกรมเวลาที่มีการแปรผันตามวัฏจักรของปริมาณการขายสินค้าหนึ่ง ..... 9   |
| 3.1    | กราฟแสดงยอดขายรถยนต์นั่งขนาดเล็กมากรายเดือนของประเทศไทย<br>ระหว่างเดือนมกราคม 2550 ถึงเดือนมกราคม 2561 ..... 29                                    |
| 4.1    | กราฟเปรียบเทียบข้อมูลอนุกรมเวลาของยอดขายรถยนต์กับข้อมูลที่ถอดผลกระทบ<br>ของสถานการณ์ออกจากข้อมูลอนุกรมเวลาของยอดขายรถยนต์ (De-event data) ..... 37 |
| 4.2    | กราฟเปรียบเทียบข้อมูลอนุกรมเวลาของยอดขายรถยนต์กับข้อมูลยอดขายรถยนต์<br>ที่ถูกถอดส่วนประกอบของอนุกรมเวลาทั้ง 3 ส่วน ..... 38                        |
| 4.3    | กราฟแสดงยอดขายรถยนต์นั่งขนาดเล็กมากเปรียบเทียบกับค่าพยากรณ์<br>ที่ได้จากวิธีการแยกส่วนประกอบอนุกรมเวลา ..... 39                                    |
| 4.4    | กราฟแสดงยอดขายรถยนต์นั่งขนาดเล็กมากรายเดือนของประเทศไทย<br>ระหว่างเดือนมกราคม 2548 ถึงเดือนกันยายน 2558 ..... 41                                   |
| 4.5    | แผนภาพการกระจายข้อมูลระหว่าง GDP และ ยอดขายรถยนต์<br>(ระหว่างเดือนมกราคม 2548 – กันยายน 2558) ..... 42   |
| 4.6    | แผนภาพการกระจายข้อมูลระหว่างอัตราดอกเบี้ยกับยอดขายรถยนต์<br>(ระหว่างเดือนมกราคม 2548 – กันยายน 2558) ..... 43                                      |
| 4.7    | กราฟเปรียบเทียบยอดขายรถยนต์กับค่าที่ได้จากตัวแบบการพยากรณ์ที่<br>ใช้ GDP และอัตราดอกเบี้ย ..... 44   |
| 4.8    | ข้อมูลเปรียบเทียบยอดขายรถยนต์รายเดือน ปี 2548-2558 ..... 45  |
| 4.9    | ข้อมูลเปรียบเทียบยอดขายรถยนต์รายเดือน ปี 2548-2553 ..... 45  |
| 4.10   | ข้อมูลเปรียบเทียบยอดขายรถยนต์รายเดือนเมื่อมีการตัดข้อมูล ..... 46  |
| 4.11   | กราฟเปรียบเทียบยอดขายรถยนต์กับค่าที่ได้จากตัวแบบการพยากรณ์<br>ที่ใช้ GDP และอัตราดอกเบี้ย เมื่อเพิ่มปัจจัยฤดูกาล ..... 48                          |
| 4.12   | กราฟเปรียบเทียบยอดขายรถยนต์กับค่าที่ได้จากตัวแบบการพยากรณ์<br>ที่ใช้ GDP และอัตราดอกเบี้ย เมื่อเพิ่มปัจจัยฤดูกาลและสถานการณ์พิเศษ ..... 53         |

# บทที่ 1

## บทนำ

### 1.1 อุตสาหกรรมยานยนต์ในประเทศไทย

อุตสาหกรรมในประเทศไทยมีการพัฒนาต่อเนื่องในช่วงเวลาหลายปีที่ผ่านมา ในอดีตนโยบายของรัฐบาลถือเป็นกลไกที่สำคัญที่สุดในการพัฒนาประเทศ รวมถึงการพัฒนาอุตสาหกรรมยานยนต์ในประเทศ โดยมีเป้าหมายหลักเพื่อทดแทนการนำเข้ารถยนต์จากต่างประเทศ เห็นได้ชัดจากการขยายฐานการผลิตเครื่องยนต์อุตสาหกรรม ชิ้นส่วนยานยนต์และอุตสาหกรรมที่เกี่ยวข้องอื่น ๆ ที่มีการเติบโตอย่างต่อเนื่อง (กลุ่มอุตสาหกรรมยานยนต์แห่งประเทศไทย, 2559) ประเทศไทยมีความสามารถด้านการผลิตสูงสุดในอาเซียน โดยเฉพาะภายหลังรัฐบาลประกาศนโยบายเปิดเสรียานยนต์ในปี 2534 ยอดจำหน่ายยานยนต์ในประเทศและการประกอบรถยนต์ขยายตัวอย่างต่อเนื่องทั้งด้านการผลิต การตลาด การจ้างงาน การพัฒนาเทคโนโลยี และความเชื่อมโยงต่ออุตสาหกรรมต่อเนื่องอื่น ๆ รวมทั้งการลงทุนในประเทศ จึงถือได้ว่าประเทศไทยเป็นฐานการผลิตของผู้ผลิตรายานยนต์ทั่วโลก และยังเป็นฐานการผลิตรถยนต์ปิกอัป รวมทั้งรถจักรยานยนต์อันดับต้นๆ ของโลก (สุนทรตรา, 2559)

ในปี 2558 การผลิตรถยนต์ของอาเซียนมีจำนวน 3.89 ล้านคัน ลดลงจากปี 2557 ร้อยละ 15 โดยมีประเทศไทยเป็นผู้ผลิตรายใหญ่ที่สุดในอาเซียน ที่มีการผลิตทั้งสิ้น 1.9 ล้านคัน เพิ่มขึ้นจากปี 2557 คิดเป็นร้อยละ 2 ซึ่งในปีนี้เป็นยอดการผลิตรถยนต์นั่งส่วนบุคคล 7.6 แสนคัน และรถยนต์เพื่อการพาณิชย์ 1.15 ล้านคัน ลำดับรองลงมาคืออินโดนีเซีย 1.1 ล้านคัน และมาเลเซีย 6.14 แสนคัน ดังตาราง 1.1

ตาราง 1.1 ปริมาณการผลิตรถยนต์ 5 ประเทศในอาเซียน ปี 2557-2558 (หน่วย : คัน)

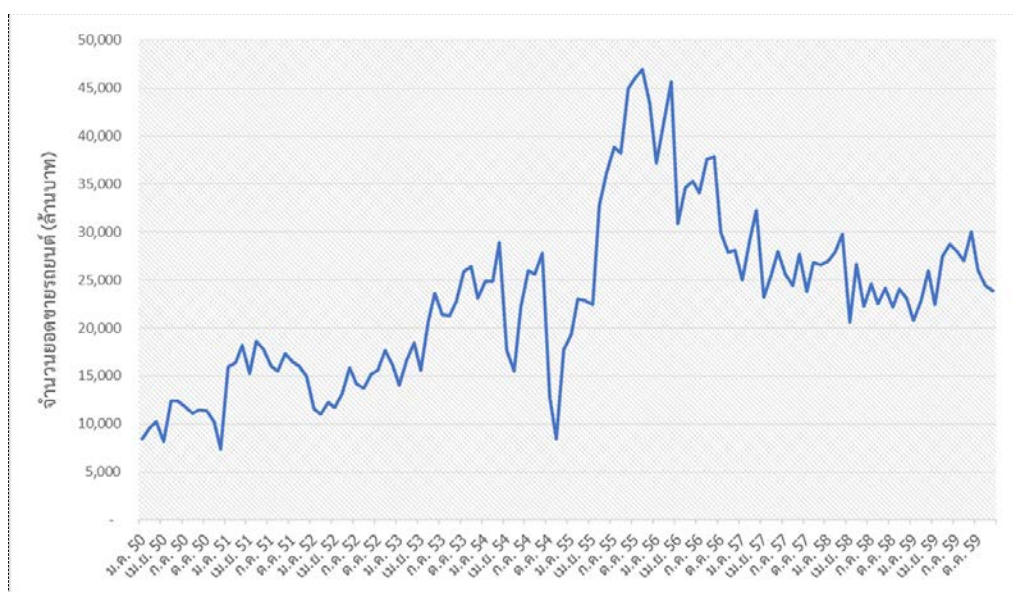
| ประเทศ         | รถยนต์นั่งส่วนบุคคล | รถยนต์เพื่อการพาณิชย์ | 2558      | 2557      | การเปลี่ยนแปลง (%) |
|----------------|---------------------|-----------------------|-----------|-----------|--------------------|
| 1. ไทย         | 760,688             | 1,152,314             | 1,913,002 | 1,880,007 | 2                  |
| 2. อินโดนีเซีย | 824,445             | 274,335               | 1,098,780 | 1,298,523 | -15                |
| 3. มาเลเซีย    | 563,883             | 50,781                | 614,664   | 596,418   | 3                  |
| 4. เวียดนาม    | 99,052              | 72,701                | 171,753   | 121,084   | 42                 |
| 5. ฟิลิปปินส์  | 36,395              | 62,373                | 98,768    | 88,854    | 11                 |
| รวม            | 2,284,463           | 1,612,504             | 3,896,967 | 3,984,877 | -2                 |

## 1.2 วัตถุประสงค์ของงานวิจัย

1. เพื่อศึกษาเทคนิคการวิเคราะห์ข้อมูลอนุกรมเวลาและการพยากรณ์เชิงปริมาณ
2. เพื่อวิเคราะห์ปัจจัยที่มีผลต่อยอดขายรถยนต์
3. เพื่อพัฒนาตัวแบบการพยากรณ์การถดถอยของยอดขายรถยนต์ที่พิจารณาฤดูกาลและสถานการณ์พิเศษ
4. เพื่อพัฒนาตัวแบบการพยากรณ์อนุกรมเวลาแบบแยกส่วนประกอบที่มีฤดูกาลและสถานการณ์พิเศษเข้ามาเกี่ยวข้อง

## 1.3 ขอบเขตของงานวิจัย

งานวิจัยนี้จะศึกษาข้อมูลยอดขายรถยนต์ประเภทรถยนต์นั่งขนาดเล็กมาก (Subcompact car) ซึ่งเป็นหนึ่งในประเภทของรถยนต์ที่นิยมมากในประเทศไทย เนื่องจากราคาไม่แพงและมีขนาดเล็กพอใช้งานได้ เป็นรถที่มักใช้ในพื้นที่เมืองที่หนาแน่น อีกทั้งยังมีระบบความปลอดภัยที่สูงอีกด้วย (CIVIC ES Group, 2555) โดยศึกษายอดขายรถยนต์ประเภทนี้ในประเทศไทย ตั้งแต่เดือนมกราคม 2548 ถึงเดือนธันวาคม 2561



รูปที่ 1.1 กราฟแสดงยอดขายรถยนต์นั่งขนาดเล็กมากรายเดือนของประเทศไทย ระหว่างเดือนมกราคม 2550 ถึงเดือนธันวาคม 2559

จะเห็นได้ว่า ข้อมูลยอดขายรถยนต์ในช่วงเดือนกันยายน 2554 ถึงเดือนมกราคม 2556 มีความผันผวนอันเนื่องมาจากภาวะน้ำท่วมใหญ่ในปี 2554 ที่ส่งผลกระทบต่ออุตสาหกรรมการผลิตรถยนต์ ตลอดจนอุตสาหกรรมอื่น ๆ ที่เกี่ยวข้อง และเป็นช่วงที่มีมาตรการคั่นเงินภาษีรถยนต์คันแรก ในปี 2554 -2556 ซึ่งเป็นโครงการส่งเสริมให้ผู้ที่ไม่มียอดรถยนต์เป็นของตนเอง สามารถใช้สิทธิ์ซื้อรถยนต์ใหม่คันแรกได้ตามนโยบายรัฐบาล และได้รับการลดหย่อนภาษีตามที่รัฐบาลกำหนด นั่นคือ

จะได้การคืนเงินภาษีเท่ากับที่จ่ายจริงในการซื้อรถยนต์คันแรก แต่จะไม่เกิน 100,000 บาท และภาครัฐจะคืนเงินภาษีได้เมื่อครบครองรถยนต์ไปแล้วเป็นเวลา 1 ปี (กรมสรรพสามิต, 2555)

ในการศึกษาครั้งนี้ใช้เทคนิคการพยากรณ์อนุกรมเวลาแบบแยกส่วนประกอบ (Decomposition Time Series Forecasting) และการวิเคราะห์การถดถอยพหุคูณ (Multiple Regression) โดยพิจารณาปัจจัยฤดูกาล (season) และสถานการณ์พิเศษ (special event) ซึ่งแบ่งเป็น 4 ช่วง ได้แก่ ช่วงก่อนเกิดน้ำท่วมใหญ่ ช่วงที่เกิดน้ำท่วมใหญ่และมีมาตรการคืนภาษีรถยนต์คันแรก ช่วงที่มีมาตรการคืนภาษีรถคันแรก และช่วงหลังจากสิ้นสุดมาตรการคืนภาษีรถยนต์คันแรก การวัดความแม่นยำของการพยากรณ์ที่ใช้คือ ค่าเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์เฉลี่ย (Mean Absolute Percentage Error : MAPE) และค่าสัมประสิทธิ์ของการตัดสินใจ ( $R^2$ ) พร้อมทั้งแสดงค่าความแม่นยำของการพยากรณ์จากค่าเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์เฉลี่ย (Mean Absolute Percentage Error : MAPE) ด้วยเช่นกัน

#### 1.4 ขั้นตอนการวิจัย

1. วิเคราะห์ข้อมูลอนุกรมเวลาของยอดขายรถยนต์ที่มีฤดูกาลและสถานการณ์พิเศษ
2. ศึกษาการเทคนิคการพยากรณ์อนุกรมเวลา
3. ศึกษาการพยากรณ์เชิงปริมาณด้วยการวิเคราะห์การถดถอยพหุคูณ
4. สร้างตัวแบบพยากรณ์ยอดขายรถยนต์ในรูปแบบเงื่อนไขต่าง ๆ
5. เปรียบเทียบความถูกต้องแม่นยำของตัวแบบพยากรณ์โดยใช้ค่าเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์เฉลี่ย (MAPE) และค่าสัมประสิทธิ์ของการตัดสินใจ ( $R^2$ )

#### 1.5 ผลที่คาดว่าจะได้รับ

1. ได้ตัวแบบพยากรณ์ยอดขายรถยนต์
2. ได้ทราบถึงเทคนิคการวิเคราะห์ข้อมูลอนุกรมเวลาที่เหมาะสมในการพยากรณ์
3. ได้ทราบถึงปัจจัยที่ส่งผลต่อยอดขายรถยนต์
4. สามารถนำผลที่ได้จากการศึกษาไปใช้ในการวางแผนการผลิตรถยนต์เพื่อรองรับสถานการณ์ต่าง ๆ

## บทที่ 2

# ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 ทฤษฎีที่ใช้ในการวิเคราะห์

#### 2.1.1 การพยากรณ์และวิธีการพยากรณ์

**การพยากรณ์ (Forecasting)** หมายถึงการคาดคะเนหรือการทำนายลักษณะการเกิดเหตุการณ์หรือสภาวะการณ์ในอนาคต โดยศึกษารูปแบบการเกิดเหตุการณ์หรือสภาวะการณ์จากข้อมูลที่เกี่ยวข้องอย่างมีระบบหรือจากความรู้ความสามารถ ประสบการณ์ และวิจญาณของผู้พยากรณ์ (ทรงศิริ, 2549) ดังนั้นการพยากรณ์จึงเป็นเครื่องมือหนึ่งที่มีบทบาทสำคัญในการสนับสนุนการตัดสินใจ และนับเป็นพื้นฐานที่สำคัญต่อการวางแผนและตัดสินใจที่มีประสิทธิภาพขององค์กร ซึ่งถ้าองค์กรสามารถที่จะคาดการณ์หรือพยากรณ์สิ่งที่อาจจะเกิดขึ้นในอนาคตได้อย่างถูกต้องแม่นยำและทันต่อสถานการณ์มากเพียงใด ก็ยิ่งจะได้รับประโยชน์และลดความเสี่ยงที่อาจจะเกิดขึ้นได้มากยิ่งขึ้นด้วย อย่างไรก็ตามการเกิดสถานการณ์ใดสถานการณ์หนึ่งในอนาคตนั้น อาจมีปัจจัยที่ส่งผลกระทบต่อหลากหลายปัจจัย ดังนั้นการเลือกเทคนิควิธีการพยากรณ์ให้เหมาะสมกับลักษณะของข้อมูลจึงเป็นสิ่งสำคัญที่จะส่งผลต่อความแม่นยำในการพยากรณ์ด้วยเช่นกัน วิธีการพยากรณ์ที่ใช้กันอยู่ทั่วไป แบ่งได้เป็น 2 ประเภทใหญ่ ๆ (มุกดา, 2549) คือ

**2.1.1.1 วิธีการพยากรณ์เชิงคุณภาพ** เป็นการพยากรณ์เหตุการณ์ที่ไม่มีข้อมูลในอดีตที่เกี่ยวข้องกับเหตุการณ์นั้น ๆ การพยากรณ์จะใช้ความคิดเห็นของผู้เชี่ยวชาญเป็นหลัก

**2.1.1.2 วิธีการพยากรณ์เชิงปริมาณ** เป็นการพยากรณ์ที่อาศัยข้อมูลในอดีตที่เกี่ยวข้องกับเหตุการณ์นั้น ๆ เพื่อนำมาสร้างตัวแบบทางคณิตศาสตร์หรือสถิติ

การพยากรณ์เชิงปริมาณอาจจำแนกตามปรัชญาของการพยากรณ์ได้เป็นสองประเภท ดังนี้ ประเภทที่ 1 เป็นวิธีที่มีแนวความคิดว่าพฤติกรรมในอดีตของสิ่งที่จะพยากรณ์เพียงพอที่จะพยากรณ์พฤติกรรมในอนาคต วิธีการพยากรณ์ประเภทนี้ ได้แก่ เทคนิคการทำให้เรียบ (Smoothing Techniques) เป็นต้น ประเภทที่ 2 เป็นวิธีที่มีแนวความคิดว่าพฤติกรรมของสิ่งที่จะพยากรณ์ถูกกำหนดขึ้นโดยสิ่งอื่น ๆ ซึ่งมีความสัมพันธ์บางลักษณะกับสิ่งที่จะพยากรณ์ ตัวอย่างการพยากรณ์ประเภทที่ 2 ได้แก่ การพยากรณ์ด้วยการวิเคราะห์การถดถอย (Regression Analysis) เป็นต้น

#### 2.1.2 การเลือกเทคนิคในการพยากรณ์

ปัจจุบันการพยากรณ์เหตุการณ์ต่าง ๆ มีวิธีการพยากรณ์ตลอดจนมีผู้พัฒนาวิธีการพยากรณ์ขึ้นหลากหลายวิธี ซึ่งแต่ละวิธีก็จะเหมาะสมกับข้อมูลที่มีลักษณะที่แตกต่างกันออกไป รวมทั้งขึ้นอยู่กับวัตถุประสงค์ของการนำค่าพยากรณ์ที่ได้ไปใช้งานด้วย จึงจำเป็นอย่างยิ่งในการเลือกวิธีการในการพยากรณ์เพื่อให้เหมาะสมกับการนำไปใช้ ดังนั้นผู้พยากรณ์จึงควรศึกษารายละเอียดต่าง ๆ ที่เกี่ยวข้องก่อนที่จะดำเนินการ ดังนี้

1. ระยะเวลาในการพยากรณ์ล่วงหน้า ผู้พยากรณ์มักจะพยากรณ์การเกิดขึ้นของเหตุการณ์แตกต่างกัน ทั้งนี้ขึ้นอยู่กับวัตถุประสงค์ของการนำค่าพยากรณ์ไปใช้งาน ระยะเวลาในการพยากรณ์ล่วงหน้าสามารถแบ่งได้เป็น ระยะสั้น ระยะกลาง และระยะยาว การพยากรณ์ระยะสั้นจะเป็นช่วงเวลาที่พยากรณ์ล่วงหน้าไม่เกิน 3 เดือน การพยากรณ์ระยะกลางจะเป็นช่วงเวลาตั้งแต่ 3 เดือนขึ้นไป จนถึง 3 ปี และการพยากรณ์ระยะยาวเป็นช่วงเวลาที่ต้องการพยากรณ์ล่วงหน้าตั้งแต่ 3 ปีขึ้นไป ระยะเวลาในการพยากรณ์ล่วงหน้านี้จะส่งผลถึงเทคนิควิธีในการเลือกใช้วิธีการในการพยากรณ์ โดยแต่ละช่วงเวลาก็จะเหมาะสมกับเทคนิคการพยากรณ์ที่แตกต่างกันออกไป เช่น หากผู้พยากรณ์ต้องการที่จะพยากรณ์ในระยะยาวแล้ว การพยากรณ์เชิงคุณภาพจะมีความเหมาะสมมากกว่า ข้อมูลที่ใช้ในการพยากรณ์นั้นอาจมีหน่วยวัดเป็น รายชั่วโมง รายวัน รายสัปดาห์ รายเดือน หรือรายไตรมาส ก็ได้ ขึ้นอยู่กับประเด็นที่ศึกษา เช่น บริษัทผลิตอาหารสำเร็จรูปต้องการพยากรณ์ยอดขายสินค้าอาหารสำเร็จรูปในแต่ละเดือน เป็นต้น อย่างไรก็ตามข้อมูลที่นำมาใช้ในกระบวนการวิเคราะห์ควรจะต้องมีขนาดใหญ่พอสมควรและมีความทันสมัยด้วย

2. รูปแบบของข้อมูล ส่วนประกอบของข้อมูลจะเป็นตัวกำหนดเทคนิคที่ใช้ในการพยากรณ์ ดังนั้นก่อนที่ผู้พยากรณ์จะเลือกเทคนิควิธีในการพยากรณ์จึงจำเป็นต้องนำข้อมูลที่มืออยู่มาพล็อตกราฟแบบจุด เพื่อพิจารณาลักษณะการเคลื่อนไหวของข้อมูลเบื้องต้นก่อน

3. ค่าใช้จ่ายในการพยากรณ์ นับเป็นเกณฑ์ที่สำคัญอีกอย่างหนึ่งที่ใช้ในการตัดสินใจเลือกเทคนิคในการพยากรณ์ อันประกอบด้วยค่าใช้จ่ายที่เกิดขึ้นเพื่อพัฒนาตัวแบบที่ใช้ในการพยากรณ์ ค่าใช้จ่ายที่เกิดจากการเก็บข้อมูลที่จะทำให้ตัวแบบมีความถูกต้องแม่นยำ และค่าใช้จ่ายที่เกิดขึ้นจากการผิดพลาดที่มาจากการพยากรณ์ ความสำคัญของค่าใช้จ่ายแต่ละประเภทจะขึ้นอยู่กับวิธีการและสภาพการณ์ เช่น วิธีการพยากรณ์เชิงคุณภาพและเชิงเทคโนโลยีโดยส่วนใหญ่จะแยกประมาณค่าใช้จ่ายเป็นส่วน ๆ รวมทั้งจะต้องประมาณการปัจจัยนำเข้าทางทรัพยากรมนุษย์ การได้ข้อมูลจากภายนอกจะเป็นค่าใช้จ่ายส่วนใหญ่ที่เกิดขึ้น และค่าใช้จ่ายเหล่านี้จะเกิดขึ้นทุกครั้งที่มีการพยากรณ์ใหม่ สำหรับการพยากรณ์เชิงปริมาณนั้นค่าใช้จ่ายส่วนใหญ่มักไม่ขึ้นอยู่กับสภาพการณ์ ทางด้านการบริหารจัดการ วิธีการพยากรณ์เชิงปริมาณที่ใช้กันในองค์กรส่วนใหญ่จะใช้คอมพิวเตอร์ ในวิเคราะห์และจัดการกับข้อมูล ค่าใช้จ่ายด้านการพัฒนาจึงเป็นเรื่องของการเขียนและดัดแปลงโปรแกรมที่ใช้ในการพยากรณ์ ซึ่งรวมถึงทรัพยากรมนุษย์ที่ต้องใช้เพื่อพัฒนาโปรแกรมและค่าใช้จ่ายเพื่อจัดระบบการทำงานของคอมพิวเตอร์ในแต่ละเทคนิควิธีที่ใช้ในการพยากรณ์ เป็นต้น

4. ระดับความแม่นยำในการพยากรณ์ และความคลาดเคลื่อนที่เกิดขึ้นกับสถานการณ์ที่ต้องการจะพยากรณ์ บางกรณีความผิดพลาด 20% ก็สามารถยอมรับได้ ในขณะที่บางกรณีความผิดพลาด 1% ก็อาจก่อให้เกิดความเสียหายได้มากมายต่อองค์กร และหากพิจารณาในแง่ของการวิเคราะห์การตัดสินใจ ก็พบว่ามีความแตกต่างระหว่างการตัดสินใจที่ดี (Good decision) และผลลัพธ์ที่ดี (Good outcome) ถ้าหากผู้พยากรณ์สามารถออกแบบจำลองสภาพการณ์ได้ท่ามกลางความไม่แน่นอน ก็น่าที่จะส่งเสริมสนับสนุนการพยากรณ์โดยไม่ต้องคำนึงถึงระดับความแม่นยำมากนัก

5. ข้อมูลในอดีตที่เก็บรวบรวมไว้ หรือข้อมูลในอดีตที่มีอยู่ จะเป็นตัวตัดสินใจหลักในการที่จะเลือกเทคนิควิธีในการพยากรณ์ นอกจากนี้ความถูกต้องของข้อมูลก็เป็นปัจจัยที่สำคัญอย่างยิ่งในการพยากรณ์ ดังนั้นหากองค์กรใดยังไม่ได้มีการจัดเก็บข้อมูลในอดีตจึงมีความจำเป็นอย่างยิ่งที่จะต้องรีบดำเนินการเพื่อให้ได้ข้อมูลที่จะใช้ในการพยากรณ์ต่อไป

6. ความยากง่ายของเทคนิควิธีการพยากรณ์ในการที่ผู้ปฏิบัติจะนำไปใช้ต่อ เป็นอีกปัจจัยหนึ่งที่สำคัญ หากผู้ที่เกี่ยวข้องขาดความเข้าใจในเทคนิคการพยากรณ์ที่นำมาใช้ ตัวแบบการพยากรณ์ที่สร้างขึ้นก็จะไม่มีความหมาย หรืออาจถูกนำไปใช้อย่างไม่ถูกต้อง

### 2.1.3 กระบวนการพยากรณ์ (Forecasting Process)

กระบวนการพยากรณ์ หมายถึง ขั้นตอนในการเลือกเทคนิควิธีที่ใช้ในการพยากรณ์ ตั้งแต่หนึ่งวิธีขึ้นไป ที่สามารถประยุกต์ใช้กับข้อมูลที่จำเป็นต่อการพยากรณ์ ซึ่งประกอบด้วย 9 ขั้นตอน (ณัฐกานต์, 2559) ดังต่อไปนี้

1. การกำหนดวัตถุประสงค์การพยากรณ์ (Specific objectives) เป็นการกำหนดจุดมุ่งหมายที่ชัดเจนในการพยากรณ์ว่าการพยากรณ์จะใช้ในการตัดสินใจอย่างไร เช่น ใช้เพื่อตัดสินใจในการลงทุน (การพยากรณ์ระยะยาว) หรือเพื่อการวางแผนกลยุทธ์ (การพยากรณ์ระยะกลาง)

2. การกำหนดสิ่งที่จะพยากรณ์ให้ชัดเจน (Determine what to forecast) เช่น การพยากรณ์ยอดขายเป็นหน่วยสินค้า หรือเป็นตัวเงิน (บาทหรือดอลลาร์) การพยากรณ์เป็นยอดขายรวม ยอดขายสายผลิตภัณฑ์ ยอดขายของแต่ละภูมิภาค หรือยอดขายในประเทศ หรือยอดขายต่างประเทศ เป็นต้น

3. การเก็บรวบรวมข้อมูล (Data collection) เป็นการเก็บรวบรวมข้อมูลที่เกี่ยวข้องกับการพยากรณ์อย่างเหมาะสม และเป็นข้อมูลที่ต้องการ ขั้นตอนนี้เป็นขั้นตอนที่ยากและใช้เวลามากที่สุด โดยในการเก็บรวบรวมข้อมูลต้องพิจารณาถึงข้อกำหนดด้านเวลา (Identify time dimensions) โดยพิจารณา 2 ประการ คือ ช่วงระยะเวลาการพยากรณ์ (Length and periodicity) เช่น ประจำปี ประจำไตรมาส ประจำเดือน ประจำวัน และความเร่งในการพยากรณ์ (Urgency) ถ้ามีความจำเป็นเร่งด่วน วิธีที่ใช้ในการพยากรณ์จะมีความซับซ้อนน้อยกว่า และข้อกำหนดเกี่ยวกับข้อมูล (Data consideration) การพิจารณาจากปริมาณและประเภทของข้อมูลที่เป็นข้อมูลภายในหรือภายนอกบริษัท เป็นข้อมูลรายปี รายเดือน เป็นข้อมูลที่เป็นตัวเงินหรือหน่วยสินค้า

4. การลดข้อมูล (Data reduction) บางครั้งข้อมูลที่เก็บรวบรวมมามีมากเกินไป และทำให้การพยากรณ์มีความถูกต้องน้อยลง จึงจำเป็นต้องลดข้อมูลบางตัวที่อาจไม่เกี่ยวข้องกับการพยากรณ์ลง

5. การเลือกแบบจำลองในการพยากรณ์ (Model selection) ขึ้นอยู่กับรูปแบบของข้อมูล จำนวนข้อมูลที่มี และระยะเวลาในการพยากรณ์ การเลือกวิธีการพยากรณ์ที่เหมาะสมกับข้อมูล จะช่วยลดความผิดพลาดในการพยากรณ์ รูปแบบพยากรณ์ที่ง่ายจะยิ่งดีต่อการยอมรับของผู้ตัดสินใจ วิธีการพยากรณ์จะต้องมีความสมดุลระหว่างความถูกต้องและเป็นวิธีที่ง่ายต่อความเข้าใจ

6. การพยากรณ์ (Model extrapolation) เป็นการพยากรณ์เหตุการณ์ที่ผ่านไป โดยใช้ข้อมูลจริงที่มีอยู่ แล้วประเมินว่าวิธีใดเหมาะสม (Fit) กับข้อมูลในอดีตก่อน โดยการวัดค่าความคลาดเคลื่อนที่เกิดขึ้น

7. การเตรียมการพยากรณ์ (Forecast preparation) เมื่อมีวิธีพยากรณ์ที่เหมาะสมมากกว่า 2 วิธีขึ้นไป การรวมค่าพยากรณ์จากวิธีเหล่านี้จะทำให้ค่าพยากรณ์ดีขึ้นกว่าการใช้วิธีเดียว

8. การนำเสนอผลการพยากรณ์ (Forecast presentation) การนำเสนอค่าการพยากรณ์ให้กับผู้บริหาร หรือผู้ใช้ด้วยการเขียนโปรแกรมเป็นลายลักษณ์อักษร หรือนำเสนอด้วยวาจา ซึ่งขั้นตอนนี้มีความสำคัญมากเช่นกัน เพราะสามารถสร้างความเข้าใจให้กับผู้บริหารหรือผู้ใช้ได้

9. การตรวจสอบผลการพยากรณ์ (Tracking results) การติดตามผลอย่างต่อเนื่องว่าผลการพยากรณ์เมื่อเปรียบเทียบกับค่าจริงแล้วมีความถูกต้องอย่างไร ซึ่งวิธีที่เคຍพยากรณ์ได้ดีที่สุด อาจมีความถูกต้องลดลงเนื่องจากสภาพแวดล้อมเปลี่ยนไป อาจต้องหาวิธีอื่นมาทดแทน

## 2.1.4 การวิเคราะห์อนุกรมเวลา (Time Series Analysis)

**2.1.4.1 ข้อมูลอนุกรมเวลา (Time Series Data)** หมายถึง ข้อมูลหรือค่าสังเกตที่เปลี่ยนแปลงไปตามลำดับของเวลา โดยข้อมูลที่เกิดขึ้นจะเป็นข้อมูลที่อยู่ในช่วงเวลาที่น่าสนใจของผู้วิเคราะห์ ซึ่งจะช่วยให้เห็นรูปแบบของการเปลี่ยนแปลงค่าสังเกตในช่วงเวลาที่ผ่านมามา เพื่อใช้ในการพยากรณ์ค่าสังเกตดังกล่าวในอนาคต

**2.1.4.2 วิธีการพยากรณ์อนุกรมเวลา (Time Series Forecasting Method)** เป็นการวิเคราะห์พฤติกรรมในอดีตของตัวแปรอนุกรมเวลา เพื่อพยากรณ์พฤติกรรมในอนาคต โดยถ้าสามารถค้นพบพฤติกรรมที่เป็นระบบบางอย่างในตัวแปรอนุกรมเวลา ผู้วิเคราะห์ก็จะสามารถสร้างแบบจำลองของพฤติกรรมของตัวแปรตาม แล้วนำมาใช้ในการพยากรณ์พฤติกรรมของตัวแปรเหล่านั้นในอนาคต เช่น สมมติว่ามีข้อมูลอนุกรมเวลาเรียงตามระยะเวลา  $t - n$  จนถึง  $t$  ดังแสดงในตาราง 2.1

ตาราง 2.1 ข้อมูลอนุกรมเวลา

| ช่วงเวลาที่ | ค่าของตัวแปร |
|-------------|--------------|
| $t - n$     | $Y_{t-n}$    |
| $\vdots$    | $\vdots$     |
| $t - 1$     | $Y_{t-1}$    |
| $t$         | $Y_t$        |

เทคนิคที่ใช้วิเคราะห์พฤติกรรมของตัวแปรอนุกรมเวลาเพื่อพยากรณ์ค่าตัวแปรในอนาคต คือ การใช้แบบจำลองการประมาณค่านอกช่วง (Extrapolation) ซึ่งมีรูปแบบทางคณิตศาสตร์โดยทั่วไป ดังนี้

$$\hat{Y}_{t+1} = f(Y_t, Y_{t-1}, \dots, Y_{t-n}) \quad (2.1)$$

โดย  $\hat{Y}_{t+1}$  คือ ค่าตัวแปรอนุกรมเวลาที่พยากรณ์ได้ในช่วงเวลา  $t + 1$

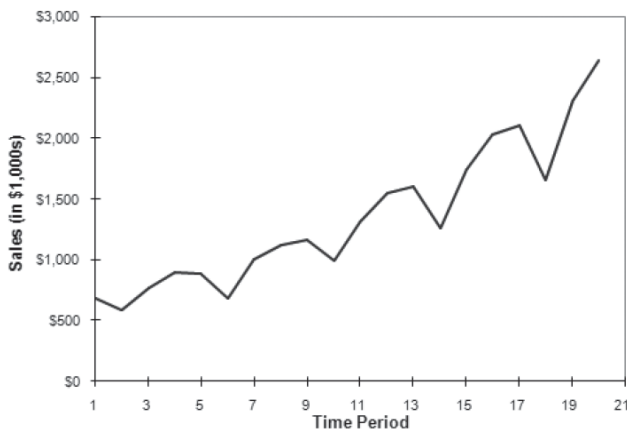
$Y_t$  คือ ค่าตัวแปรอนุกรมเวลาในช่วงเวลา  $t$

$Y_{t-1}$  คือ ค่าตัวแปรอนุกรมเวลาในช่วงเวลา  $t-1$

จุดประสงค์ของแบบจำลองการประมาณค่านอกช่วง คือ การระบุฟังก์ชันตามสมการ (2.1) เพื่อใช้พยากรณ์ค่าของตัวแปรอนุกรมเวลาในอนาคต

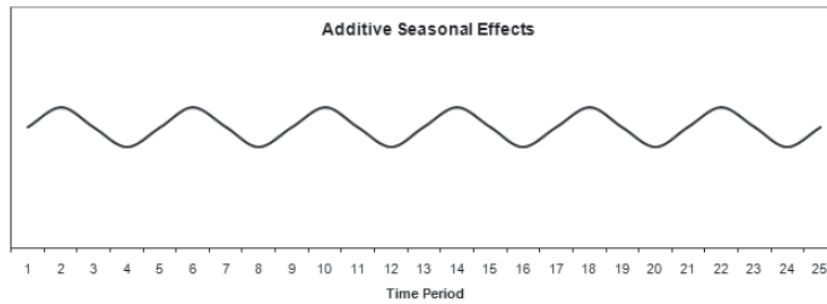
ข้อมูลอนุกรมเวลาที่เก็บรวบรวมได้ในช่วงระยะเวลาที่เท่ากันจะมีค่าเปลี่ยนแปลงซึ่งจะเป็นตัวแปรตาม ส่วนระยะเวลาที่เปลี่ยนแปลงจะเป็นตัวแปรอิสระ สาเหตุที่ทำให้อนุกรมเวลาเปลี่ยนแปลงจะเรียกว่า ส่วนประกอบของอนุกรมเวลา (Components of a time series) แบ่งได้เป็น 4 ส่วนประกอบด้วยกัน ดังนี้

**1. แนวโน้ม (Trend :  $T$ )** เป็นสาเหตุที่ทำให้ข้อมูลอนุกรมเวลาเปลี่ยนแปลงในระยะเวลา มากกว่าหนึ่งปี ซึ่งการเปลี่ยนแปลงอาจจะเพิ่มขึ้นหรือลดลง อาจจะเป็นการเปลี่ยนแปลงอย่างรวดเร็วหรือช้า ๆ ซึ่งลักษณะแนวโน้มของการเปลี่ยนแปลงอาจจะเป็นเชิงเส้น หรือเส้นโค้ง พาราโบลา หรือเส้นโค้งเอกโพเนนเชียล รูปที่ 2.1 แสดงยอดขายของสินค้าชนิดหนึ่งซึ่งมีแนวโน้มสูงขึ้นเมื่อเวลาผ่านไป

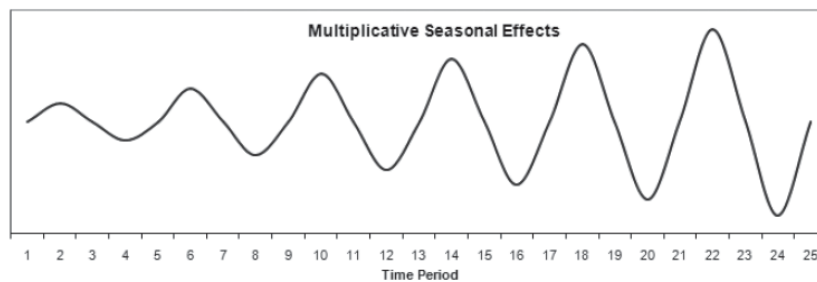


รูปที่ 2.1 แนวโน้มยอดขายของสินค้าชนิดหนึ่ง

**2. การแปรผันตามฤดูกาล (Seasonal Variation :  $S$ )** เป็นสาเหตุที่ทำให้ข้อมูลอนุกรมเวลาเปลี่ยนแปลงในระยะเวลาสั้น ๆ อาจจะเป็นรายวัน รายสัปดาห์ รายเดือน หรือรายไตรมาส จะเรียกช่วงระยะเวลานี้ว่า ฤดูกาล การเปลี่ยนแปลงอาจจะเพิ่มขึ้นหรือลดลงเกิดขึ้นซ้ำ ๆ กัน ในรอบของช่วงเวลาที่ไม่เกินหนึ่งปี ค่าที่วัดการเปลี่ยนแปลงตามฤดูกาล เรียกว่า ดัชนีฤดูกาล (Seasonal Index) สามารถแบ่งเป็น 2 ลักษณะ คือ การแปรผันตามฤดูกาลเชิงบวก และการแปรผันตามฤดูกาลเชิงลบ ดังรูปที่ 2.2 และรูปที่ 2.3 ตามลำดับ การแปรผันตามฤดูกาลเชิงบวกเป็นการเปลี่ยนแปลงที่มีฤดูกาล โดยมีค่าซ้ำ ๆ ไม่ต่างจากฤดูกาลก่อนหน้ามากนัก ส่วนการแปรผันตามฤดูกาลเชิงลบเป็นการเปลี่ยนแปลงที่เพิ่มขึ้นเป็นสัดส่วนกับเวลา

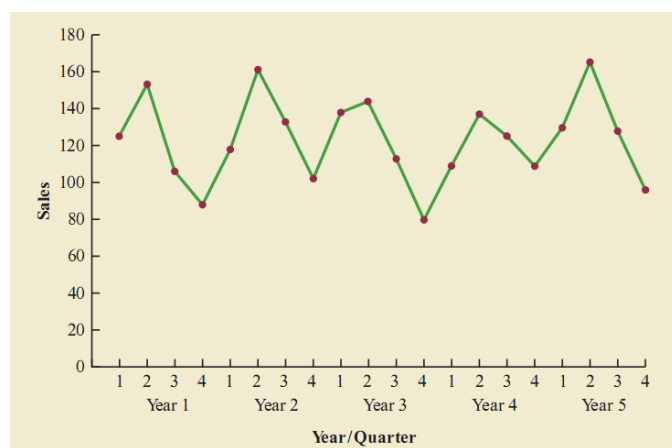


รูปที่ 2.2 ข้อมูลอนุกรมเวลาที่มีการแปรผันตามฤดูกาลเชิงบวก



รูปที่ 2.3 ข้อมูลอนุกรมเวลาที่มีการแปรผันตามฤดูกาลเชิงคูณ

3. การแปรผันตามวัฏจักร (Cyclical Variation :  $C$ ) เป็นสาเหตุที่ทำให้ข้อมูลอนุกรมเวลาเปลี่ยนแปลงขึ้นและลงซ้ำ ๆ กันในช่วงระยะเวลาที่ยาวนานกว่าการแปรผันตามฤดูกาล ช่วงระยะเวลาการเปลี่ยนแปลงไม่แน่นอน รูปที่ 2.4 แสดงปริมาณการขายสินค้าหนึ่งที่มีการแปรผันตามวัฏจักร



รูปที่ 2.4 ข้อมูลอนุกรมเวลาที่มีการแปรผันตามวัฏจักรของปริมาณการขายสินค้าหนึ่ง

4. การเปลี่ยนแปลงของข้อมูล (Irregular Variation :  $I$ ) เป็นสาเหตุที่ทำให้ข้อมูลอนุกรมเวลาเปลี่ยนแปลงขึ้นและลงในระยะเวลาที่ไม่แน่นอน ไม่เป็นไปตามที่ควรจะเป็น มักจะเกิดในช่วงระยะเวลาสั้น ๆ และเป็นการเปลี่ยนแปลงข้อมูลเพียงเล็กน้อย ซึ่งจะเกิดจากเหตุการณ์ที่ไม่สามารถคาดคะเนได้ล่วงหน้า

## 2.2 การวิเคราะห์แนวโน้มและความแปรผันตามฤดูกาล

### 2.2.1 การวิเคราะห์แนวโน้ม

2.2.1.1 การทดสอบแนวโน้ม เมื่อมีการเก็บรวบรวมข้อมูลแล้ว และต้องการทราบว่าอนุกรมเวลาชุดนั้นมีส่วนประกอบใดบ้าง วิธีที่เป็นที่นิยมใช้ก็คือ การนำข้อมูลชุดดังกล่าวมาเขียนเป็นแผนภาพการกระจาย (scatter-diagram) แล้วพิจารณาว่าข้อมูลชุดนั้นมีแนวโน้มหรือไม่ และแนวโน้มมีลักษณะเป็นอย่างไร ในกรณีที่พิจารณาจากแผนภาพการกระจายแล้วยังมีความคลุมเครือว่าข้อมูลชุดนั้นจะมีส่วนประกอบเป็นแนวโน้มจริงหรือไม่ เพื่อให้เกิดความแน่ใจยิ่งขึ้น จึงต้องทดสอบสมมติฐานว่าข้อมูลมีแนวโน้มหรือไม่ ซึ่งในการทดสอบสมมติฐานมีทั้งการทดสอบแบบใช้พารามิเตอร์ (parametric test) และแบบไม่ใช้พารามิเตอร์ (nonparametric test) ในที่นี้จะใช้วิธีแบบไม่ใช้พารามิเตอร์ คือ การทดสอบแบบเครื่องหมาย (sign test)

การทดสอบแบบเครื่องหมาย เป็นการทดสอบแนวโน้มที่พิจารณาว่าค่าสังเกตที่อยู่ติดกันจะมีค่าต่างกันทางบวกหรือทางลบ นั่นคือพิจารณาว่าค่าสังเกต  $Y_i$  มีค่ามากกว่าหรือน้อยกว่า  $Y_{i-1}$  (เมื่อ  $Y_i, Y_{i-1}$  คือค่าข้อมูลอนุกรมเวลาตัวที่  $i$  และ  $i-1$ ) ซึ่งจะพิจารณาได้จากเครื่องหมายของผลต่างครั้งที่หนึ่งของ  $Y_i - Y_{i-1}$  ว่าเป็นบวกหรือลบ กรณีที่ผลต่างเป็นบวกมากกว่าครั้ง หรือเป็นลบมากกว่าครั้ง เป็นผลจากการที่อนุกรมเวลามีแนวโน้มขึ้นหรือแนวโน้มลง การทดสอบมีขั้นตอนดังนี้

1. กำหนดสมมติฐานหลักและสมมติฐานทางเลือก

$H_0$  : อนุกรมเวลาไม่มีแนวโน้ม

$H_1$  : อนุกรมเวลามีแนวโน้มอาจจะขึ้นหรือทางลง

2. สถิติที่ใช้ทดสอบ

$V$  : จำนวนผลต่างครั้งที่หนึ่งของ  $Y_i - Y_{i-1}$  ของอนุกรมเวลาที่เป็นบวก

3. เกณฑ์การตัดสินใจ เมื่ออนุกรมเวลามีขนาดใหญ่  $n \geq 20$  ใช้

$$Z = \frac{V - \mu_V}{\sigma_V}$$

เมื่อ  $\mu_V = \frac{n}{2}$ ,  $\sigma_V = \sqrt{\frac{n}{4}}$ ,  $n$  = จำนวนข้อมูลทั้งหมดของอนุกรมเวลา

ส่วนการยอมรับหรือปฏิเสธ  $H_0$  จะใช้ค่าวิกฤต  $Z_{\frac{\alpha}{2}}$  และ  $-Z_{\frac{\alpha}{2}}$  โดยที่ยอมรับ

$H_0$  เมื่อ  $-Z_{\frac{\alpha}{2}} \leq Z \leq Z_{\frac{\alpha}{2}}$  และปฏิเสธ  $H_0$  เมื่อ  $Z < -Z_{\frac{\alpha}{2}}$  หรือ  $Z > Z_{\frac{\alpha}{2}}$

2.2.1.2 การสร้างเส้นแนวโน้มโดยการประมาณด้วยสายตา (freehand method) การหาเส้นแนวโน้มโดยวิธีนี้ จะต้องนำเอาข้อมูลอนุกรมเวลาที่ต้องการหาเส้นแนวโน้มมา

เขียนแผนภาพการกระจายของข้อมูล โดยให้แกน X แทนระยะเวลา และแกน Y แทนข้อมูลอนุกรมเวลาชุดนั้น แล้วจึงลากเส้นโดยวิธีการประมาณด้วยสายตา การลากเส้นดังกล่าวต้องพยายามลากผ่านจุดเหล่านั้นให้ได้มากที่สุดหรือใกล้เคียงที่สุด และต้องพยายามให้จุดที่อยู่เหนือเส้นและใต้เส้นนี้มีจำนวนเท่ากันหรือเกือบเท่ากัน การหาเส้นแนวโน้มโดยวิธีนี้เป็นวิธีที่ง่าย สะดวกไม่สิ้นเปลืองเวลา แต่มีข้อเสียที่เป็นวิธีที่อาศัยดุลยพินิจของผู้ทำการวิเคราะห์เป็นสำคัญ เนื่องจากการประมาณโดยวิธีนี้เป็นวิธีการที่ง่าย การหาเส้นแนวโน้มโดยวิธีนี้จึงมักจะนำมาใช้เป็นจุดเริ่มต้นในการพิจารณาว่าข้อมูลอนุกรมเวลาชุดนั้นมีแนวโน้มเป็นเส้นตรงหรือเส้นโค้ง เพื่อจะนำไปคำนวณหาแนวโน้มที่ถูกต้องและเหมาะสมต่อไป

**2.2.1.3 การสร้างเส้นแนวโน้มด้วยวิธีกำลังสองน้อยที่สุด (least-square method)** เป็นวิธีสร้างเส้นแนวโน้มโดยการสร้างสมการทางคณิตศาสตร์ ผู้สร้างจำเป็นต้องตัดสินใจในการเลือกตัวแบบ (สมการแนวโน้ม) และมีจำนวนข้อมูลมากพอที่จะใช้ในการวิเคราะห์ เป็นวิธีการหาค่าแนวโน้มที่ให้ผลรวมของผลต่างระหว่างค่าข้อมูลที่เก็บรวบรวมได้ ( $Y$ ) กับค่าแนวโน้ม ( $\hat{Y}$ ) ยกกำลังสอง หรือ  $\sum(Y - \hat{Y})^2$  มีค่าน้อยที่สุด (least-square error)

## 2.2.2 การวิเคราะห์ความแปรผันตามฤดูกาล

ความแปรผันตามฤดูกาล (seasonal variation) เป็นส่วนประกอบหนึ่งของอนุกรมเวลา ที่แสดงการเคลื่อนไหวของข้อมูลอนุกรมเวลาที่มีผลมาจากอิทธิพลของฤดูกาล โดยทั่วไปความผันแปรตามฤดูกาลนั้นจะมีลักษณะการเปลี่ยนแปลงอยู่ 2 ประการ คือ การเปลี่ยนแปลงที่คงที่ในช่วงเวลาหนึ่ง ๆ เรียกว่าฤดูกาลคงที่ (constant seasonal) และการเปลี่ยนแปลงที่เป็นการเปลี่ยนแปลงรูปแบบตลอดเวลาในช่วงเวลาใดเวลาหนึ่ง เรียกว่าฤดูกาลเปลี่ยนแปลง (changing seasonal) การคำนวณความผันแปรตามฤดูกาลจะคำนวณเป็นปริมาณสัมพัทธ์หรือร้อยละ ซึ่งเรียกค่าที่คำนวณได้ว่าดัชนีฤดูกาล

### การทดสอบความผันแปรตามฤดูกาล

การจะดูว่าข้อมูลอนุกรมมีความผันแปรตามฤดูกาลหรือไม่ นอกจากจะนำข้อมูลมาเขียนแผนภาพการกระจายแล้ว ยังสามารถนำข้อมูลอนุกรมเวลาที่ถูกกำจัดค่าแนวโน้มไปทดสอบว่ามีความแปรผันตามฤดูกาลหรือไม่ การทดสอบนั้นมีทั้งการทดสอบแบบใช้และแบบไม่ใช้พารามิเตอร์ แต่ในที่นี้จะใช้การทดสอบแบบไม่ใช้พารามิเตอร์ คือวิธีการทดสอบของครุสคัลและวัลลิส (Kruskal-Wallis Test)

การทดสอบของครุสคัลและวัลลิส นำมาประยุกต์ใช้กับการทดสอบว่าอนุกรมเวลาที่กำจัดค่าแนวโน้มแล้วมีความผันแปรตามฤดูกาลหรือไม่ มีขั้นตอนดังนี้

1. หาอนุกรมเวลาที่กำจัดแนวโน้มทั้งหมด แล้วเรียงลำดับข้อมูลดังกล่าวจากค่าต่ำสุด (ให้เป็นอันดับที่ 1) และเพิ่มขึ้นเรื่อย ๆ จนครบข้อมูลทุกตัว ซึ่งจะเท่ากับ  $n$  เมื่อ  $n$  คือจำนวนข้อมูลทั้งหมด

2. ตั้งสมมติฐานหลักและสมมติฐานทางเลือก ดังนี้

$H_0$  : อนุกรมเวลาไม่มีส่วนประกอบความผันแปรตามฤดูกาล

$H_1$  : อนุกรมเวลามีส่วนประกอบความผันแปรตามฤดูกาล

3. ตัวสถิติที่ใช้ทดสอบ

$$H = \frac{12}{n(n+1)} SS'_B$$

เมื่อ  $SS'_B$  แทนผลรวมกำลังสองของข้อมูลที่ใช้อันดับข้อมูลฤดูกาล

4. การตัดสินใจที่จะยอมรับหรือปฏิเสธสมมติฐานหลัก ( $H_0$ ) จะพิจารณาค่า  $p$  (p-value) ถ้า  $p < 0.05$  จะยอมรับสมมติฐาน  $H_0$  ปฏิเสธสมมติฐาน  $H_1$  นั่นคือข้อมูลไม่มีความแปรผันตามฤดูกาล แต่ถ้า  $p \geq 0.05$  จะปฏิเสธสมมติฐาน  $H_0$  ยอมรับสมมติฐาน  $H_1$  นั่นคือข้อมูลมีความแปรผันตามฤดูกาล

## 2.3 ตัวแบบอนุกรมเวลา (Model of Time Series)

การวิเคราะห์อนุกรมเวลาแบบแยกส่วน (Decomposition Time series Analysis) ตัวแบบที่นิยมใช้กันมี 2 ตัวแบบ คือ

### 2.3.1 ตัวแบบการบวก (Additive Model)

$$\text{กำหนดให้ } Y = T + S + C + I \quad (2.2)$$

โดยที่  $Y$  แทน อนุกรมเวลา

$T$  แทน แนวโน้ม (Trend)

$S$  แทน ฤดูกาล (Season)

$C$  แทน วัฏจักร (Cycle)

$I$  แทน ความไม่สม่ำเสมอของข้อมูล (Irregular Variation)

ตัวแบบการบวกของอนุกรมเวลาเกิดขึ้นจากแนวคิดที่ว่า ส่วนประกอบทั้งสี่ของอนุกรมเวลาจะต้องเป็นอิสระต่อกัน นั่นคือส่วนประกอบใดส่วนประกอบหนึ่งมีค่าเปลี่ยนแปลงไปไม่ว่าจะเพิ่มขึ้นหรือลดลง จะไม่มีผลกระทบต่อเปลี่ยนแปลงของส่วนประกอบอื่น ๆ ที่เหลือ

### 2.3.2 ตัวแบบการคูณ (Multiplicative model)

$$\text{กำหนดให้ } Y = T * S * C * I \quad (2.3)$$

ตัวแบบการคูณของอนุกรมเวลาเกิดขึ้นจากแนวคิดที่ว่า ส่วนประกอบทั้งสี่ของอนุกรมเวลามีความสัมพันธ์กัน หรือไม่เป็นอิสระต่อกัน นั่นคือส่วนประกอบใดส่วนประกอบหนึ่งมีค่าเปลี่ยนแปลงไปไม่ว่าจะเพิ่มขึ้นหรือลดลง จะมีผลกระทบต่อเปลี่ยนแปลงของส่วนประกอบอื่น ๆ ที่เหลือ ตัวแบบการคูณมักนิยมใช้กับข้อมูลทางเศรษฐศาสตร์ และธุรกิจ เพราะการเปลี่ยนแปลงปัจจัยปัจจัยหนึ่ง ในทางธุรกิจหรือเศรษฐศาสตร์ จะมีผลกระทบต่อปัจจัยอื่น ๆ ด้วย เช่น การศึกษาการเปลี่ยนแปลงของอัตราเงินกู้ในระบบธนาคาร จะส่งผลกระทบต่อปัจจัยอื่น ๆ อีกมากมายในทางธุรกิจ เป็นต้น ข้อสังเกตของการใช้ตัวแบบการคูณ คือ ค่าอนุกรมเวลาแต่ละค่าไม่จำเป็นต้อง

ประกอบด้วยสี่ส่วนเสมอไป ค่าบางค่าอาจจะประกอบด้วยแนวโน้มเพียงอย่างเดียว บางค่าอาจจะประกอบด้วยแนวโน้มและการแปรผันตามฤดูกาล เป็นต้น

## 2.4 การวิเคราะห์การถดถอยเชิงเส้น (Linear Regression Analysis)

การวิเคราะห์การถดถอย (Regression Analysis) เป็นวิธีการทางสถิติสำหรับวิเคราะห์รูปแบบความสัมพันธ์ระหว่างตัวแปร 2 ชุด โดยอาศัยข้อมูลที่เก็บรวบรวมไว้ในอดีต หรือข้อมูลจากการสำรวจ หรือข้อมูลจากการทดลอง (ธิดาเดียว, 2558) เมื่อทราบรูปแบบความสัมพันธ์ระหว่างตัวแปร 2 ชุดที่ทำการศึกษา จะทำการสร้างเป็นตัวแบบแสดงความสัมพันธ์ที่เรียกว่า ตัวแบบการถดถอย (Regression model) หรือฟังก์ชันการถดถอย (Regression function) ตัวแปร 2 ชุดนี้ ประกอบด้วย ตัวแปรชุดที่ 1 เรียกว่า ตัวแปรตาม (Dependent variable) หรือตัวแปรตอบสนอง (Response variable) แทนด้วยตัวแปร  $Y$  ชุดที่ 2 เรียกว่า ตัวแปรอิสระ (Independent variable) หรือตัวแปรทำนาย (Predictor) หรือตัวแปรถดถอย (Regression) แทนด้วยตัวแปร  $X$  ความสัมพันธ์ที่ได้จะเป็นความสัมพันธ์ที่อธิบายการเปลี่ยนแปลงของตัวแปรตามด้วยตัวแปรอิสระนั่นเอง ในงานวิจัยนี้จะใช้การวิเคราะห์ความสัมพันธ์ของตัวแปรในรูปเชิงเส้น โดยการวิเคราะห์การถดถอยเชิงเส้น แบ่งออกเป็น 2 ประเภทดังนี้

### 2.4.1 การถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression)

การวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปร 2 ตัว (ในที่นี้คือ ตัวแปร  $X$  และ  $Y$ ) ที่มีความสัมพันธ์ในลักษณะเชิงเส้น หรือเมื่อกราฟของความสัมพันธ์ระหว่างค่าเฉลี่ยของตัวแปรตาม  $Y$  และตัวแปรอิสระ  $X$  เป็นแบบเส้นตรง ตัวแบบการถดถอยเชิงเส้นอย่างง่ายหรือสมการถดถอยเชิงเส้นอย่างง่ายของข้อมูล จะมีรูปแบบดังนี้

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (2.4)$$

- เมื่อ  $Y_i$  คือ ตัวแปรตาม หรือตัวแปรตอบสนอง  
 $X_i$  คือ ตัวแปรอิสระ หรือตัวแปรทำนาย หรือตัวแปรถดถอย  
 $\beta_0$  คือ ระยะเวลาตัดแกน  $Y$  หรือค่าเฉลี่ยของ  $Y$  เมื่อ  $X = 0$  ซึ่งเป็นค่าพารามิเตอร์ของตัวแบบถดถอย  
 $\beta_1$  คือ ความชัน (Slope) ของเส้นถดถอย หรืออัตราการเปลี่ยนแปลงในค่าเฉลี่ยของ  $Y$  เมื่อ  $X$  เปลี่ยนไป 1 หน่วย ซึ่งเป็นค่าพารามิเตอร์ของตัวแบบการถดถอย  
 $\varepsilon_i$  คือ ความคลาดเคลื่อน (Error term) อันเนื่องมาจากความแตกต่างระหว่างค่า  $Y$  จริง กับค่า  $Y$  ที่อยู่บนเส้นถดถอย หรือความแตกต่างระหว่างค่า  $Y$  จริง กับค่า  $Y$  ที่ได้จากการประมาณด้วยตัวแบบการถดถอย นั่นเอง

วิธีประมาณค่าพารามิเตอร์  $\beta_0$  และ  $\beta_1$  หรือสัมประสิทธิ์การถดถอยที่นิยมใช้คือ วิธีกำลังสองน้อยที่สุดสามัญ (Ordinary Least Squares Method : OLS) หรือ วิธีกำลังสองน้อย

**ที่สุต** (Least Squares Method) โดยตัวประมาณค่าที่ได้จะทำให้ผลบวกของกำลังสองของความคลาดเคลื่อนระหว่างค่าสังเกตและเส้นถดถอยมีค่าน้อยที่สุด

การประมาณค่าสัมประสิทธิ์ การถดถอยจากตัวแบบการถดถอยของข้อมูลประชากร คือ  $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$  ตัวแบบการถดถอยของข้อมูลตัวอย่างจากการประมาณด้วยวิธีกำลังสองน้อยที่สุด กำหนดดังนี้

$$\hat{Y}_i = b_0 + b_1 X_i; i = 1, 2, \dots, n \quad (2.5)$$

เขียนรูปย่อ ๆ ได้ตั้งสมการต่อไปนี้

$$\hat{Y} = b_0 + b_1 X \quad (2.6)$$

เมื่อ  $\hat{Y}_i$  คือ ค่าประมาณของ  $Y_i$  ที่ได้จากตัวแบบการถดถอยเชิงเส้นอย่างง่าย  $b_0, b_1$  คือ ค่าประมาณของ  $\beta_0$  และ  $\beta_1$  ตามลำดับ

ผลต่างระหว่างค่าสังเกต  $Y_i$  และ  $\hat{Y}_i$  ที่ตรงกันเรียกว่า Residual หรือส่วนเกิน ซึ่งอาจเรียกว่าความคลาดเคลื่อน ค่า Residual ที่  $i$  เขียนแทนด้วย  $e_i$  นั่นคือ

$$e_i = Y_i - \hat{Y}_i = Y_i - (b_0 + b_1 X_i); i = 1, 2, \dots, n \quad (2.7)$$

โดย  $e_i$  เป็นค่าประมาณของ  $\varepsilon_i$  ตามสมการ (2.4) ซึ่งเป็นค่าสำคัญที่ใช้วิเคราะห์ว่าตัวแบบการถดถอยที่ได้มีความเหมาะสมกับข้อมูลหรือไม่

การหา  $b_0$  และ  $b_1$  ด้วยวิธีกำลังสองน้อยที่สุด จะทำให้ผลบวกของค่าคลาดเคลื่อนยกกำลังสอง  $\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n (Y_i - b_0 - b_1 X_i)^2$  มีค่าน้อยที่สุด ถ้ากำหนดให้

$$S = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n (Y_i - b_0 - b_1 X_i)^2 \quad (2.8)$$

$S$  เรียกว่าฟังก์ชันผลบวกกำลังสอง (Sum of squares function) การหาค่า  $b_0$  และ  $b_1$  จะทำโดยการหาค่าอนุพันธ์ย่อยของ  $S$  เทียบกับ  $\beta_0$  และ  $\beta_1$  ทีละค่าตามลำดับ แล้วแทน  $\beta_0$  ด้วย  $b_0$  และแทน  $\beta_1$  ด้วย  $b_1$  แล้วทำให้ผลลัพธ์เท่ากับ 0 จะได้สมการ 2 สมการ ดังนี้

$$\frac{\partial S}{\partial b_0} = -2 \sum_{i=1}^n (Y_i - b_0 - b_1 X_i) = 0$$

จะได้ 
$$-\sum_{i=1}^n Y_i + n b_0 + b_1 \sum_{i=1}^n X_i = 0$$

และ 
$$\frac{\partial S}{\partial b_1} = -2 \sum_{i=1}^n (Y_i - b_0 - b_1 X_i)(X_i) = 0$$

จะได้ 
$$-\sum_{i=1}^n X_i Y_i + b_0 \sum_{i=1}^n X_i + b_1 \sum_{i=1}^n X_i^2 = 0$$

จากสมการข้างต้น เขียนใหม่ได้เป็น

$$\begin{aligned} n b_0 + b_1 \sum_{i=1}^n X_i &= \sum_{i=1}^n Y_i \\ b_0 \sum_{i=1}^n X_i + b_1 \sum_{i=1}^n X_i^2 &= \sum_{i=1}^n X_i Y_i \end{aligned} \quad (2.9)$$

สมการ (2.9) เรียกว่าสมการปกติกำลังสองน้อยที่สุด (Least squares normal equations) แก้สมการเพื่อหาค่า  $b_0$  และ  $b_1$  ได้ดังนี้

$$\text{จากสมการ (2.9) จะได้} \quad b_0 = \bar{Y} - b_1\bar{X} \quad (2.10)$$

ซึ่ง  $b_0$  คือจุดที่เส้นตรงตัดแกน  $Y$  เมื่อ  $X = 0$  และ

$$b_1 = \frac{\sum_{i=1}^n X_i Y_i - \frac{(\sum_{i=1}^n X_i)(\sum_{i=1}^n Y_i)}{n}}{\sum_{i=1}^n X_i^2 - \frac{(\sum_{i=1}^n X_i)^2}{n}} \quad (2.11)$$

หรือจัดรูปใหม่ได้เป็น

$$b_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} \quad (2.12)$$

$$\text{เมื่อ } \bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{\sum_{i=1}^n X_i}{n} \quad \text{และ} \quad \bar{Y} = \frac{Y_1 + Y_2 + \dots + Y_n}{n} = \frac{\sum_{i=1}^n Y_i}{n}$$

ดังนั้น เมื่อแทนค่า  $b_0 = \bar{Y} - b_1\bar{X}$  จากสมการ (2.10) ในสมการ (2.6) จะได้ตัวแบบการถดถอยของข้อมูลตัวอย่างจากการประมาณด้วยวิธีกำลังสองน้อยที่สุด ดังนี้

$$\hat{Y} = \bar{Y} - b_1\bar{X} + b_1X = \bar{Y} + b_1(X - \bar{X}) \quad (2.13)$$

#### 2.4.2 การถดถอยเชิงเส้นแบบพหุ (Multiple Linear Regression)

การถดถอยเชิงเส้นแบบพหุ เป็นการศึกษาความสัมพันธ์ของตัวแปรตาม ที่เกิดจากตัวแปรอิสระตั้งแต่ 2 ตัวแปรขึ้นไป ถ้าตัวแปรอิสระมี  $k$  ตัว ( $X_1, X_2, \dots, X_k$ ) โดยความสัมพันธ์อยู่ในรูปเชิงเส้น จะได้ตัวแบบการถดถอยเชิงเส้นแบบพหุ หรือสมการถดถอยเชิงเส้นแบบพหุของข้อมูลประชากร ซึ่งแสดงความสัมพันธ์ระหว่างตัวแปร  $Y$  และ  $X_1, X_2, \dots, X_k$  ดังนี้

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i \quad (2.14)$$

หรือเขียนในรูปแบบย่อ ๆ คือ

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \quad (2.15)$$

เมื่อ  $Y_i$  คือ ค่าของตัวแปรตาม หรือตัวแปรตอบสนอง ตัวที่  $i$

$X_{ij}$  คือ ค่าสังเกตตัวที่  $i$  ของตัวแปรอิสระตัวที่  $j$  ;  $j = 1, 2, \dots, k$

$\beta_0$  คือ ระยะเวลาตัดแกน  $Y$  เมื่อกำหนดให้  $X_1 = X_2 = \dots = X_k = 0$

$\beta_1, \beta_2, \dots, \beta_k$  คือ สัมประสิทธิ์การถดถอยเชิงเส้นแบบพหุ โดย  $\beta_i$  เป็นค่าที่แสดงถึงการเปลี่ยนแปลงของตัวแปร  $Y$  เมื่อตัวแปรอิสระ  $X_i$  เปลี่ยนไป 1 หน่วย โดยที่ตัวแปรอิสระ  $X$  ตัวอื่น ๆ มีค่าคงที่

$\varepsilon_i$  คือ ค่าความคลาดเคลื่อนตัวที่  $i$  ;  $i = 1, 2, \dots, n$

การประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นแบบพหุจากตัวแบบการถดถอยของข้อมูลประชากร ซึ่งมีพารามิเตอร์  $k+1$  ตัว คือ  $\beta_0, \beta_1, \beta_2, \dots, \beta_k$  จะต้องใช้ข้อมูลตัวอย่าง ของตัวแปร  $Y, X_1, X_2, \dots, X_k$  โดยใช้ตัวอย่างขนาด  $n$  จาก ตัวแบบการถดถอยเชิงเส้นแบบพหุ

$$\text{จากสมการ (2.14)} \quad Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i$$

จะประมาณค่า  $Y$  หรือประมาณสมการ (2.14) ด้วยสมการ (2.16) ดังนี้

$$\hat{Y}_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + \dots + b_k X_{ik} \quad (2.16)$$

หรือ สมการ (2.16) เขียนในรูปแบบย่อ ๆ คือ

$$\hat{Y}_i = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_k X_k \quad (2.17)$$

ดังนั้น ค่าคลาดเคลื่อนในการประมาณค่า  $Y_i$  ด้วย  $\hat{Y}_i$  คือ  $e_i = Y_i - \hat{Y}_i$

ในการประมาณค่า  $\beta_0, \beta_1, \beta_2, \dots, \beta_k$  ด้วยค่า  $b_0, b_1, b_2, \dots, b_k$  ตามลำดับนั้น เช่นเดียวกันกับการถดถอยเชิงเส้นอย่างง่าย นั่นคือ หาค่า  $b_0, b_1, b_2, \dots, b_k$  โดยพิจารณาผลบวกของค่าคลาดเคลื่อนยกกำลังสอง  $\sum_{i=1}^n e_i^2$  ที่ทำให้  $\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$  มีค่าน้อยที่สุด จะหาสมการปกติได้ดังนี้

$$\text{ให้ } S = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n (Y_i - b_0 - \sum_{j=1}^k b_j X_{ij})^2 \quad (2.18)$$

ทำให้  $S$  มีค่าน้อยที่สุดโดยการหาอนุพันธ์ย่อยของ  $S$  เทียบกับ  $\beta_0, \beta_1, \dots, \beta_k$  ทีละค่าตามลำดับ แล้วแทนค่า  $\beta_0, \beta_1, \dots, \beta_k$  ด้วย  $b_0, b_1, b_2, \dots, b_k$  ตามลำดับ จากนั้นให้ผลลัพธ์ที่ได้แต่ละค่าเท่ากับ 0 จะได้สมการทั้งหมด  $k+1$  สมการ ดังนี้

$$\frac{\partial S}{\partial b_0} = -2 \sum_{i=1}^n (Y_i - b_0 - \sum_{j=1}^k b_j X_{ij}) = 0 \quad (2.19)$$

$$\text{และ} \quad \frac{\partial S}{\partial b_1} = -2 \sum_{i=1}^n (Y_i - b_0 - \sum_{j=1}^k b_j X_{ij}) = 0; j = 1, 2, \dots, k \quad (2.20)$$

นั่นคือ ถ้ามีตัวแปรอิสระ  $k$  ตัว จะได้สมการจำนวน  $k+1$  สมการ เมื่อจัดสมการใหม่ จะได้สมการปกติของวิธีกำลังสองน้อยที่สุด ดังนี้

$$\begin{aligned} nb_0 + b_1 \sum_{i=1}^n X_{i1} + b_2 \sum_{i=1}^n X_{i2} + \dots + b_k \sum_{i=1}^n X_{ik} &= \sum_{i=1}^n Y_i \\ 0 \sum_{i=1}^n X_{i1} + b_1 \sum_{i=1}^n X_{i1}^2 + b_2 \sum_{i=1}^n X_{i1} X_{i2} + \dots + b_k \sum_{i=1}^n X_{i1} X_{ik} &= \sum_{i=1}^n X_{i1} Y_i \\ &\vdots \\ b_0 \sum_{i=1}^n X_{ik} + b_1 \sum_{i=1}^n X_{ik} X_{i1} + b_2 \sum_{i=1}^n X_{ik} X_{i2} + \dots + b_k \sum_{i=1}^n X_{ik}^2 &= \sum_{i=1}^n X_{ik} Y_i \end{aligned} \quad (2.21)$$

ในสมการปกติทั้งหมด  $k+1$  สมการ จะสามารถหาค่า  $b_0, b_1, b_2, \dots, b_k$  ได้ด้วยวิธีพีชคณิตหรือด้วยวิธีการใช้เมทริกซ์ในการวิเคราะห์สมการ โดยพิจารณาตัวแบบการถดถอยเชิงเส้นแบบพหุจากสมการ (2.14) เขียนในรูปของเมทริกซ์ได้ ดังนี้

$$\underline{Y} = \underline{X}\underline{\beta} + \underline{\varepsilon} \quad (2.22)$$

$$\text{เมื่อ } \underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad \underline{X} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix},$$

$$\underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \underline{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}, \quad \underline{b} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{bmatrix}$$

โดยที่  $\underline{Y}$  คือ เวกเตอร์ของค่าสังเกตหรือตัวแปรตาม ขนาด  $n \times 1$

$\underline{X}$  คือ เมทริกซ์ของตัวแปรอิสระ ขนาด  $n \times (k + 1)$

$\underline{\beta}$  คือ เวกเตอร์ของค่าสัมประสิทธิ์การถดถอย ขนาด  $(k + 1) \times 1$

$\underline{b}$  คือ เวกเตอร์ของตัวประมาณของสัมประสิทธิ์การถดถอย ขนาด  $(k + 1) \times 1$

และ  $\underline{\varepsilon}$  คือ เวกเตอร์ของความคลาดเคลื่อน ขนาด  $n \times 1$

ต้องการหาตัวประมาณ  $b$  โดยวิธีกำลังสองน้อยที่สุด ความคลาดเคลื่อนของสมการถดถอยจึงเขียนได้เป็น

$$\underline{\varepsilon} = \underline{Y} - \underline{X}\underline{\beta} \quad (2.23)$$

และผลบวกกำลังสองของความคลาดเคลื่อน เขียนได้ดังนี้

$$\begin{aligned} S &= \sum \varepsilon^2 = \underline{\varepsilon}'\underline{\varepsilon} = (\underline{Y} - \underline{X}\underline{\beta})'(\underline{Y} - \underline{X}\underline{\beta}) \\ &= \underline{Y}'\underline{Y} - \underline{\beta}'\underline{X}'\underline{Y} - \underline{Y}'\underline{X}\underline{\beta} + \underline{\beta}'\underline{X}'\underline{X}\underline{\beta} \\ &= \underline{Y}'\underline{Y} - 2\underline{\beta}'\underline{X}'\underline{Y} + \underline{\beta}'\underline{X}'\underline{X}\underline{\beta} \end{aligned} \quad (2.24)$$

$$\text{ซึ่งจะมีค่าต่ำสุด เมื่อ } \frac{\partial \sum \varepsilon^2}{\partial \underline{\beta}} = -2\underline{X}'\underline{Y} + 2\underline{X}'\underline{X}\underline{\beta} = 0 \quad (2.25)$$

ประมาณค่าเวกเตอร์  $\underline{\beta}$  ด้วยเวกเตอร์  $\underline{b}$  จะได้สมการอยู่ในรูป

$$(\underline{X}'\underline{X})\underline{b} = \underline{X}'\underline{Y} \quad (2.26)$$

$$\text{ดังนั้น } \underline{b} = (\underline{X}'\underline{X})^{-1} \underline{X}'\underline{Y} \quad (2.27)$$

เมื่อ  $\underline{X}'$  คือ ทรานสโพสเมทริกซ์ของเมทริกซ์  $\underline{X}$

$\underline{X}'\underline{X}$  คือ เมทริกซ์จัตุรัสขนาด  $(k + 1) \times (k + 1)$

$\underline{X}'\underline{Y}$  คือ เวกเตอร์คอลัมน์ขนาด  $(k + 1) \times 1$

$(\underline{X}'\underline{X})^{-1}$  คือ เมทริกซ์ผกผันของ  $\underline{X}'\underline{X}$

ดังนั้น จะสามารถคำนวณค่า  $\underline{b}$  ได้ก็ต่อเมื่อ เมทริกซ์ผกผัน  $\underline{X}'\underline{X}$  สามารถหาค่าได้จริง นั่นคือตัวแปรอิสระทั้งหมดเป็นอิสระเชิงเส้นกัน (Linearly Independent) สมการ (2.26) สามารถเขียนให้ละเอียดได้ดังนี้

$$\underline{X}'\underline{X} \quad \underline{b} = \underline{X}'\underline{Y}$$

$$\begin{bmatrix} n & \sum_{i=1}^n X_{i1} & \sum_{i=1}^n X_{i2} & \cdots & \sum_{i=1}^n X_{ik} \\ \sum_{i=1}^n X_{i1} & \sum_{i=1}^n X_{i1}^2 & \sum_{i=1}^n X_{i1}X_{i2} & \cdots & \sum_{i=1}^n X_{i1}X_{ik} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n X_{ik} & \sum_{i=1}^n X_{ik}X_{i1} & \sum_{i=1}^n X_{ik}X_{i2} & \cdots & \sum_{i=1}^n X_{ik}^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_{i1}Y_i \\ \vdots \\ \sum_{i=1}^n X_{ik}Y_i \end{bmatrix}$$

จะสังเกตได้ว่า สมาชิกในแนวเส้นทแยงมุมหลักของเมทริกซ์  $\underline{X}'\underline{X}$  คือผลบวกกำลังสองของสมาชิกแต่ละคอลัมน์ในเมทริกซ์  $\underline{X}$  ส่วนสมาชิกที่อยู่นอกแนวเส้นทแยงมุมหลัก คือผลบวกของผลคูณระหว่างสมาชิกในแต่ละคอลัมน์ของเมทริกซ์  $\underline{X}$  ส่วนสมาชิกของเมทริกซ์  $\underline{X}'\underline{Y}$  คือผลบวกของผลคูณระหว่างสมาชิกในแต่ละคอลัมน์ของเมทริกซ์  $\underline{X}$  กับสมาชิกของเวกเตอร์  $\underline{Y}$

รูปแบบของตัวแบบการถดถอย คือ

$$\hat{Y} = \underline{X} \underline{b} = \underline{X} (\underline{X}'\underline{X})^{-1} \underline{X}' \underline{Y} = \underline{H} \underline{Y} \quad (2.28)$$

เมื่อ เวกเตอร์ของ  $\underline{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$  กำหนดโดย

$$\underline{e} = \underline{Y} - \hat{Y} = \underline{Y} - \underline{X} \underline{b} = (\underline{I} - \underline{H})\underline{Y} \quad (2.29)$$

เมื่อ  $\underline{H} = \underline{X} (\underline{X}'\underline{X})^{-1} \underline{X}'$  เรียกว่า เมทริกซ์แฮท (Hat matrix)

### 2.4.3 การวิเคราะห์การถดถอยสำหรับตัวแปรเชิงคุณภาพ (Regression Analysis for Qualitative Independent Variables)

การสร้างตัวแบบการถดถอยเพื่ออธิบายความสัมพันธ์ของตัวแปรอิสระกับตัวแปรตามนั้น บางครั้งตัวแปรอิสระที่นำมาวิเคราะห์การถดถอยอาจไม่ใช่ตัวแปรเชิงปริมาณ แต่อาจเป็นตัวแปรเชิงคุณภาพ (Qualitative Variables) หรือเรียกอีกอย่างหนึ่งว่าตัวแปรเชิงกลุ่ม (Categorical Variables) ซึ่งค่าของตัวแปรประเภทนี้จะแบ่งแยกออกเป็นกลุ่ม เช่น ตัวแปรเพศ แบ่งแยกเป็นชายและหญิง ตัวแปรสถานะภาพการสมรส แบ่งแยกเป็น โสด สมรส หม้าย หย่าร้าง และแยกกันอยู่ ตัวแปรอาชีพ แบ่งแยกเป็น รับราชการ รัฐวิสาหกิจ เอกชน และอาชีพอิสระ เป็นต้น ตัวแปรอิสระที่นำมาวิเคราะห์การถดถอยในคราวเดียวกันอาจจะมีทั้งตัวแปรเชิงปริมาณและตัวแปรเชิงคุณภาพ กรณีที่มีตัวแปรเชิงคุณภาพอาจมีตัวแปรเดียวหรือหลายตัวแปร การวิเคราะห์การถดถอยสามารถสร้างตัวแบบการถดถอยแสดงความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระเชิงคุณภาพได้เช่นกัน การนำตัวแปรเชิงคุณภาพมาวิเคราะห์การถดถอยจะต้องกำหนดตัวเลขให้กับแต่ละกลุ่ม ซึ่งตัวเลขที่กำหนดขึ้นมานั้นไม่มีความหมายในเชิงตัวเลข ตัวเลขนี้จะเป็นค่าตัวแปรอิสระที่กำหนด

ขึ้นมาใหม่ เรียกตัวแปรอิสระนี้ว่า **ตัวแปรหุ่น** (Dummy Variables) ดังนั้นตัวแปรหุ่นจึงเป็นตัวแปรอิสระที่ไม่มีอยู่จริงตั้งแต่ต้น แต่ถูกกำหนดขึ้นมาใหม่เพื่อระบุกลุ่มให้กับตัวแปรเชิงคุณภาพ (ธิตา เดียว, 2559)

การกำหนดค่าของตัวแปรหุ่นจะกำหนดเป็นรหัสตัวเลขที่เรียงกัน หากตัวแปรเชิงคุณภาพถูกแบ่งเป็น  $m$  กลุ่ม จะใช้ตัวแปรหุ่นจำนวน  $m-1$  ตัวแปร เพื่อกำหนดรหัสตัวเลขให้กับกลุ่มของตัวแปรเชิงคุณภาพ เช่น ตัวแปรเพศแบ่งเป็น 2 กลุ่ม จะใช้ตัวแปรหุ่น 1 ตัวแปรแทนตัวแปรเพศ ค่าของตัวแปรหุ่นนี้อาจกำหนดรหัสเป็น 1 แทนกลุ่มเพศชายและกำหนดรหัสเป็น 0 แทนกลุ่มเพศหญิง หรืออาจกำหนดรหัสตัวเลขให้กับกลุ่มสลับกัน ตัวอย่างของตัวแปรเชิงคุณภาพ การกำหนดตัวแปรหุ่นและค่าของตัวแปรหุ่นแสดงดังในตาราง 2.2

ตาราง 2.2 ตัวอย่างตัวแปรเชิงคุณภาพ การกำหนดตัวแปรหุ่นและค่าของตัวแปรหุ่น

| ตัวแปรเชิงคุณภาพ  | จำนวนตัวแปรหุ่น | การกำหนดค่าของตัวแปรหุ่น  |
|---|-----------------|---|
| เพศ แบ่งเป็น 2 กลุ่ม<br>(ชาย, หญิง)   | 1               | $X = 1$ ; ชาย<br>$= 0$ ; หญิง   |
| ระดับการศึกษา แบ่งเป็น 3 กลุ่ม<br>(ต่ำกว่าปริญญาตรี, ปริญญาตรี, สูงกว่าปริญญาตรี)                                 | 2               | $X_1 = 1$ ; ต่ำกว่าปริญญาตรี<br>$= 0$ ; ปริญญาตรี หรือสูงกว่า         |
|   |                 | $X_2 = 1$ ; ปริญญาตรี<br>$= 0$ ; ต่ำกว่าปริญญาตรี หรือสูงกว่า         |
| สถานะภาพการทำงาน แบ่งเป็น 4 กลุ่ม<br>(ไม่เคยทำงาน, ไม่มีงานทำไม่เกิน 3 เดือน, ไม่มีงานทำมากกว่า 3 เดือน, มีงานทำ) | 3               | $X_1 = 1$ ; ไม่เคยทำงาน<br>$= 0$ ; สำหรับสถานะภาพอื่น ๆ               |
|   |                 | $X_2 = 1$ ; ไม่มีงานทำไม่เกิน 3 เดือน<br>$= 0$ ; สำหรับสถานะภาพอื่น ๆ |
|   |                 | $X_3 = 1$ ; ไม่มีงานทำมากกว่า 3 เดือน<br>$= 0$ ; สำหรับสถานะภาพอื่น ๆ |
| สถานะภาพการสมรส แบ่งเป็น 5 กลุ่ม<br>(โสด, สมรส, หม้าย, หย่าร้าง, แยกกันอยู่)                                      | 4               | $X_1 = 1$ ; โสด<br>$= 0$ ; สำหรับสถานะภาพอื่น ๆ                       |
|   |                 | $X_2 = 1$ ; สมรส  |

|  |  |                            |
|--|--|----------------------------|
|  |  | = 0 ; สำหรับสถานะภาพอื่น ๆ |
|  |  | $X_3 = 1$ ; หม้าย          |
|  |  | = 0 ; สำหรับสถานะภาพอื่น ๆ |
|  |  | $X_4 = 1$ ; หย่าร้าง       |
|  |  | = 0 ; สำหรับสถานะภาพอื่น ๆ |

จากตาราง 2.2 จะพบว่าตัวแปรเชิงคุณภาพที่แบ่งเป็น 2 กลุ่ม จะต้องการตัวแปรหุ่น 1 ตัวแปรในการกำหนดรหัสให้กับกลุ่ม คือ  $X$  ส่วนตัวแปรเชิงคุณภาพที่แบ่งเป็น 3 กลุ่ม จะต้องการตัวแปรหุ่น 2 ตัวแปรในการกำหนดรหัส คือ  $X_1$  และ  $X_2$  เช่นตัวแปรระดับการศึกษา เมื่อนำ  $X_1$  และ  $X_2$  มากำหนดรหัสรวมกันจะได้ดังนี้

| กลุ่มที่ | $X_1$ | $X_2$ |  |
|----------|-------|-------|--|
| 1        | 1     | 0     | เมื่อตัวอย่างมีระดับการศึกษาต่ำกว่าปริญญาตรี |
| 2        | 0     | 1     | เมื่อตัวอย่างมีระดับการศึกษาปริญญาตรี        |
| 3        | 0     | 0     | เมื่อตัวอย่างมีระดับการศึกษาสูงกว่าปริญญาตรี |

ส่วนกรณีตัวแปรเชิงคุณภาพที่แบ่งเป็น 4 กลุ่ม จะต้องการตัวแปรหุ่น 3 ตัวแปรในการกำหนดรหัส คือ  $X_1, X_2$  และ  $X_3$  เช่นตัวแปรสถานะภาพการทำงาน เมื่อนำ  $X_1, X_2$  และ  $X_3$  มากำหนดรหัสรวมกันจะได้ดังนี้

| กลุ่มที่ | $X_1$ | $X_2$ | $X_3$ |  |
|----------|-------|-------|-------|--|
| 1        | 1     | 0     | 0     | เมื่อตัวอย่างไม่เคยทำงาน               |
| 2        | 0     | 1     | 0     | เมื่อตัวอย่างไม่มีงานทำไม่เกิน 3 เดือน |
| 3        | 0     | 0     | 1     | เมื่อตัวอย่างไม่มีงานทำมากกว่า 3 เดือน |
| 4        | 0     | 0     | 0     | เมื่อตัวอย่างมีงานทำ                   |

ตัวอย่างการวิเคราะห์การถดถอยของข้อมูลที่มีตัวแปรอิสระเชิงคุณภาพ 1 ตัวแปร เมื่อมีการสร้างตัวแปรหุ่นเพื่อกำหนดรหัสให้กับกลุ่มของตัวแปรอิสระเชิงคุณภาพแล้ว การนำตัวแปรหุ่นเข้าสู่การวิเคราะห์การถดถอยด้วยวิธีกำลังสองน้อยสุด ซึ่งมีตัวแปรอิสระเชิงปริมาณอื่นร่วมด้วย สมมุติว่าในการวิเคราะห์การถดถอยเชิงเส้นแบบพหุระหว่างตัวแปรตาม  $Y$  กับตัวแปรอิสระ 2 ตัว โดยเป็นตัวแปรอิสระเชิงปริมาณ 1 ตัวแปร และเป็นตัวแปรอิสระเชิงคุณภาพ 1 ตัวแปรที่แบ่งเป็น 2 กลุ่ม สมมุติให้ตัวแปรอิสระเชิงปริมาณคือ  $X_1$  และตัวแปรหุ่นคือ  $X_2$  โดย  $X_2$  กำหนดเป็นรหัส 0 ให้กับกลุ่มที่ 1 และกำหนดรหัสเป็น 1 ให้กับกลุ่มที่ 2 ของตัวแปรเชิงคุณภาพ ตัวแบบการถดถอยเชิงเส้นแบบพหุของประชากร เป็นดังนี้

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \quad (2.30)$$

สามารถแยกเป็นตัวแบบการถดถอยสำหรับแต่ละกลุ่มได้ ดังนี้

**กลุ่มที่ 1** ( $X_2 = 0$ ) ตัวแบบการถดถอยตามสมการ (2.30) เปลี่ยนเป็นดังนี้

$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1 + \beta_2 (0) + \varepsilon \\ &= \beta_0 + \beta_1 X_1 + \varepsilon \end{aligned} \quad (2.31)$$

พบว่าตัวแบบการถดถอยสำหรับกลุ่มที่ 1 มีจุดตัดแกน  $Y$  เท่ากับ  $\beta_0$  และมีความชันเท่ากับ  $\beta_1$

**กลุ่มที่ 2** ( $X_2 = 1$ ) ตัวแบบการถดถอยตามสมการ (2.30) เปลี่ยนเป็นดังนี้

$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1 + \beta_2 (1) + \varepsilon \\ &= (\beta_0 + \beta_2) + \beta_1 X_1 + \varepsilon \end{aligned} \quad (2.32)$$

พบว่าตัวแบบการถดถอยสำหรับกลุ่มที่ 2 มีจุดตัดแกน  $Y$  เท่ากับ  $\beta_0 + \beta_2$  และมีความชันเท่ากับ  $\beta_1$

เส้นถดถอยสองเส้นของตัวแบบการถดถอยในสมการ (2.31) และ (2.32) จะขนานกัน เนื่องจากมีความชันเท่ากันคือ  $\beta_1$  แต่มีจุดตัดแกน  $Y$  ต่างกัน โดยเส้นถดถอยเส้นหนึ่งจะอยู่สูงกว่าเส้นถดถอยอีกเส้นหนึ่งเท่ากับ  $\beta_2$

หากตัวแปรอิสระเชิงคุณภาพแบ่งเป็น 3 กลุ่ม ตามที่แสดงในตาราง 2.2 ตัวแปรเชิงคุณภาพ จะต้องการตัวแปรหุ่น 2 ตัวแปร เพื่อกำหนดรหัสให้กับกลุ่มตัวแปรเชิงคุณภาพ สมมติว่าในการวิเคราะห์การถดถอยเชิงเส้นแบบพหุระหว่างตัวแปร  $Y$  กับตัวแปรอิสระเชิงปริมาณคือ  $X_1$  และตัวแปรหุ่นคือ  $X_2$  และ  $X_3$  เมื่อนำ  $X_2$  และ  $X_3$  มากำหนดรหัสรวมกันจะได้ดังนี้

| แบบที่ 1 | กลุ่มที่ | $X_2$ | $X_3$ | แบบที่ 2 | กลุ่มที่ | $X_2$ | $X_3$ | แบบที่ 3 | กลุ่มที่ | $X_2$ | $X_3$ |
|----------|----------|-------|-------|----------|----------|-------|-------|----------|----------|-------|-------|
|          | 1        | 1     | 0     |          | 1        | 0     | 1     |          | 1        | 0     | 0     |
|          | 2        | 0     | 1     |          | 2        | 1     | 0     |          | 2        | 1     | 0     |
|          | 3        | 0     | 0     |          | 3        | 0     | 0     |          | 3        | 0     | 1     |

ตัวแบบการถดถอยเชิงเส้นแบบพหุของประชากร เป็นดังนี้

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon \quad (2.33)$$

หากกำหนดรหัสของตัวแปรหุ่นตามแบบที่ 1 ตัวแบบการถดถอยสำหรับแต่ละกลุ่มจะเป็นดังนี้

**กลุ่มที่ 1** ( $X_2 = 1, X_3 = 0$ ) ตัวแบบการถดถอยตามสมการ (2.33) เปลี่ยนเป็นดังนี้

$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1 + \beta_2 (1) + \beta_3 (0) + \varepsilon \\ &= (\beta_0 + \beta_2) + \beta_1 X_1 + \varepsilon \end{aligned} \quad (2.34)$$

พบว่าตัวแบบการถดถอยสำหรับกลุ่มที่ 1 มีจุดตัดแกน  $Y$  เท่ากับ  $\beta_0 + \beta_2$  และมีความชันเท่ากับ  $\beta_1$

กลุ่มที่ 2 ( $X_2 = 0, X_3 = 1$ ) ตัวแบบการถดถอยตามสมการ (2.33) เปลี่ยนเป็นดังนี้

$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1 + \beta_2 (0) + \beta_3 (1) + \varepsilon \\ &= (\beta_0 + \beta_3) + \beta_1 X_1 + \varepsilon \end{aligned} \quad (2.35)$$

พบว่าตัวแบบการถดถอยสำหรับกลุ่มที่ 2 มีจุดตัดแกน  $Y$  เท่ากับ  $\beta_0 + \beta_3$  และมีความชันเท่ากับ  $\beta_1$

กลุ่มที่ 3 ( $X_2 = 0, X_3 = 0$ ) ตัวแบบการถดถอยตามสมการ (2.33) เปลี่ยนเป็นดังนี้

$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1 + \beta_2 (0) + \beta_3 (0) + \varepsilon \\ &= \beta_0 + \beta_1 X_1 + \varepsilon \end{aligned} \quad (2.36)$$

พบว่าตัวแบบการถดถอยสำหรับกลุ่มที่ 3 มีจุดตัดแกน  $Y$  เท่ากับ  $\beta_0$  และมีความชันเท่ากับ  $\beta_1$

เส้นถดถอยทั้ง 3 เส้นของตัวแบบการถดถอยในสมการ (2.34) ถึง (2.36) จะขนานกัน เนื่องจากมีความชันเท่ากันคือ  $\beta_1$  แต่เส้นถดถอยทั้ง 3 เส้นนี้จะมีจุดตัดแกน  $Y$  ต่างกัน โดยเส้นถดถอยของตัวแบบตามสมการ (2.34) จะห่างจากเส้นถดถอยของตัวแบบตามสมการ (2.36) เท่ากับ  $\beta_2$  และเส้นถดถอยของตัวแบบตามสมการ (2.35) จะห่างจากเส้นถดถอยของตัวแบบตามสมการ (2.36) เท่ากับ  $\beta_3$

ในการศึกษาครั้งนี้เป็นการหาตัวแบบการถดถอยที่อธิบายความสัมพันธ์ของยอดขายรถยนต์ที่มีอิทธิพลของฤดูกาลร่วมด้วย ซึ่งเป็นตัวแปรเชิงคุณภาพ จึงมีการกำหนดรหัสสำหรับตัวแปรกลุ่มนี้โดยชุดตัวแปรเชิงคุณภาพฤดูกาล แบ่งเป็น 12 กลุ่ม (ตามรายเดือน) และเมื่อเพิ่มชุดตัวแปรของสถานการณ์พิเศษอีก 4 กลุ่ม รวมเป็น 16 กลุ่ม ก็จะมีจำนวนตัวแปรหุ่นรวมเป็น 15 ตัวแปร

## 2.5 การพิจารณาความสัมพันธ์และการเปรียบเทียบค่าพยากรณ์

### 2.5.1 การวิเคราะห์สหสัมพันธ์ (Correlation Analysis)

สหสัมพันธ์ (Correlation) เป็นสถิติที่ใช้หาความสัมพันธ์ระหว่างตัวแปรสองกลุ่ม ซึ่งค่าสหสัมพันธ์ที่คำนวณได้ เรียกว่า ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation coefficient) ซึ่งสถิติสำหรับการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์มีหลายชนิด การเลือกใช้แบบใดนั้นขึ้นอยู่กับเงื่อนไขหลายประการ ในการวัดความสัมพันธ์แต่ละแบบจะต้องมีการทดสอบนัยสำคัญก่อนจึงจะสรุปได้ว่าตัวแปรคู่ใดมีความสัมพันธ์กันจริงหรือไม่ และมีความสัมพันธ์กันมากน้อยเพียงใด ค่าสัมประสิทธิ์สหสัมพันธ์ที่ใช้กันอยู่แพร่หลายได้แก่ ค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน (Pearson's Correlation Coefficient)

การคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน หรือบางครั้งเรียกว่า สหสัมพันธ์อย่างง่าย (Simple Correlation) โดยใช้สัญลักษณ์  $r$  แทน สัมประสิทธิ์สหสัมพันธ์ของกลุ่มตัวอย่าง ค่าสัมประสิทธิ์สหสัมพันธ์ที่ใช้วัดขนาดของความสัมพันธ์ระหว่างตัวแปรมีค่าอยู่ระหว่าง -1 ถึง 1 (กัลยา, 2545) การบอกระดับความสัมพันธ์ สามารถพิจารณาได้ดังนี้

ตาราง 2.3 ระดับความสัมพันธ์ของข้อมูลจากค่าสัมประสิทธิ์สหสัมพันธ์

| ค่า $r$                   | ระดับความสัมพันธ์               |
|---------------------------|---------------------------------|
| $\pm 0.81$ ถึง $\pm 1.00$ | มีความสัมพันธ์กันสูงมาก         |
| $\pm 0.61$ ถึง $\pm 0.80$ | มีความสัมพันธ์กันในระดับสูง     |
| $\pm 0.41$ ถึง $\pm 0.60$ | มีความสัมพันธ์กันในระดับปานกลาง |
| $\pm 0.21$ ถึง $\pm 0.40$ | มีความสัมพันธ์กันในระดับต่ำ     |
| $\pm 0.00$ ถึง $\pm 0.20$ | มีความสัมพันธ์กันในระดับต่ำมาก  |

ค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันสามารถคำนวณได้ดังนี้

$$r = \frac{n \sum_{i=1}^n X_i Y_i - (\sum_{i=1}^n X_i)(\sum_{i=1}^n Y_i)}{\sqrt{(n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2)(n \sum_{i=1}^n Y_i^2 - (\sum_{i=1}^n Y_i)^2)}} \quad (2.37)$$

เมื่อ  $r$  แทน สัมประสิทธิ์สหสัมพันธ์  
 $\sum_{i=1}^n X_i$  แทน ผลรวมของตัวแปร  $X$   
 $\sum_{i=1}^n Y_i$  แทน ผลรวมของตัวแปร  $Y$   
 $\sum_{i=1}^n X_i^2$  แทน ผลรวมของกำลังสองของตัวแปร  $X$   
 $\sum_{i=1}^n Y_i^2$  แทน ผลรวมของกำลังสองของตัวแปร  $Y$   
 $\sum_{i=1}^n X_i Y_i$  แทน ผลรวมของผลคูณของตัวแปร  $X$  และ  $Y$   
 $n$  แทน จำนวนกลุ่มตัวอย่าง

ค่าสัมประสิทธิ์สหสัมพันธ์จะใช้ได้อย่างเหมาะสมกับข้อมูลที่มีความสัมพันธ์เชิงเส้นเครื่องหมาย + และ - หน้าค่าสัมประสิทธิ์สหสัมพันธ์แสดงถึงทิศทางของความสัมพันธ์ หากค่า  $r$  เป็นลบ แสดงว่า  $X$  และ  $Y$  มีความสัมพันธ์ในทิศทางตรงกันข้าม คือ ถ้า  $X$  มีค่าเพิ่มขึ้น  $Y$  จะมีค่าลดลง ในทางตรงกันข้าม หาก  $X$  มีค่าลดลง  $Y$  จะมีค่าเพิ่มขึ้น ค่า  $r$  เป็นบวก แสดงว่า  $X$  และ  $Y$  มีความสัมพันธ์ในทิศทางเดียวกัน นั่นคือ ถ้า  $X$  มีค่าเพิ่มขึ้น  $Y$  จะมีค่าเพิ่มขึ้น และ หาก  $X$  มีค่าลดลง  $Y$  จะมีค่าลดลง ถ้าค่า  $r$  เข้าใกล้ 1 แสดงว่า  $X$  และ  $Y$  มีความสัมพันธ์ในทิศทางเดียวกัน และมีความสัมพันธ์กันมาก ถ้าค่า  $r$  เข้าใกล้ -1 แสดงว่า  $X$  และ  $Y$  มีความสัมพันธ์ในทิศทางตรงกันข้าม และมีความสัมพันธ์กันมาก แต่ถ้าค่า  $r$  เท่ากับ 0 แสดงว่า  $X$  และ  $Y$  ไม่มี ความสัมพันธ์ในลักษณะเชิงเส้นต่อกัน

ปัจจัยที่ใช้ในการศึกษาครั้งนี้ประกอบด้วย 2 ปัจจัยหลัก ได้แก่ ผลิตภัณฑ์มวลรวมในประเทศ (Gross Domestic Products : GDP) และอัตราดอกเบี้ยเงินกู้ 12 เดือน (12-month Loan's Interest Rate)

## 2.5.2 การประเมินค่าความผิดพลาดจากการพยากรณ์

การประเมินประสิทธิภาพของการพยากรณ์ เป็นการวัดค่าความถูกต้องของการพยากรณ์ที่วัดจากขนาดของความคลาดเคลื่อนของการพยากรณ์เทียบกับค่าจริง ความถูกต้องของการพยากรณ์จะมีมากหรือน้อยขึ้นอยู่กับว่าค่าพยากรณ์ใกล้เคียงกับค่าจริงมากน้อยเพียงใด หรือค่าคลาดเคลื่อนของการพยากรณ์มีมากน้อยเพียงใด (ทรงศิริ, 2549) ค่าวัดความถูกต้องนี้เป็นค่าที่ไม่มีหน่วยจึงเหมาะที่จะใช้ในการเปรียบเทียบอนุกรมเวลาหลายชุดที่อาจจะมิหนวยของการวัดต่างกันเมื่อใช้วิธีการพยากรณ์เดียวกัน เนื่องจากการพยากรณ์อนุกรมเวลานั้นมีหลายวิธี โดยแต่ละวิธีมีรูปแบบหรือแบบจำลองทางคณิตศาสตร์เพื่อพยากรณ์อนุกรมเวลาที่แตกต่างกันออกไป กำหนดให้  $Y_i$  คือ ค่าที่แท้จริงของค่าสังเกตตัวที่  $i$  ในอนุกรมเวลา  $Y_i$  คือ ค่าที่ของตัวแปรพยากรณ์จากค่าสังเกตตัวที่  $i$  และ  $n$  แทนจำนวนชุดข้อมูล วิธีการตรวจสอบความแม่นยำของการพยากรณ์ที่นำมาใช้ในงานวิจัย มีดังนี้

**ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean Square Error : MSE)** เป็นวิธีวัดความแม่นยำโดยแก้ปัญหาวิธีค่าเฉลี่ยความคลาดเคลื่อน โดยพิจารณาความแตกต่างระหว่างค่าจริงกับค่าพยากรณ์โดยวิธีกำลังสอง โดยที่

$$MSE = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n} \quad (2.38)$$

**ค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์ (Mean Absolute Percent Error : MAPE)** เป็นวิธีวัดค่าความแม่นยำโดยคำนวณเปอร์เซ็นต์ความผิดพลาดในการพยากรณ์ โดยไม่คำนึงถึงเครื่องหมาย ค่าวัดความถูกต้องนี้เป็นค่าที่ไม่มีหน่วยจึงเหมาะที่จะใช้ในการเปรียบเทียบอนุกรมเวลาหลายชุดเมื่อใช้วิธีการพยากรณ์เดียวกัน หรือเปรียบเทียบวิธีการพยากรณ์หลายวิธีเมื่อใช้ข้อมูลอนุกรมเวลาชุดเดียวกัน ถ้าค่า MAPE คำนวณค่าได้ต่ำหมายถึงตัวแบบพยากรณ์ความแม่นยำสูง เช่น MAPE = 5% แสดงว่าตัวแบบพยากรณ์ที่ใช้มีความคลาดเคลื่อนในการพยากรณ์เฉลี่ยอยู่ที่ร้อยละ 5 คำนวณได้จากสูตรดังนี้

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \times 100 \quad (2.39)$$

**รากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Root Mean Square Error : RMSE)** เป็นวิธีวัดค่าความแม่นยำโดยคำนวณความผิดพลาดในการพยากรณ์ จากสูตร

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}} \quad (2.40)$$

## 2.5.3 สัมประสิทธิ์การตัดสินใจ (Coefficient of Determination : $R^2$ )

สัมประสิทธิ์การตัดสินใจ หมายถึง สัดส่วนที่ตัวแปร  $X$  สามารถอธิบายการเปลี่ยนแปลงของตัวแปร  $Y$  ได้มากหรือน้อยเพียงใด หรืออาจกล่าวอีกนัยหนึ่งได้ว่า การเปลี่ยนแปลงของตัวแปรตามเป็นผลมาจากอิทธิพลของตัวแปรอิสระมากน้อยเพียงใด ตัวแบบการถดถอยที่ได้จะ

สามารถพยากรณ์การเปลี่ยนแปลงของตัวแปรตามได้ดีหรือไม่ ขึ้นอยู่กับค่าสัมประสิทธิ์การตัดสินใจ ซึ่งเกิดจากค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation coefficient) ยกกำลังสอง ใช้ตัวย่อ  $R^2$  ดังนั้น ถ้า  $R^2$  มีค่ามาก แสดงว่า  $X$  และ  $Y$  มีความสัมพันธ์กันมาก หรือ  $X$  สามารถอธิบายการเปลี่ยนแปลงของค่า  $Y$  ได้มาก ถ้า  $R^2$  มีค่าน้อย แสดงว่า  $X$  และ  $Y$  มีความสัมพันธ์กันน้อย หรือ  $X$  สามารถอธิบายการเปลี่ยนแปลงของค่า  $Y$  ได้น้อยนั่นเอง โดยที่  $R^2$  จะไม่มีหน่วย และ  $0 \leq R^2 \leq 1$  สูตรที่ใช้ในการหาค่าสัมประสิทธิ์การตัดสินใจ ( $R^2$ ) คือ

$$R^2 = \frac{(\sum_{i=1}^n X_i Y_i - n\bar{X}\bar{Y})^2}{(\sum_{i=1}^n X_i^2 - n\bar{X}^2)(\sum_{i=1}^n Y_i^2 - n\bar{Y}^2)} \quad (2.41)$$

หรือ

$$R^2 = \frac{(n\sum_{i=1}^n X_i Y_i - (\sum_{i=1}^n X_i)(\sum_{i=1}^n Y_i))^2}{(n\sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2)(n\sum_{i=1}^n Y_i^2 - (\sum_{i=1}^n Y_i)^2)} \quad (2.42)$$

**ความหมายของค่า  $R^2$**

- $R^2$  มีค่าเข้าใกล้ 1 หมายถึง เปอร์เซนต์ที่  $X$  สามารถอธิบายการเปลี่ยนแปลงของ  $Y$  ได้มาก หรือ  $X$  และ  $Y$  มีความสัมพันธ์กันมาก
- $R^2$  มีค่าเข้าใกล้ 0 หมายถึง เปอร์เซนต์ที่  $X$  สามารถอธิบายการเปลี่ยนแปลงของ  $Y$  ได้น้อย หรือ  $X$  และ  $Y$  มีความสัมพันธ์กันน้อย
- $R^2$  มีค่าอยู่ระหว่าง เปอร์เซนต์ที่ 0.7 และ 0.3 หมายถึง  $X$  สามารถอธิบายการเปลี่ยนแปลงของ  $Y$  ได้ปานกลาง หรือ  $X$  และ  $Y$  มีความสัมพันธ์กันปานกลาง

## 2.6 ทบทวนวรรณกรรม

จากงานวิจัยของ Pindyck และ Rotemberg (1983), Hamilton (1988) รวมทั้งงานวิจัยของ Barber และคณะ (1999) พบว่าราคาน้ำมันที่เพิ่มขึ้นอย่างมีนัยสำคัญ ส่งผลกระทบต่อยอดขายรถยนต์ อีกทั้งยังเป็นปัจจัยสำคัญที่ส่งผลกระทบต่อต้นทุนในการผลิตรถยนต์ มูลค่าการลงทุน และความต้องการของผู้บริโภคด้วย

Ludvigson (1998) ได้ทำการศึกษาผลกระทบจากนโยบายการบริหารด้านการเงินของรัฐบาล ที่มีนโยบายการเพิ่มอัตราดอกเบี้ย พบว่าการเพิ่มขึ้นของอัตราดอกเบี้ยมีผลกระทบในทางลบต่อยอดขายรถยนต์

ในงานวิจัยของ Dargay และ Gately (1999) ได้ทำการศึกษาการเป็นเจ้าของรถยนต์ของประชากรของ 26 ประเทศ ในกลุ่มองค์กรเพื่อความร่วมมือทางเศรษฐกิจและการพัฒนา (OECD) พบว่า อัตราการเป็นเจ้าของรถยนต์สำหรับประเทศที่มีรายได้ต่ำนั้น สูงกว่าที่คาดการณ์ไว้ ในขณะที่ประเทศจีน อินเดีย และปากีสถาน มีสัดส่วนการถือครองรถยนต์เพิ่มขึ้นตามการเพิ่มขึ้นของรายได้ของประชากรต่อหัว

Dargay (2001) ได้ทำการสำรวจค่าใช้จ่ายในครัวเรือน พบว่าสถิติการเป็นเจ้าของรถยนต์มีแนวโน้มสูงขึ้นเมื่อครอบครัวมีรายได้เพิ่มขึ้น และมีความสัมพันธ์เชิงลบ นั่นคือมีแนวโน้มการเป็นเจ้าของรถยนต์ลดลงเมื่อมีการลดลงของรายได้

T. Apiwattanachai และ J. Pichitlamken (2008) ได้นำเสนอตัวแบบการพยากรณ์ที่นำเอารูปแบบการพยากรณ์ของ Holt-Winter's มาปรับใช้ เพื่อประมาณความต้องการของลูกค้าที่ต้องการซื้อรถยนต์ แทนการใช้ข้อมูลการขายในอดีต เป็นการเพิ่มประสิทธิภาพในการจัดการสินค้าคงคลังของตัวแทนจำหน่ายรถยนต์ เพื่อลดปัญหาการกักตุนสินค้าและสินค้าตกค้างในคลังเก็บสินค้า

Shahabudin (2009) ได้นำเสนอตัวแบบการถดถอยเพื่ออธิบายยอดขายรถยนต์ ที่ได้รับอิทธิพลจากตัวแปรต่าง ๆ ได้แก่ ระดับรายได้ของประชากร อัตราดอกเบี้ยเงินกู้ ยอดรวมทางการเงิน และอัตราการว่างงาน พบว่าตัวแปรทั้งหมดนี้ส่งผลกระทบต่อยอดขายรถยนต์อย่างมีนัยสำคัญ

W. Chaimongkol และ C. Tansathit (2010) ได้พัฒนาตัวแบบการพยากรณ์ความต้องการรถยนต์ของลูกค้า โดยปรับปรุงตัวแบบการพยากรณ์ของ Holt-Winter's เพื่อประเมินความต้องการรถยนต์ของลูกค้า โดยการเปรียบเทียบค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์ (MAPE) พบว่าตัวแบบการพยากรณ์ที่ปรับปรุงขึ้นใหม่นั้นมีค่า MAPE น้อยกว่าตัวแบบเดิม

Wirotheewan P. และคณะ (2011) ได้ทำการศึกษาตัวแบบการพยากรณ์การส่งออกยานยนต์ที่ผลิตในประเทศไทยไปยังตลาดต่างประเทศที่มีความถูกต้องและแม่นยำ โดยใช้รูปแบบโปรแกรมเชิงเส้น และการวิเคราะห์รูปแบบการพยากรณ์เพื่อให้ได้ผลกำไรสูงสุด

ในงานวิจัยของ Muhammad และคณะ (2012) พบว่า นอกจากผลิตภัณฑ์มวลรวมในประเทศ (GDP) ที่มีความสัมพันธ์เชิงบวกกับยอดขายรถยนต์ใน 5 ประเทศ ได้แก่ มาเลเซีย อินโดนีเซีย สิงคโปร์ ฟิลิปปินส์ และประเทศไทย แล้วยังมีปัจจัยของอัตราเงินเฟ้อ อัตราการว่างงาน และอัตราดอกเบี้ยเงินกู้ ที่เป็นปัจจัยที่ส่งผลกระทบต่อยอดขายรถยนต์ในกลุ่มประเทศอาเซียน เหล่านี้ในระยะยาวด้วยเช่นกัน

Dean Fantazzini และ Zhamal Toktamysova (2015) ได้นำเสนอตัวแบบการพยากรณ์ยอดขายรถยนต์รายเดือน โดยใช้ตัวแปรทางเศรษฐกิจและข้อมูลการค้นหาออนไลน์ในเวปไซด์ Google

C. Leenawong และ P. Netisopakul (2016) ได้นำเสนอตัวแบบการพยากรณ์ยอดขายรถยนต์ในประเทศไทยที่มีการวิเคราะห์ผลกระทบจากภาวะน้ำท่วมครั้งรุนแรงในช่วงฤดูมรสุมปลายปี 2554 และโครงการคืนภาษีรถยนต์สำหรับผู้ซื้อรถยนต์คันแรกที่มีรายได้ต่ำและสำหรับรถยนต์ที่มีกำลังเครื่องยนต์ไม่เกิน 1,500 ซีซี เพื่อกระตุ้นยอดขายรถยนต์ในประเทศไทย นโยบายคืนเงินภาษีรถยนต์เริ่มเปิดจองรถยนต์ตั้งแต่วันที่ 16 กันยายน 2554 ถึง 31 ธันวาคม 2555 โดยการคืนเงินภาษีสูงสุดไม่เกิน 100,00 บาท และผู้ซื้อจะต้องครอบครองรถยนต์ครบ 1 ปี โดยการวิเคราะห์ข้อมูลอนุกรมเวลาทั้งก่อนและหลังสถานการณ์พิเศษจากข้อมูลยอดขายรถยนต์ในอดีต โดยมีองค์ประกอบของแนวโน้มและฤดูกาล จากการวิเคราะห์ข้อมูลพบว่าข้อมูลแสดงให้เห็นว่ายอดขาย

รถยนต์รายเดือนสามารถแบ่งออกเป็นช่วงเวลา ได้แก่ ช่วงเวลาปกติ ช่วงเวลาน้ำท่วม ช่วงเวลาที่มีนโยบายคืนเงินภาษีรถยนต์เพื่อเพิ่มแรงจูงใจรวมทั้งกระตุ้นยอดขายรถยนต์ และช่วงหลังสถานการณ์พิเศษนี้แล้ว โดยใช้วิธีปรับรูปแบบเอกโพเนนเชียล Holt-Winter ที่มีเดือนเป็นฤดูกาล ผลการศึกษาพบว่าตัวแบบการพยากรณ์ที่ได้นำเสนอนั้นมีความแม่นยำสูง โดยพิจารณาจากค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์ (MAPE) ที่ต่ำถึง 9.97% ซึ่งภายหลังจากสถานการณ์พิเศษนี้พบว่ายอดขายรถยนต์ลดน้อยลง อาจเนื่องมาจากมีผู้ซื้อรถยนต์จำนวนมากแล้วในช่วงเวลาที่มีนโยบายส่งเสริมและกระตุ้นยอดขายรถยนต์

## บทที่ 3

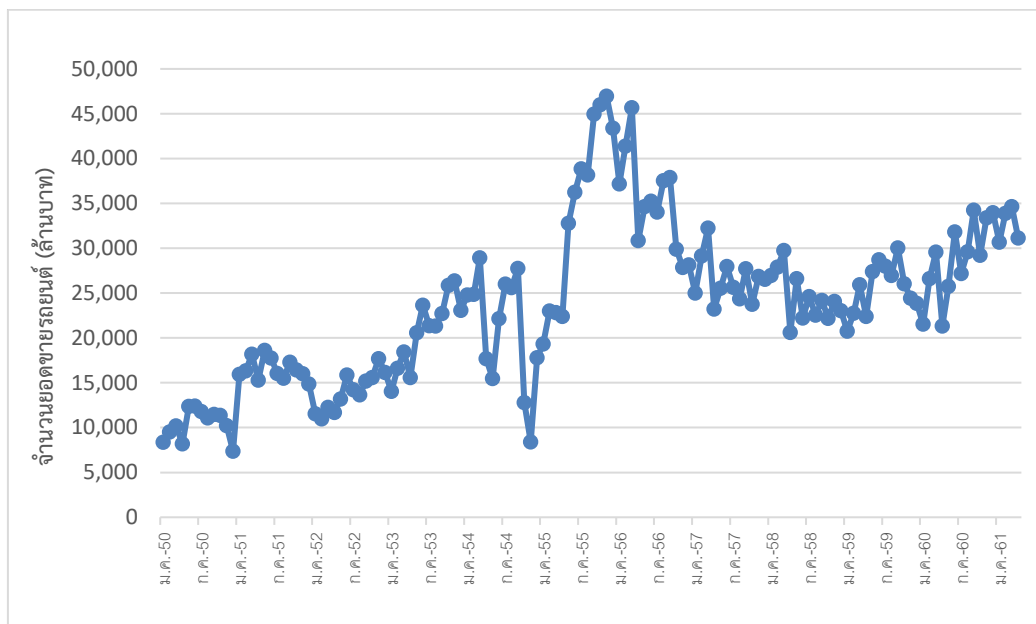
### วิธีการดำเนินงานวิจัย

การวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาตัวแบบการพยากรณ์ยอดขายรถยนต์ขนาดเล็กมาก (Subcompact car) ซึ่งเป็นรถยนต์นั่งส่วนบุคคลที่ขนาดไม่เกิน 1,500 ซีซี ความยาวไม่เกิน 3.4 เมตร และกำลังเป็นที่นิยมในประเทศไทย เนื่องจากมีขนาดกะทัดรัด มีความคล่องตัวในการขับขี่ในเขตชุมชนหรือเขตเมืองที่มีการจราจรหนาแน่น อีกทั้งในปัจจุบันอุตสาหกรรมรถยนต์มีการพัฒนาระบบเครื่องยนต์แบบประหยัดพลังงานส่งผลให้มียอดขายรถยนต์นั่งส่วนบุคคลขนาดเล็กในตลาดเพิ่มมากขึ้น ในการวิจัยนี้จะใช้การวิเคราะห์อนุกรมเวลา (Time Series Analysis) และการวิเคราะห์การถดถอย (Regression Analysis) เพื่อสร้างตัวแบบการพยากรณ์ในการอธิบายยอดขายรถยนต์ โดยทำการปรับแก้ตัวแบบอนุกรมเวลาและตัวแบบการถดถอยด้วยการเพิ่มปัจจัยสถานการณ์ (Event Factor) และใช้เกณฑ์การเปรียบเทียบจากการวัดค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์ (MAPE) ค่าสัมประสิทธิ์การตัดสินใจ  $R^2$  เพื่อให้ได้ตัวแบบการพยากรณ์ที่มีความคลาดเคลื่อนน้อยที่สุด นั่นคือเพื่อให้ได้ตัวแบบพยากรณ์ที่มีความแม่นยำในการพยากรณ์สูงขึ้นนั่นเอง โดยมีขั้นตอนในการดำเนินการวิจัยและเครื่องมือที่ใช้ในการวิเคราะห์ ดังนี้

**ขั้นที่ 1** เก็บรวบรวมข้อมูลที่ใช้ในการวิจัย ได้แก่ ข้อมูลยอดขายรถยนต์ขนาดเล็กมาก ผลิตภัณฑ์มวลรวมในประเทศ (Gross Domestic Products : GDP) และอัตราดอกเบี้ยเงินกู้ 12 เดือน (12-month Loan's Interest Rate) จากข้อมูลในเว็บไซต์ของสำนักงานเศรษฐกิจอุตสาหกรรมและธนาคารแห่งประเทศไทยที่ได้มีเก็บรวบรวมไว้เป็นรายเดือน ย้อนหลังตั้งแต่เดือนมกราคม 2548 ถึงเดือนธันวาคม 2561

**ขั้นที่ 2** สร้างตัวแบบพยากรณ์อนุกรมเวลา โดยนำข้อมูลอนุกรมเวลาของยอดขายรถยนต์มาสร้างกราฟเพื่อศึกษาลักษณะทั่วไปของข้อมูลอนุกรมเวลา และสร้างตัวแบบอนุกรมเวลาโดยวิธีการแยกส่วนประกอบ (Decomposition Method) ดังนี้

พิจารณาข้อมูลยอดขายรถยนต์นั่งขนาดเล็กมากในประเทศไทยย้อนหลัง ตั้งแต่ มกราคม 2550 ถึงธันวาคม 2561 ดังแสดงในรูปที่ 3.1



รูปที่ 3.1 กราฟแสดงยอดขายรถยนต์นั่งขนาดเล็กมารายเดือนของประเทศไทย ระหว่างเดือนมกราคม 2550 ถึงเดือนมกราคม 2561

จากข้อมูลที่แสดงในกราฟตามรูปที่ 3.1 ข้างต้น จะพบว่าข้อมูลยอดขายรถยนต์มีความผันผวนในช่วง ปี 2554 – 2556 ซึ่งในช่วงเวลาดังกล่าวประเทศไทยประสบปัญหาอุทกภัยครั้งใหญ่ ส่งผลกระทบต่อเศรษฐกิจไทยในทุก ๆ ภาคส่วน โดยเฉพาะอย่างยิ่งอุตสาหกรรมยานยนต์ที่ประเทศไทยถือได้ว่าเป็นฐานกำลังผลิตที่สำคัญในภูมิภาคอาเซียน และเป็นช่วงเวลาที่รัฐบาลมีมาตรการกระตุ้นเศรษฐกิจ โดยออกนโยบายค้ำเงินภาษีสำหรับผู้ซื้อรถยนต์คันแรกสูงสุดไม่เกิน 100,000 บาท เมื่อผู้ซื้อครอบครองรถยนต์ครบ 1 ปี เริ่มจองรถยนต์ตามโครงการนี้ตั้งแต่ 16 กันยายน 2554 ถึง 31 ธันวาคม 2555 และจากข้อมูลอนุกรมเวลาข้างต้นนี้ หากพิจารณาข้อมูลองค์ประกอบของอนุกรมเวลา ก็จะพบว่าการเปลี่ยนแปลงของข้อมูลอย่างรวดเร็วในช่วงเวลาหนึ่ง ซึ่งไม่เป็นไปอย่างปกติ จึงจะสร้างตัวแบบอนุกรมเวลาเพื่อพยากรณ์ยอดขายโดยใช้ตัวแบบอนุกรมเวลาแบบการคูณ ซึ่งเป็นความสัมพันธ์ระหว่างส่วนประกอบทั้ง 4 ของอนุกรมเวลา ตามสมการ (3.1)

$$Y = T * S * C * I \quad (3.1)$$

- เมื่อ  $T$  แทน แนวโน้มของข้อมูล (Trend)  
 $S$  แทน ฤดูกาล (Season)  
 $C$  แทน วัฏจักร (Cycle)  
 และ  $I$  แทน ความไม่สม่ำเสมอของข้อมูล (Irregularity)

แต่เนื่องจากข้อมูลมีความผิดปกติจากสถานการณ์พิเศษ จึงปรับเปลี่ยนการสร้างตัวแบบพยากรณ์โดยเพิ่มองค์ประกอบของสถานการณ์พิเศษ หรือสถานการณ์ที่ไม่ปกติ (Event ;  $E$ ) เพื่อให้สอดคล้องกับข้อมูลอนุกรมเวลาของยอดขายรถยนต์ และใช้ตัวแบบอนุกรมเวลาแบบการคูณ ดังสมการ (3.2)

$$Y = T * S * C * I * E \quad (3.2)$$

เมื่อ  $T$  แทน แนวโน้มของข้อมูล  
 $S$  แทน ฤดูกาล  
 $C$  แทน วัฏจักร  
 $I$  แทน ความไม่สม่ำเสมอของข้อมูล  
 และ  $E$  แทน สถานการณ์พิเศษ (Event)

กำหนดให้  $X_{y,m}^k$  แทนข้อมูลอนุกรมเวลาของยอดขายรถยนต์ของปีที่  $y$ , เดือน  $m$  ที่สอดคล้องกับสถานการณ์  $k$

เมื่อ  $y=1,2,\dots$  แทน ปี ในที่งานวิจัยนี้ใช้ข้อมูลอนุกรมเวลาของยอดขายรถยนต์ปีแรก

คือ พ.ศ. 2554, พ.ศ. 2555, ... ,พ.ศ. 2561 ตามลำดับ

$m=1,2,3,\dots,12$  แทน เดือนมกราคม, กุมภาพันธ์, ... , ธันวาคม ตามลำดับ และ

สถานการณ์  $k=0,1,2,3$  กำหนดโดย

$k=0$  หมายถึง ช่วงเวลาที่ยอดขายรถยนต์อยู่ในสถานการณ์ปกติที่ยังไม่ได้รับผลกระทบจากสถานการณ์น้ำท่วมและนโยบายคืนเงินภาษีรถยนต์ (ม.ค. 2554 - ก.ย. 2554 และ ต.ค. 2557 - มิ.ย. 2561)

$k=1$  หมายถึง ช่วงเวลาที่ยอดขายรถยนต์ได้รับผลกระทบจากสถานการณ์น้ำท่วม (ต.ค. 2554 - ม.ค. 2555)

$k=2$  หมายถึง ช่วงเวลาที่ยอดขายรถยนต์ได้รับผลกระทบจากสถานการณ์น้ำท่วมและนโยบายคืนเงินภาษีรถยนต์ (ก.พ. 2555 - ก.ย. 2556)

$k=3$  หมายถึง ช่วงเวลาหลังจากสิ้นสุดนโยบายคืนเงินภาษีรถยนต์ (ต.ค. 2556 - ก.ย. 2557)

จากตัวแบบอนุกรมเวลาในสมการ (3.2) จะสร้างตัวแบบพยากรณ์สถานการณ์ (Forecasting Event Modelling) เพื่อพยากรณ์ยอดขายรถยนต์โดยวิธีการแยกส่วนประกอบอนุกรมเวลา (Time-Series Decomposition Method : TSD) ตามขั้นตอนต่อไปนี้

**ขั้นตอนที่ 2.1** หาค่าเฉลี่ยเคลื่อนที่ (Centered Moving Averages : CMAs) ของข้อมูลยอดขายรถยนต์ ดังนี้

ให้  $\bar{X}_{y,m}^k$  แทนค่าเฉลี่ยเคลื่อนที่ของ  $X_{y,m}^k$  โดยเฉลี่ยครั้งละ 12 ชุดข้อมูล จะได้ค่าเฉลี่ยเคลื่อนที่ชุดที่ 1 คือ  $\bar{X}_{1,7}^k$  และค่าเฉลี่ยเคลื่อนที่ชุดสุดท้าย คือ  $\bar{X}_{8,6}^k$  ตามลำดับดังนี้

$$\begin{aligned}
\bar{X}_{1,7}^k &= CMA(X_{1,7}^k) = \frac{X_{1,1}^k + X_{1,2}^k + \dots + X_{1,12}^k}{12}, \\
\bar{X}_{1,8}^k &= CMA(X_{1,8}^k) = \frac{X_{1,2}^k + X_{1,3}^k + \dots + X_{1,12}^k + X_{2,1}^k}{12}, \\
\bar{X}_{1,9}^k &= CMA(X_{1,9}^k) = \frac{X_{1,3}^k + X_{1,4}^k + \dots + X_{1,12}^k + X_{2,1}^k + X_{2,2}^k}{12}, \\
&\vdots \\
\bar{X}_{8,6}^k &= CMA(X_{8,6}^k) = \frac{X_{8,1}^k + X_{8,2}^k + \dots + X_{8,12}^k}{12}
\end{aligned} \tag{3.3}$$

ในที่นี้ข้อมูล  $X_{y,m}^k$  และ  $\bar{X}_{y,m}^k$  จะสอดคล้องกับช่วงเวลาที่เกิดสถานการณ์  $k$  เมื่อ  $k=0,1,2,3$  ตามช่วงเวลาที่กำหนดข้างต้น

**ขั้นตอนที่ 2.2** พิจารณาข้อมูลค่าเฉลี่ยเคลื่อนที่ของสถานการณ์ปกติ ( $k=0$ ) ที่ได้จากขั้นตอนที่ 1 ตามสมการ (3.3) โดยนำค่าเฉลี่ยเคลื่อนที่  $\bar{X}_{y,m}^k$  ของแต่ละเดือน ( $m=1,2,\dots,12$ ) เฉพาะที่มีช่วงเวลาสถานการณ์ปกติ  $k=0$  มาหาค่าเฉลี่ยของแต่ละเดือน จะได้ค่าเฉลี่ยเคลื่อนที่ของสถานการณ์ปกติ ( $k=0$ ) ของแต่ละเดือน ดังสมการ

$$\bar{X}_m^0 = avg \bar{X}_{y,m}^0 \tag{3.4}$$

**ขั้นตอนที่ 2.3** หาปัจจัยสถานการณ์ (Event Factors) ของข้อมูลยอดขายรถยนต์

ให้  $EF_{y,m}^k$  แทนปัจจัยสถานการณ์ในปีที่  $y$  เดือน  $m$  ซึ่งสอดคล้องกับสถานการณ์  $k$  เดียวกัน ปัจจัยสถานการณ์นี้ คำนวณได้จากอัตราส่วนของค่าเฉลี่ยเคลื่อนที่ของข้อมูล ( $\bar{X}_{y,m}^k$ ) กับค่าเฉลี่ยเคลื่อนที่ของสถานการณ์ปกติ ( $k=0$ ) ของแต่ละเดือน ( $\bar{X}_m^0$ ) ดังสมการ

$$EF_{y,m}^k = \frac{\bar{X}_{y,m}^k}{\bar{X}_m^0} \tag{3.5}$$

**ขั้นตอนที่ 2.4** หาดัชนีสถานการณ์ (Event Indices) ของข้อมูลยอดขายรถยนต์

ให้  $E^k$  แทนดัชนีสถานการณ์สำหรับสถานการณ์  $k$  เมื่อ  $k=0,1,2,3$  ดัชนีสถานการณ์ คำนวณได้จากการหาค่าเฉลี่ยของปัจจัยสถานการณ์ ( $EF_{y,m}^k$ ) ที่มีสถานการณ์  $k$  เดียวกัน ดังสมการ

$$E^k = avg EF_{y,m}^k \tag{3.6}$$

**ขั้นตอนที่ 2.5** ถอดผลกระทบของสถานการณ์ออกจากข้อมูลอนุกรมเวลาของยอดขายรถยนต์

ให้  $D_{y,m}^k$  แทนข้อมูลที่ถอดผลกระทบของสถานการณ์ออกจากข้อมูลอนุกรมเวลาของยอดขายรถยนต์ (De-evented data) ที่สอดคล้องกับพารามิเตอร์ 3 ตัว นั่นคือ ปีที่  $y$ , เดือน  $m$  และอยู่ในช่วงเวลาที่เกิดสถานการณ์  $k$

คำนวณหา  $D_{y,m}^k$  ได้จากผลหารของข้อมูลอนุกรมเวลาของยอดขายรถยนต์แต่ละค่า ( $X_{y,m}^k$ ) กับดัชนีสถานการณ์  $E^k$  ที่มีสถานการณ์  $k$  เดียวกัน ดังสมการ

$$D_{y,m}^k = \frac{X_{y,m}^k}{E^k} \tag{3.7}$$

จากสมการที่ (3.7) จะพบว่าช่วงเวลาของสถานการณ์  $k$  ไม่ส่งผลต่อข้อมูลที่ถูกถอดผลกระทบของสถานการณ์ออกจากข้อมูลอนุกรมเวลาแล้ว ดังนั้นจึงปรับเปลี่ยนตัวแปรเพื่อให้สะดวกและเข้าใจง่ายขึ้น นั่นคือ  $D_{y,m} = D_{y,m}^k$

จากขั้นตอนที่ 2.1 ถึงขั้นตอนที่ 2.5 จะได้ว่า ข้อมูลอนุกรมเวลาของยอดขายรถยนต์ถูกถอดผลกระทบของสถานการณ์ออกไปแล้ว ดังนั้นข้อมูลที่เหลืออยู่จึงมีเพียงตัวประกอบอนุกรมเวลาของฤดูกาลและวัฏจักรเท่านั้น ขั้นตอนต่อจากนี้จะทำการถอดผลกระทบจากตัวประกอบอนุกรมเวลาทั้งสองออกจากข้อมูล

**ขั้นตอนที่ 2.6** หาค่าเฉลี่ยเคลื่อนที่ (Centered Moving Averages; CMAs) ของข้อมูลยอดขายรถยนต์ที่ถอดผลกระทบจากสถานการณ์ออกไปแล้ว ดังนี้

ให้  $\bar{D}_{y,m}$  แทนค่าเฉลี่ยเคลื่อนที่ของข้อมูลยอดขายรถยนต์ที่ถูกถอดผลกระทบจากสถานการณ์ออกไปแล้ว ( $D_{y,m}$ ) โดยเฉลี่ยครั้งละ 12 ชุดข้อมูล จะได้ค่าเฉลี่ยค่าแรกคือ  $\bar{D}_{2,1}$  และค่าเฉลี่ยค่าสุดท้าย คือ  $\bar{D}_{8,6}$  ตามลำดับ ดังนี้

$$\begin{aligned}\bar{D}_{2,1} &= CMA(D_{2,1}) = \frac{D_{1,7}+D_{1,8}+\dots+D_{1,12}+D_{2,1}+D_{2,2}+\dots+D_{2,6}}{12}, \\ \bar{D}_{1,8} &= CMA(D_{1,8}) = \frac{D_{1,2}+D_{1,3}+\dots+D_{1,11}+D_{1,12}}{12}, \\ &\vdots \\ \bar{D}_{8,6} &= CMA(D_{8,6}) = \frac{D_{8,1}+D_{8,2}+\dots+D_{8,11}+D_{8,12}}{12}\end{aligned}\tag{3.8}$$

**ขั้นตอนที่ 2.7** หาปัจจัยฤดูกาล (Seasonal Factors) ของข้อมูลยอดขายรถยนต์ที่ถูกถอดผลกระทบจากสถานการณ์ออกไปแล้ว

ให้  $SF_{y,m}$  แทน ปัจจัยฤดูกาลของข้อมูลยอดขายรถยนต์ที่ถูกถอดผลกระทบจากสถานการณ์ออกไปแล้วในปีที่  $y$  และเดือน  $m$  คำนวณได้จากผลหารของข้อมูลยอดขายรถยนต์ที่ถูกถอดปัจจัยสถานการณ์ออกไปแล้วกับค่าเฉลี่ยเคลื่อนที่ของข้อมูลยอดขายรถยนต์ที่ถูกถอดผลกระทบจากสถานการณ์ออกไปแล้ว นั่นคือ

$$SF_{y,m} = \frac{D_{y,m}}{\bar{D}_{y,m}}\tag{3.9}$$

**ขั้นตอนที่ 2.8** หาดัชนีฤดูกาล (Seasonal Indices) ของข้อมูลยอดขายรถยนต์ที่ถูกถอดผลกระทบจากสถานการณ์ออกไปแล้ว

ให้  $S_m$  แทนดัชนีฤดูกาลของเดือน  $m$  ซึ่งคำนวณได้จากการหาค่าเฉลี่ยของปัจจัยฤดูกาลในเดือนเดียวกันของแต่ละปี ดังสมการ

$$S_m = avgSF_{y,m}\tag{3.10}$$

จากขั้นตอนที่ 2.6 ถึงขั้นตอนที่ 2.8 เป็นขั้นตอนในการคำนวณหาดัชนีฤดูกาลของข้อมูล ยอดขายรถยนต์ที่ถูกถอดปัจจัยสถานการณ์ออกไปแล้ว ขั้นตอนต่อไปจากนี้จะพิจารณา ส่วนประกอบอนุกรมเวลาในส่วนของวัฏจักรเพื่อหาปัจจัยวัฏจักรของข้อมูลยอดขายรถยนต์ที่ถูกถอด ปัจจัยสถานการณ์ออกไปแล้ว

**ขั้นตอนที่ 2.9** หาสมการแนวโน้มของข้อมูลยอดขายรถยนต์ที่ถูกถอดปัจจัยสถานการณ์ออกไปแล้ว

ให้  $TD_{y,m}$  แทน สมการแนวโน้มของข้อมูลยอดขายรถยนต์ที่ถูกถอดปัจจัยสถานการณ์ ออกไปแล้ว ( $\bar{D}_{y,m}$ ) ที่สอดคล้องกับเวลา  $t$  กำหนดโดย

$$TD_{y,m} = b_0 + b_1 t \quad (3.11)$$

เมื่อ  $b_0$  คือจุดตัดแกน Y ของเส้นแนวโน้ม และ  $b_1$  คือความชันของเส้นแนวโน้ม

**ขั้นตอนที่ 2.10** หาปัจจัยวัฏจักร (Cycle Factors) ของข้อมูลยอดขายรถยนต์ที่ถูกถอดปัจจัย สถานการณ์ออกไปแล้ว

ให้  $CF_{y,m}$  แทนปัจจัยวัฏจักรของข้อมูลยอดขายรถยนต์ที่ถูกถอดปัจจัยสถานการณ์ออกไป แล้ว ของปีที่  $y$  เดือนที่  $m$  คำนวณจากสมการ

$$CF_{y,m} = \frac{\bar{D}_{y,m}}{TD_{y,m}} \quad (3.12)$$

**ขั้นตอนที่ 2.11** ถอดส่วนประกอบฤดูกาลและวัฏจักรออกจากข้อมูลยอดขายรถยนต์ที่ถูกถอด ปัจจัยสถานการณ์ออกไปแล้วก่อนหน้านี้

ให้  $DA_{y,m}$  แทน ข้อมูลยอดขายรถยนต์ที่ถูกถอดส่วนประกอบอนุกรมเวลาทั้ง 3 ส่วน ได้แก่ ปัจจัยสถานการณ์ ฤดูกาล และวัฏจักร ซึ่งได้จากรนำข้อมูลที่ถูกถอดปัจจัยสถานการณ์จาก ขั้นตอนที่ 2.5 ตามสมการ (3.7) หาคด้วยดัชนีฤดูกาลที่ได้จากขั้นตอนที่ 2.8 ตามสมการ (3.10) และ ปัจจัยวัฏจักรที่ได้จากขั้นตอนที่ 2.10 ตามสมการ (3.12) จะได้

$$DA_{y,m} = \frac{D_{y,m}}{S_m * CF_{y,m}} \quad (3.13)$$

**ขั้นตอนที่ 2.12** หาสมการแนวโน้มของข้อมูลที่ถูกถอดส่วนประกอบอนุกรมเวลาทั้ง 3 ส่วน ออกไปแล้วในขั้นตอนที่ 2.11

ให้  $TDA_{y,m}$  แทน สมการแนวโน้มของข้อมูลที่ถูกถอดส่วนประกอบอนุกรมเวลาทั้ง 3 ส่วนออกไปแล้วตามสมการ (3.13) ที่สอดคล้องกับปีที่  $y$  เดือน  $m$  จะได้รูปแบบทั่วไปของสมการ แนวโน้มที่สอดคล้องกับเวลา  $t$  ดังสมการ

$$TDA_{y,m} = c_0 + c_1 t \quad (3.14)$$

เมื่อ  $c_0$  คือจุดตัดแกน Y ของเส้นแนวโน้ม และ  $c_1$  คือความชันของเส้นแนวโน้ม

**ขั้นตอนที่ 2.13** รวมส่วนประกอบอนุกรมเวลาทั้ง 3 ส่วน กลับเข้าไปในข้อมูลพยากรณ์ เพื่อสร้างตัวแบบพยากรณ์อนุกรมเวลาของยอดขายรถยนต์

ให้  $TF_{y,m}$  แทนสมการตัวแบบพยากรณ์อนุกรมเวลา นำสมการแนวโน้มของข้อมูลที่ถูกถอดส่วนประกอบอนุกรมเวลาทั้ง 3 ส่วนออกตามสมการ (3.14) และส่วนประกอบปัจจัยอนุกรมเวลาฤดูกาล (สมการ 3.10) วัฏจักร (สมการ 3.12) และสถานการณ์ (สมการ 3.6) มาสร้างตัวแบบพยากรณ์อนุกรมเวลาแบบการคูณ จะได้

$$TF_{y,m} = TDA_{y,m} * S_m * CF_{y,m} * E^k \quad (3.15)$$

สมการ (3.15) เป็นสมการตัวแบบอนุกรมเวลาที่พยากรณ์ยอดขายรถยนต์โดยใช้วิธีการแยกส่วนประกอบ

**ขั้นที่ 3** นำข้อมูลอนุกรมเวลามาพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างปัจจัยหลัก 2 ปัจจัย ได้แก่ ผลิตภัณฑ์มวลรวมในประเทศ (GDP) และอัตราดอกเบี้ยเงินกู้ 12 เดือน แล้วสร้างตัวแบบพยากรณ์ถดถอยเชิงเส้นแบบพหุ ตามขั้นตอนต่อไปนี้

**ขั้นตอนที่ 3.1** สร้างตัวแบบพยากรณ์ถดถอยเชิงเส้นแบบพหุ ตัวแบบที่ 1 คือตัวแบบการถดถอยเชิงเส้นแบบพหุเพื่ออธิบายความสัมพันธ์ของยอดขายรถยนต์กับ GDP และอัตราดอกเบี้ยเงินกู้ พร้อมทั้งพิจารณาค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์ (MAPE) และค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของตัวแบบที่ได้

**ขั้นตอนที่ 3.2** พิจารณากราฟความสัมพันธ์ของข้อมูลยอดขายรถยนต์รายเดือนในแต่ละปี เพื่อวิเคราะห์ปัจจัยฤดูกาลซึ่งเป็นอีกปัจจัยหนึ่งที่มีอิทธิพลต่อยอดขายรถยนต์ ดังนี้

1. สร้างกราฟข้อมูลยอดขายรายเดือนของทุกปี (มกราคม 2548 ถึงกันยายน 2558)
2. สร้างกราฟข้อมูลยอดขายรายเดือนก่อนเกิดสถานการณ์น้ำท่วมใหญ่ (มกราคม 2548 ถึงธันวาคม 2553)
3. สร้างกราฟข้อมูลยอดขายรายเดือนโดยลดจำนวนข้อมูลหรือตัดข้อมูล (Cut Data) บางส่วน ในที่นี้จะทำการตัดข้อมูลในช่วงเวลาที่เกิดสถานการณ์น้ำท่วมใหญ่และช่วงเวลาที่รัฐบาลมีการใช้นโยบายมาตรการการเงินภาษีรถยนต์คันแรก นั่นคือจะตัดข้อมูลระหว่างช่วงเดือนกันยายน 2554 ถึงเดือนมกราคม 2557 จำนวน 29 ชุดข้อมูล จากข้อมูลทั้งหมด 129 ชุดข้อมูลที่นำมาวิเคราะห์ซึ่งเป็นข้อมูลรายเดือนที่เก็บรวบรวมข้อมูลย้อนหลัง 10 ปี ตั้งแต่เดือนมกราคม 2548 ถึงเดือนกันยายน 2558

**ขั้นตอนที่ 3.3** สร้างตัวแบบพยากรณ์ถดถอยเชิงเส้นแบบพหุ ตัวแบบที่ 2 คือตัวแบบการถดถอยเชิงเส้นแบบพหุเพื่ออธิบายความสัมพันธ์ ของยอดขายรถยนต์กับ GDP และอัตราดอกเบี้ยเงินกู้ เมื่อเพิ่มปัจจัยฤดูกาลซึ่งเป็นตัวแปรอิสระเชิงคุณภาพ โดยชุดตัวแปรเชิงคุณภาพฤดูกาล แบ่งเป็น 12 กลุ่ม (ตามรายเดือน) จึงมีตัวแปรหุ่น 11 ตัวแปร เพิ่มขึ้นมาในตัวแบบนอกเหนือจากปัจจัย GDP และอัตราดอกเบี้ยเงินกู้ พร้อมทั้งพิจารณาค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์ (MAPE) และค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของตัวแบบที่ได้

**ขั้นตอนที่ 3.4** สร้างตัวแบบพยากรณ์ถดถอยเชิงเส้นแบบพหุ ตัวแบบที่ 3 คือตัวแบบการถดถอยเชิงเส้นแบบพหุเพื่ออธิบายความสัมพันธ์ ของยอดขายรถยนต์กับ GDP และอัตราดอกเบี้ยเงินกู้ เมื่อเพิ่มปัจจัยฤดูกาลและสถานการณ์พิเศษที่รัฐบาลมีนโยบายมาตรการการคืนเงินภาษีรถยนต์คันแรกเพื่อกระตุ้นเศรษฐกิจระหว่างช่วงที่ยังมีสถานการณ์น้ำท่วมและหลังสถานการณ์น้ำท่วมอีก 4 กลุ่ม จะได้ตัวแปรทั้งหมดเป็น 16 กลุ่ม ดังนั้นในตัวแบบที่ 3 จะมีตัวแปรหุ่น 15 ตัวแปร พร้อมทั้งพิจารณาค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์ (MAPE) และค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของตัวแบบที่ได้

**ขั้นตอนที่ 3.5** สร้างตารางเปรียบเทียบค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์ (MAPE) และ  $R^2$  ของตัวแบบการพยากรณ์ถดถอยเชิงเส้นแบบพหุทั้ง 3 ตัวแบบ

#### ขั้นที่ 4 สรุปผลการวิจัยและข้อเสนอแนะ

#### เครื่องมือที่ใช้ในการวิเคราะห์

การวิเคราะห์ตัวแบบอนุกรมเวลา และการวิเคราะห์ตัวแบบการถดถอยเชิงเส้นแบบพหุในงานวิจัยนี้ ใช้โปรแกรม Microsoft Excel และโปรแกรมสำเร็จรูปอื่น

## บทที่ 4

### ผลการวิจัย

ในบทนี้จะแสดงผลการสร้างตัวแบบพยากรณ์โดยวิเคราะห์ลักษณะข้อมูลยอดขายรถยนต์นั่งขนาดเล็กมาก เพื่ออธิบายและพยากรณ์ยอดขายที่เหมาะสมกับข้อมูลอนุกรมเวลา พร้อมทั้งพิจารณาความแม่นยำของตัวแบบพยากรณ์ โดยจะสร้างตัวแบบพยากรณ์ใน 2 รูปแบบ ดังนี้

#### 4.1 ตัวแบบพยากรณ์จากการวิเคราะห์ข้อมูลอนุกรมเวลา

ขั้นตอนที่ 2.1 เป็นขั้นตอนในการปรับข้อมูลอนุกรมเวลาของยอดขายรถยนต์ให้เรียบ โดยใช้วิธีการหาค่าเฉลี่ยเคลื่อนที่ (Centered Moving Averages : CMAs) ของข้อมูลยอดขายรถยนต์ จะได้ค่าตามสมการ (3.3)

ขั้นตอนที่ 2.2 ถึงขั้นตอนที่ 2.4 เป็นขั้นตอนการหาปัจจัยสถานการณ์ (Event Factors) และดัชนีสถานการณ์ (Event Indices) ได้ผลลัพธ์ดังตาราง 4.1

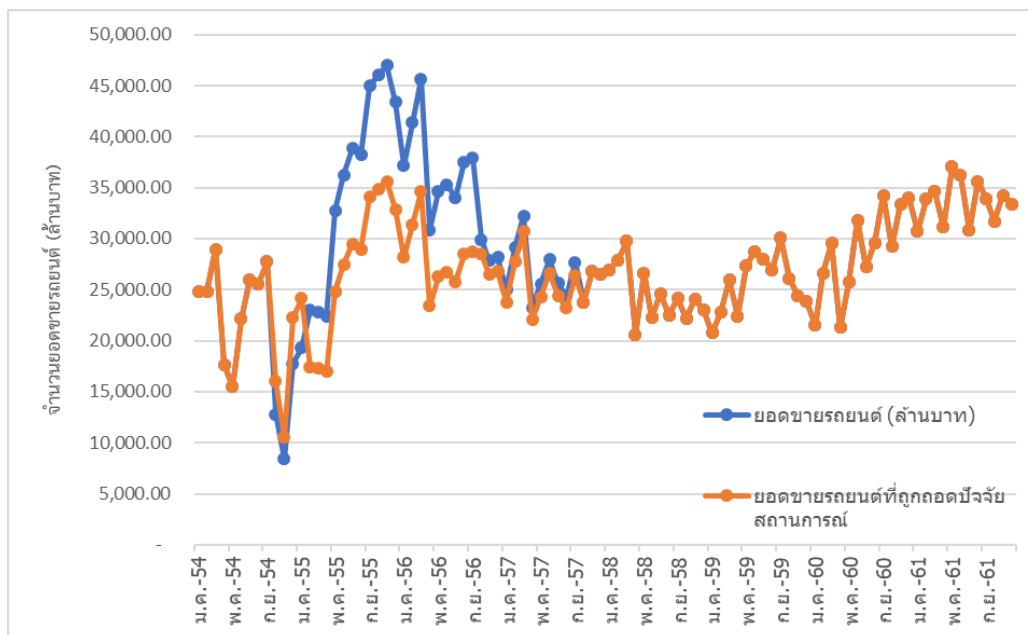
ตาราง 4.1 ตารางแสดงค่าดัชนีสถานการณ์ที่ได้จากขั้นตอนที่ 2.2 – 2.4

| สถานการณ์ ( $k$ )        | $k=0$ | $k=1$ | $k=2$ | $k=3$ |
|--------------------------|-------|-------|-------|-------|
| ดัชนีสถานการณ์ ( $E^k$ ) | 1.00  | 0.80  | 1.32  | 1.05  |

จากตาราง 4.1 จะพบว่า ในช่วงสถานการณ์ปกติ ( $k=0$ ) ดัชนีสถานการณ์ในช่วงสถานการณ์ปกติ ( $E^0$ ) มีค่าเท่ากับ 1 แสดงว่ายอดขายรถยนต์ยังเป็นไปอย่างปกติ แต่ในช่วงสถานการณ์น้ำท่วม ( $k=1$ ) ดัชนีสถานการณ์  $E^1$  มีค่าลดลงเท่ากับ 0.80 สอดคล้องกับยอดขายรถยนต์ที่ลดลงในช่วงเวลาที่เกิดสถานการณ์น้ำท่วมหรืออาจกล่าวได้ว่าสถานการณ์น้ำท่วมส่งผลกระทบต่อยอดขายรถยนต์ โดยมีผลทำให้ยอดขายรถยนต์ลดลงจากเดิม ต่อมาในช่วงปลายของระยะเวลาที่เกิดสถานการณ์น้ำท่วม รัฐบาลได้ประกาศใช้นโยบายคืนเงินภาษีรถยนต์สำหรับผู้ซื้อรถยนต์คันแรก ซึ่งเป็นช่วงเวลาที่ยอดขายรถยนต์ได้รับผลกระทบจากสถานการณ์น้ำท่วม และนโยบายคืนเงินภาษีรถยนต์ ( $k=2$ ) จะพบว่าดัชนีสถานการณ์  $E^2$  มีค่าเพิ่มขึ้นจากเดิมเป็น 1.32 สอดคล้องกับสถานการณ์ที่ยอดขายรถยนต์ตลอดจนยอดการจองรถยนต์สูงขึ้นในช่วงสถานการณ์นั้นอย่างเห็นได้ชัด และเมื่อสิ้นสุดนโยบายคืนเงินภาษีรถยนต์ ( $k=3$ ) จะพบว่าดัชนีสถานการณ์  $E^3$  ก็ลดลงจากเดิมเป็น 1.05 ใกล้เคียงกับสถานการณ์ปกติ

ขั้นตอนที่ 2.5 เป็นขั้นตอนในการถอดผลกระทบของสถานการณ์ออกจากข้อมูลอนุกรมเวลาของยอดขายรถยนต์ ดังนั้นข้อมูลที่เหลืออยู่จึงมีเพียงตัวประกอบอนุกรมเวลาของฤดูกาลและวัฏจักรเท่านั้น พิจารณากราฟเปรียบเทียบข้อมูลอนุกรมเวลาของยอดขายรถยนต์

กับข้อมูลที่ถอดผลกระทบของสถานการณ์ออกจากข้อมูลอนุกรมเวลาของยอดขายรถยนต์ (De-event data) ดังรูปที่ 4.1



รูปที่ 4.1 กราฟเปรียบเทียบข้อมูลอนุกรมเวลาของยอดขายรถยนต์กับข้อมูลที่ถอดผลกระทบของสถานการณ์ออกจากข้อมูลอนุกรมเวลาของยอดขายรถยนต์ (De-event data)

ภายหลังจากที่ได้มีการปรับข้อมูลยอดขายรถยนต์เดิมโดยถอดปัจจัยสถานการณ์แล้ว ตามขั้นตอนที่ 2.1 ถึงขั้นตอนที่ 2.5 ในลำดับต่อไปจะเป็นขั้นตอนในการพิจารณาปัจจัยฤดูกาลและดัชนีฤดูกาล ดังในขั้นตอนที่ 2.6 ถึงขั้นตอนที่ 2.8 ซึ่งจะพบได้ว่าดัชนีฤดูกาลของแต่ละเดือนมีค่าใกล้เคียงกัน ดังแสดงในตาราง 4.2 จึงอาจกล่าวได้ว่ายอดขายรถยนต์ในแต่ละเดือนภายหลังจากที่ถูกถอดปัจจัยสถานการณ์แล้วนั้นไม่ได้มีความแตกต่างกันมากเท่าใด

ตาราง 4.2 ตารางแสดงค่าดัชนีฤดูกาลที่ได้จากขั้นตอนที่ 2.6 – 2.8

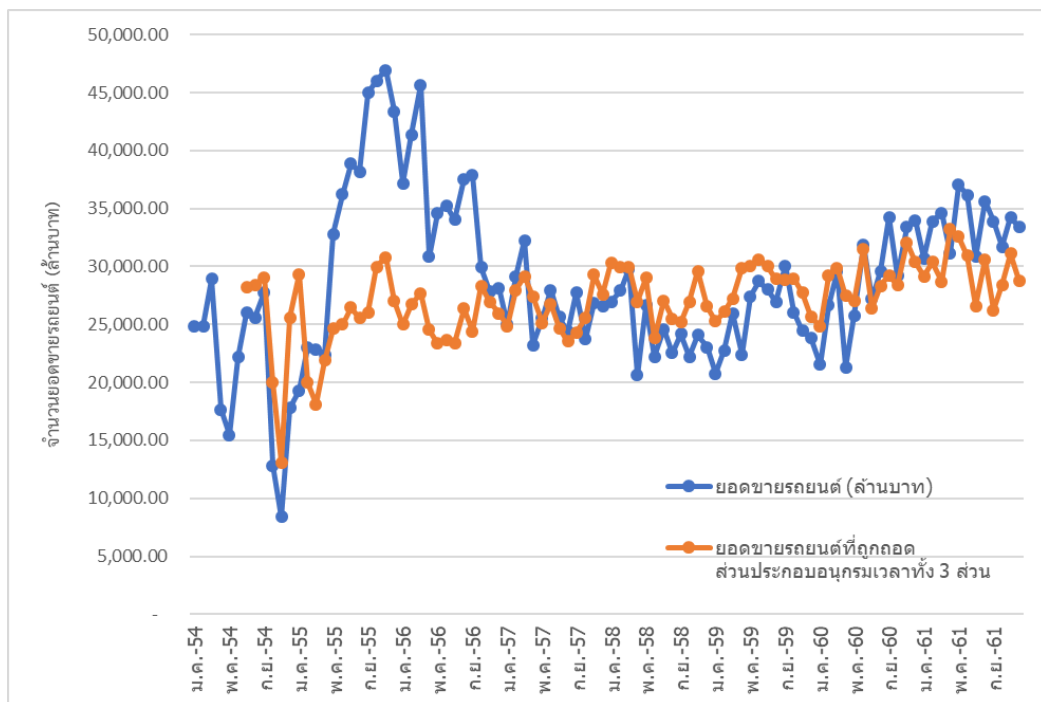
| เดือน ( $m$ ) | ดัชนีฤดูกาล ( $S_m$ ) |
|---------------|-----------------------|
| 1             | 0.95                  |
| 2             | 0.99                  |
| 3             | 1.07                  |
| 4             | 0.83                  |
| 5             | 1.00                  |
| 6             | 1.04                  |

| เดือน ( $m$ ) | ดัชนีฤดูกาล ( $S_m$ ) |
|---------------|-----------------------|
| 7             | 1.03                  |
| 8             | 1.02                  |
| 9             | 1.13                  |
| 10            | 0.97                  |
| 11            | 0.96                  |
| 12            | 1.01                  |

ในขั้นตอนที่ 2.9 ซึ่งเป็นขั้นตอนในการหาสมการแนวโน้มจากข้อมูลยอดขายรถยนต์ภายหลังจากที่ถูกถอดปัจจัยสถานการณ์ออกไปแล้ว จะได้สมการดังนี้

$$TD_{y,m} = 65.732t + 23,452 \quad (4.1)$$

ในขั้นตอนที่ 2.10 เป็นการหาปัจจัยวัฏจักรของข้อมูลยอดขายรถยนต์ภายหลังจากที่ถูกถอดปัจจัยสถานการณ์ออกไปแล้ว และดำเนินการถอดส่วนประกอบฤดูกาลและวัฏจักรออกจากข้อมูลยอดขายรถยนต์ที่ถูกถอดปัจจัยสถานการณ์ออกไปแล้วตามขั้นตอนที่ 2.11 ทำให้ได้ข้อมูลยอดขายรถยนต์ที่ถูกถอดส่วนประกอบอนุกรมเวลาทั้ง 3 ส่วน ดังในรูปที่ 4.2 ที่เป็นกราฟแสดงการเปรียบเทียบข้อมูลอนุกรมเวลาของยอดขายรถยนต์กับข้อมูลที่ถอดผลกระทบของส่วนประกอบอนุกรมเวลาทั้ง 3 ส่วน ได้แก่ ปัจจัยสถานการณ์ ฤดูกาล และวัฏจักร ออกจากข้อมูลอนุกรมเวลาของยอดขายรถยนต์

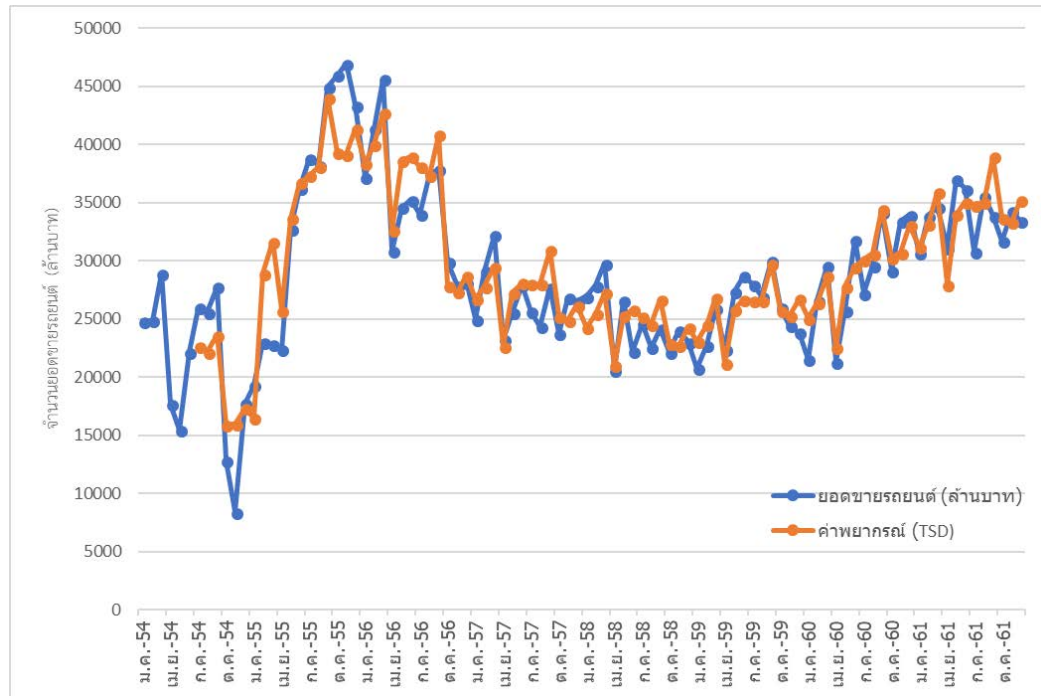


รูปที่ 4.2 กราฟเปรียบเทียบข้อมูลอนุกรมเวลาของยอดขายรถยนต์กับข้อมูลยอดขายรถยนต์ที่ถูกถอดส่วนประกอบของอนุกรมเวลาทั้ง 3 ส่วน

ในขั้นตอนที่ 2.12 ซึ่งเป็นขั้นตอนในการหาสมการแนวโน้มจากข้อมูลยอดขายรถยนต์ที่ถูกถอดส่วนประกอบของอนุกรมเวลาทั้ง 3 ส่วนออกไปแล้ว ตามขั้นตอนที่ 2.11 จะได้สมการดังนี้

$$TDA_{y,m} = 66.612t + 23,349 \quad (4.2)$$

เมื่อดำเนินการตามขั้นตอนวิธีการพยากรณ์ด้วยอนุกรมเวลาแบบแยกส่วนประกอบแล้ว จะได้ค่าพยากรณ์ยอดขายรถยนต์จากตัวแบบอนุกรมเวลาแบบแยกส่วนประกอบเปรียบเทียบกับยอดขายรถยนต์ ดังรูปที่ 4.3



รูปที่ 4.3 กราฟแสดงยอดขายรถยนต์นั่งขนาดเล็มากเปรียบเทียบกับค่าพยากรณ์ที่ได้จากวิธีการแยกส่วนประกอบอนุกรมเวลา

พิจารณาค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์ (MAPE) และรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของตัวแบบอนุกรมเวลาที่ได้จากการใช้วิธีการแยกส่วนประกอบอนุกรมเวลา คือ

$$\begin{aligned}
 MAPE &= \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \times 100 \\
 &= \frac{1}{90} \sum \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \times 100 \\
 &= \frac{1}{90} (744.6373) \\
 &= 8.2737\%
 \end{aligned}
 \tag{4.3}$$

$$\begin{aligned}
 RMSE &= \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}} \\
 &= \sqrt{\frac{673,160,187.73}{90}} \\
 &= \sqrt{7,479,557.64} \\
 &= 2,734.88
 \end{aligned} \tag{4.4}$$

จากการดำเนินการตามลำดับขั้นตอนทั้งหมดข้างต้น จะพบว่าดัชนีสถานการณ์ของยอดขายรถยนต์ที่อยู่ในสถานการณ์ปกติ ยังไม่ได้รับผลกระทบจากสถานการณ์น้ำท่วมและสถานการณ์ของนโยบายค้ำเงินภาษีรถยนต์ นั่นคือ สถานการณ์ที่  $k=0$  มีค่าเท่ากับ 1 หรือ  $E^0 = 1$  แสดงได้ดังนี้

**ทฤษฎีบท 4.1** ให้  $E^k$  แทนดัชนีสถานการณ์ สำหรับสถานการณ์  $k=0,1,2,3$  กำหนดโดยสมการ (3.6) นั่นคือ ถ้า  $E^k = avgEF_{y,m}^k$  แล้ว  $E^0 = 1$

**พิสูจน์** ให้  $n_m^0$  แทนจำนวนข้อมูลทั้งหมดที่มีสถานการณ์  $k=0$  ของทุก ๆ ปี สำหรับเดือน  $m$  ( $m=1,2,\dots,12$ )

จากขั้นตอนที่ 2.4 จะได้ว่า ดัชนีสถานการณ์ของช่วงเวลาที่สถานการณ์ปกติ  $k=0$  คือ

$$E^0 = avgEF_{y,m}^0,$$

หรือ

$$E^0 = avgEF_{y,m}^0 = \frac{\sum EF_{y,m}^0}{n_1^0 + n_2^0 + \dots + n_{12}^0}.$$

จากสมการ (3.5) ในขั้นตอนที่ 2.3 จะได้

$$E^0 = \frac{\sum \left( \frac{\bar{x}_{y,m}^0}{\bar{x}_m^0} \right)}{n_1^0 + n_2^0 + \dots + n_{12}^0},$$

หรือ

$$E^0 = \frac{\sum \left( \frac{\bar{x}_{y,1}^0}{\bar{x}_1^0} \right) + \sum \left( \frac{\bar{x}_{y,2}^0}{\bar{x}_2^0} \right) + \dots + \sum \left( \frac{\bar{x}_{y,12}^0}{\bar{x}_{12}^0} \right)}{n_1^0 + n_2^0 + \dots + n_{12}^0}.$$

จากสมการ (3.4) ในขั้นตอนที่ 2.2 จะได้

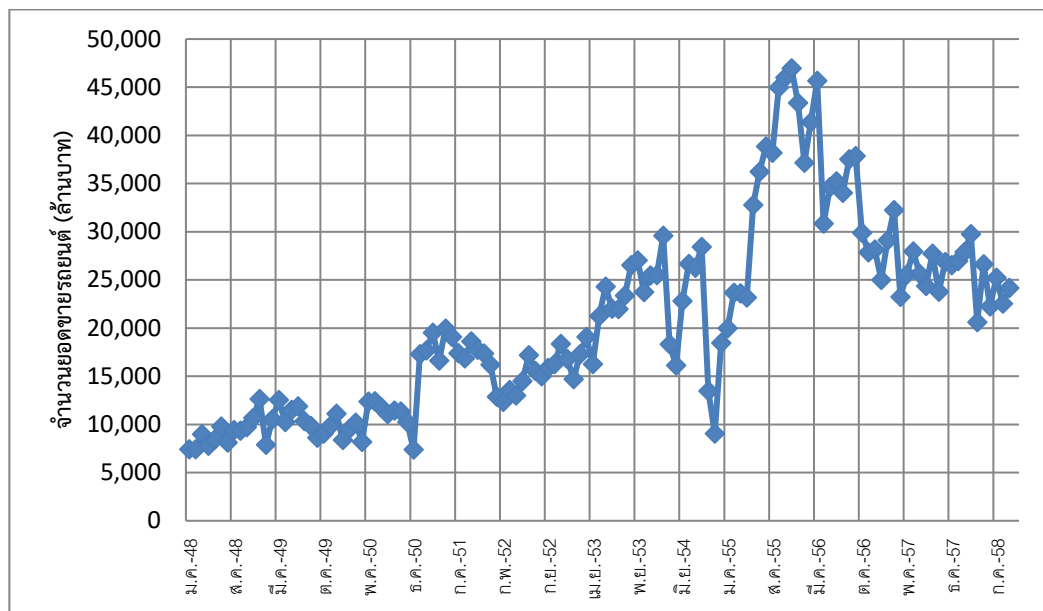
$$\begin{aligned}
 E^0 &= \frac{\sum \left( \frac{\bar{X}_{y,1}^0}{avg \bar{X}_{y,1}^0} \right) + \sum \left( \frac{\bar{X}_{y,2}^0}{avg \bar{X}_{y,2}^0} \right) + \dots + \sum \left( \frac{\bar{X}_{y,12}^0}{avg \bar{X}_{y,12}^0} \right)}{n_1^0 + n_2^0 + \dots + n_{12}^0} \\
 &= \frac{\left( \frac{\sum \bar{X}_{y,1}^0}{(\sum \bar{X}_{y,1}^0) / n_1^0} \right) + \left( \frac{\sum \bar{X}_{y,2}^0}{(\sum \bar{X}_{y,2}^0) / n_2^0} \right) + \dots + \left( \frac{\sum \bar{X}_{y,12}^0}{(\sum \bar{X}_{y,12}^0) / n_{12}^0} \right)}{n_1^0 + n_2^0 + \dots + n_{12}^0},
 \end{aligned}$$

$$= \frac{\left(\frac{\sum \bar{X}_{y,1}^0}{\sum \bar{X}_{y,1}^0}\right) n_1^0 + \left(\frac{\sum \bar{X}_{y,2}^0}{\sum \bar{X}_{y,2}^0}\right) n_2^0 + \dots + \left(\frac{\sum \bar{X}_{y,12}^0}{\sum \bar{X}_{y,12}^0}\right) n_{12}^0}{n_1^0 + n_2^0 + \dots + n_{12}^0}$$

ดังนั้น  $E^0 = 1$ . □

## 4.2 ตัวแบบพยากรณ์จากการวิเคราะห์การถดถอยเชิงเส้นแบบพหุ

พิจารณาข้อมูลยอดขายรถยนต์นั่งขนาดเล็กมากในประเทศไทยย้อนหลัง 10 ปี ตั้งแต่เดือนมกราคม 2548 ถึงเดือนกันยายน 2558 ซึ่งแสดงความสัมพันธ์แบบกราฟได้ดังรูปที่ 4.4



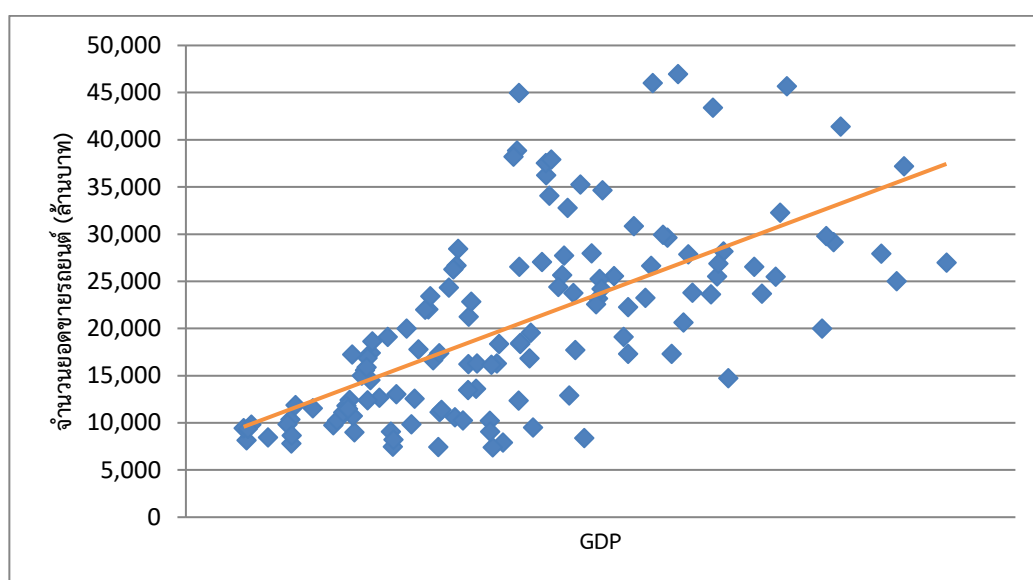
รูปที่ 4.4 กราฟแสดงยอดขายรถยนต์นั่งขนาดเล็กมากรายเดือนของประเทศไทย ระหว่างเดือนมกราคม 2548 ถึงเดือนกันยายน 2558

จากงานวิจัยที่มีการศึกษาเกี่ยวกับยอดขายรถยนต์ จะพบว่าในภาพรวมส่วนใหญ่แล้ว ยอดขายรถยนต์จะขึ้นอยู่กับค่าผลิตภัณฑ์มวลรวมในประเทศ (GDP) และอัตราดอกเบี้ยเงินกู้ 12 เดือน (Loan's Interest Rate) ดังนั้นในงานวิจัยนี้จึงทำการวิเคราะห์ความสัมพันธ์ของสองตัวแปรนี้ร่วมด้วย โดยพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ของยอดขายรถยนต์กับค่า GDP และอัตราดอกเบี้ยเงินกู้ 12 เดือน ดังตาราง 4.3

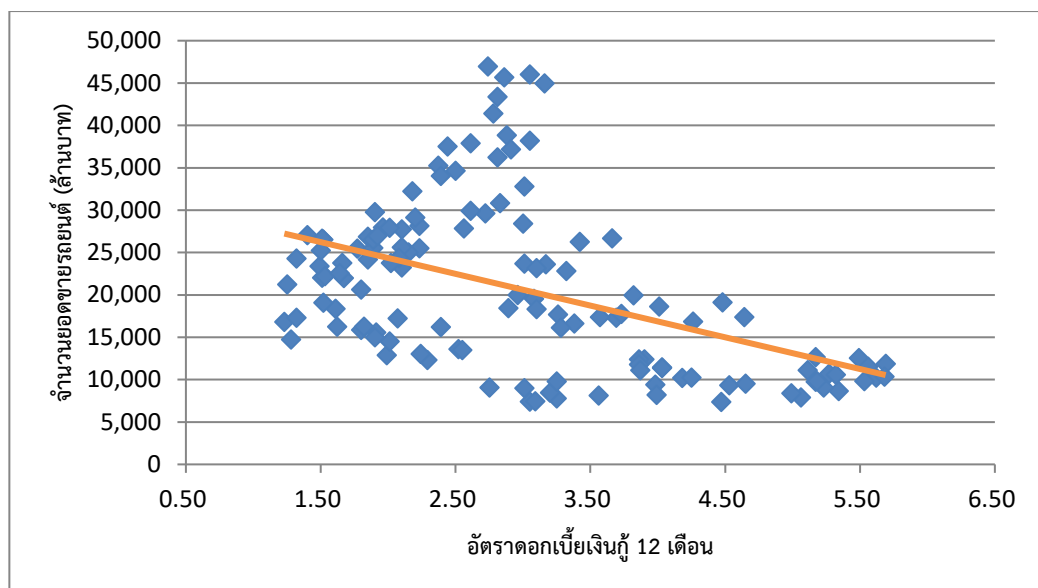
ตาราง 4.3 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของยอดขายรถยนต์กับค่า GDP และอัตราดอกเบี้ย

| ค่าสัมประสิทธิ์สหสัมพันธ์ | ยอดขายรถยนต์ | GDP      | อัตราดอกเบี้ยเงินกู้ 12 เดือน |
|---------------------------|--------------|----------|-------------------------------|
| ยอดขายรถยนต์              | 1.00000      |          |                               |
| GDP                       | 0.64258      | 1.00000  |                               |
| อัตราดอกเบี้ย             | -0.46425     | -0.51391 | 1.00000                       |

จากตาราง 4.3 จะพบว่ายอดขายรถยนต์และ GDP มีความสัมพันธ์ในทิศทางเดียวกัน และมีความสัมพันธ์กันในระดับสูง ส่วนยอดขายรถยนต์และอัตราดอกเบี้ยเงินกู้ 12 เดือน มีความสัมพันธ์ในทิศทางตรงกันข้ามและมีความสัมพันธ์กันในระดับปานกลาง ดังแสดงในรูปที่ 4.5 และรูปที่ 4.6 ตามลำดับ



รูปที่ 4.5 แผนภาพการกระจายข้อมูลระหว่าง GDP และ ยอดขายรถยนต์ (ระหว่างเดือนมกราคม 2548 - กันยายน 2558)



รูปที่ 4.6 แผนภาพการกระจายข้อมูลระหว่างอัตราดอกเบี้ยกับยอดขายรถยนต์  
(ระหว่างเดือนมกราคม 2548 – กันยายน 2558)

เมื่อทำการวิเคราะห์ข้อมูลยอดขายรถยนต์ กับ GDP และอัตราดอกเบี้ยเงินกู้ 12 เดือน โดยใช้คำสั่งใน Data Analysis ของโปรแกรม Microsoft Excel จะได้ค่าความสัมพันธ์ดังตาราง 4.4

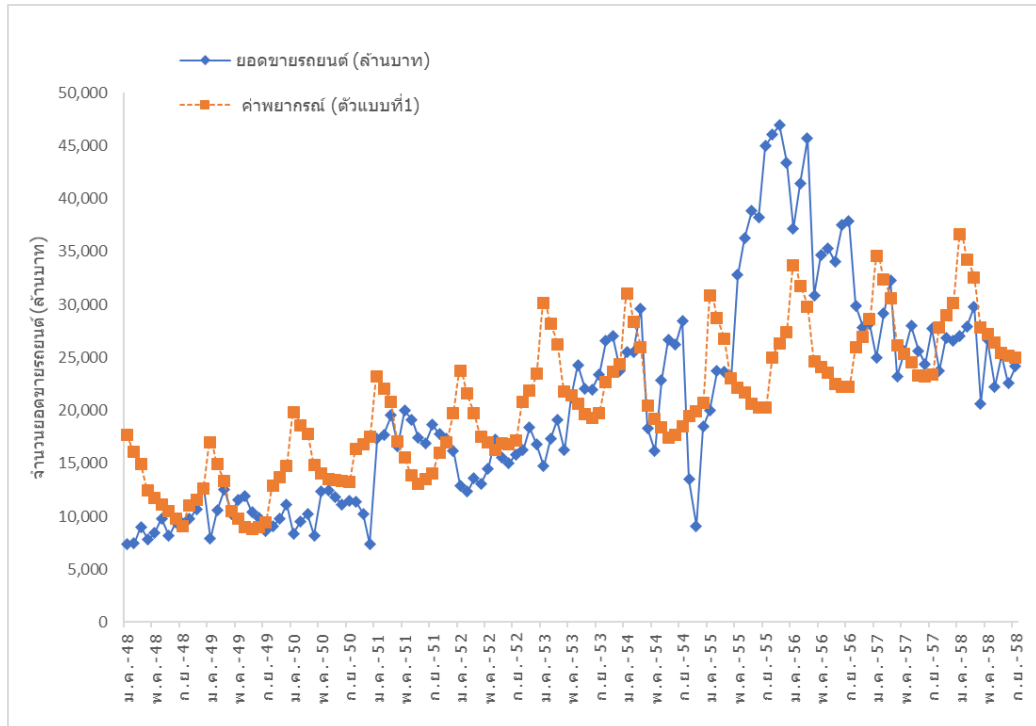
ตาราง 4.4 แสดงค่าสัมประสิทธิ์การถดถอยของตัวแบบพยากรณ์ถดถอยเชิงเส้นแบบพหุ  
ตัวแบบที่ 1

| ตัวแปร       | จุดตัดแกน Y | GDP    | อัตราดอกเบี้ย |
|--------------|-------------|--------|---------------|
| สัมประสิทธิ์ | -17571.8786 | 0.0624 | -1468.6843    |

จากตาราง 4.4 จะได้ตัวแบบพยากรณ์ถดถอยเชิงเส้นแบบพหุ ตัวแบบที่ 1 ซึ่งเป็นตัวแบบการถดถอยเชิงเส้นแบบพหุที่แสดงความสัมพันธ์ระหว่างยอดขายกับ GDP และอัตราดอกเบี้ยเงินกู้ 12 เดือน คือ

$$Y = (-17,571.8786) + (0.0624) \times (\text{GDP}) + (-1,468.6843) \times (\text{อัตราดอกเบี้ย}) \quad (4.5)$$

เมื่อนำค่าพยากรณ์ที่ได้จากตัวแบบที่ 1 ตามสมการ (4.5) มาเปรียบเทียบกับค่าจริงของยอดขายรถยนต์ แสดงได้ดังรูปที่ 4.7



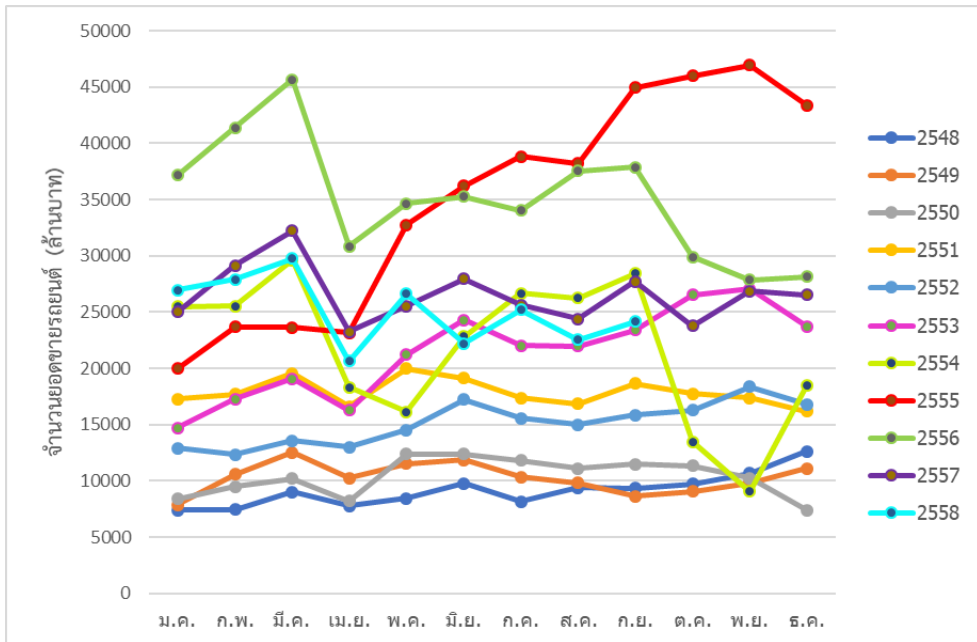
รูปที่ 4.7 กราฟเปรียบเทียบยอดขายรถยนต์กับค่าที่ได้จากตัวแบบการพยากรณ์ที่ใช้ GDP และอัตราดอกเบี้ย

ค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์ (MAPE) และรากที่สองของความคลาดเคลื่อนกำลังสอง (RMSE) ของตัวแบบที่ 1 ที่เป็นตัวแบบการถดถอยเชิงเส้นแบบพหุที่แสดงความสัมพันธ์ระหว่างยอดขายกับ GDP และอัตราดอกเบี้ยเงินกู้ 12 เดือน ตามสมการ (4.5) คือ

$$\begin{aligned}
 MAPE &= \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \times 100 \\
 &= \frac{1}{129} \sum_{i=1}^{129} \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \times 100 \quad (4.6) \\
 &= \frac{1}{129} (3,847.2527) \\
 &= 29.8236\%
 \end{aligned}$$

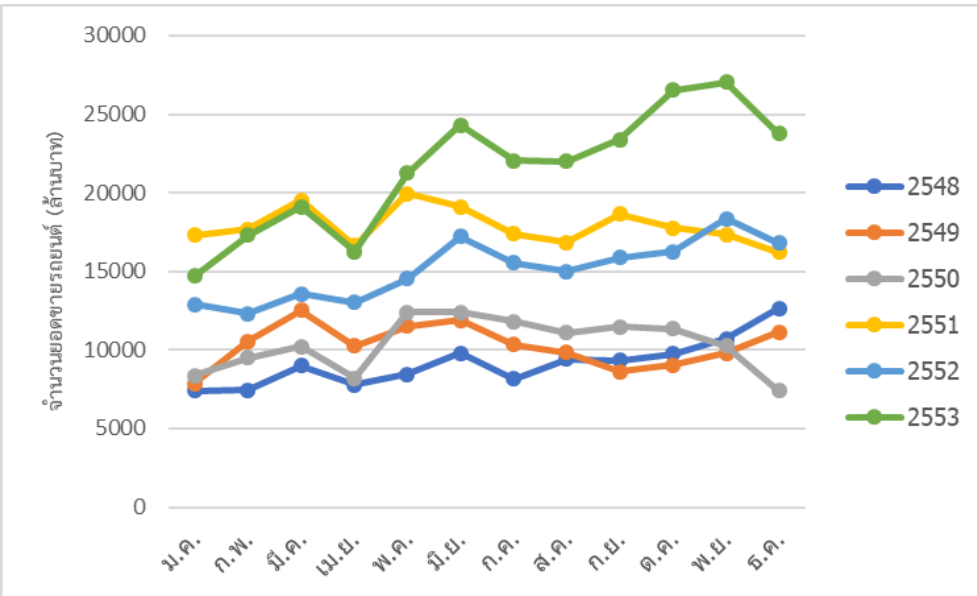
$$\begin{aligned}
 RMSE &= \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}} \\
 &= \sqrt{\frac{6,985,418,228}{129}} \quad (4.7) \\
 &= \sqrt{54,150,528.9} \\
 &= 7,358.70
 \end{aligned}$$

เมื่อพิจารณาข้อมูลยอดขายรถยนต์จากรูปที่ 4.3 จะพบว่าข้อมูลยอดขายรถยนต์มีแนวโน้มเชิงฤดูกาลเข้ามาเกี่ยวข้อง ซึ่งจะพบว่าแนวโน้มยอดขายจะสูงขึ้นในช่วงเดือนมีนาคม มิถุนายน และพฤศจิกายน ของทุกปี ดังรูปที่ 4.8



รูปที่ 4.8 ข้อมูลเปรียบเทียบยอดขายรถยนต์รายเดือน ปี 2548-2558

หากพิจารณาข้อมูลย้อนหลังในปี 2548 ถึงปี 2553 ที่ยังไม่เกิดเหตุการณ์น้ำท่วมใหญ่ในประเทศไทยและรัฐบาลยังไม่มึนโยบายมาตรการการค้เงินภาษีรถยนต์คันแรกมาเกี่ยวข้อง จะเห็นได้ว่าแนวโน้มเชิงฤดูกาลของข้อมูลมีความชัดเจนมากยิ่งขึ้น ดังรูปต่อไปนี้



รูปที่ 4.9 ข้อมูลเปรียบเทียบยอดขายรถยนต์รายเดือน ปี 2548-2553



จากนั้นทำการสร้างตัวแบบที่ 2 ที่เป็นตัวแบบการถดถอยเชิงเส้นแบบพหุแสดงความสัมพันธ์ระหว่างยอดขายรถยนต์กับ GDP และอัตราดอกเบี้ยเงินกู้ 12 เดือน โดยเพิ่มปัจจัยฤดูกาลใช้ข้อมูลตามตาราง 4.6

ตาราง 4.6 แสดงข้อมูลยอดขายรถยนต์รายเดือน GDP อัตราดอกเบี้ยและปัจจัยฤดูกาล

| เดือน    | ยอดขายรถยนต์<br>(ล้านบาท) | GDP        | อัตราดอกเบี้ย | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|----------|---------------------------|------------|---------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| ม.ค.-48  | 7,417                     | 636,870.12 | 3.05          | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   |
| ก.พ.-48  | 7,440                     | 612,185.23 | 3.09          | 1   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   |
| มี.ค.-48 | 8,995                     | 591,359.13 | 3.01          | 0   | 1   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   |
| เม.ย.-48 | 7,793                     | 557,133.71 | 3.25          | 0   | 0   | 1   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   |
| พ.ค.-48  | 8,455                     | 544,419.04 | 3.20          | 0   | 0   | 0   | 1   | 0   | 0   | 0   | 0   | 0   | 0   | 0   |
| มิ.ย.-48 | 9,784                     | 535,447.22 | 3.25          | 0   | 0   | 0   | 0   | 1   | 0   | 0   | 0   | 0   | 0   | 0   |
| ก.ค.-48  | 8,139                     | 532,801.46 | 3.56          | 0   | 0   | 0   | 0   | 0   | 1   | 0   | 0   | 0   | 0   | 0   |
| ส.ค.-48  | 9,418                     | 531,308.09 | 3.98          | 0   | 0   | 0   | 0   | 0   | 0   | 1   | 0   | 0   | 0   | 0   |
| ก.ย.-48  | 9,349                     | 533,575.80 | 4.53          | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 1   | 0   | 0   | 0   |
| ต.ค.-48  | 9,739                     | 579,896.20 | 5.17          | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 1   | 0   | 0   |
| พ.ย.-48  | 10,700                    | 590,417.07 | 5.27          | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 1   | 0   |
| ธ.ค.-48  | 12,641                    | 604,979.86 | 5.17          | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 1   |
| ม.ค.-49  | 7,885                     | 671,961.88 | 5.06          | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   |
| .        |                           |            |               |     |     | .   |     |     |     |     |     |     |     | .   |
| .        |                           |            |               |     |     | .   |     |     |     |     |     |     |     | .   |
| .        |                           |            |               |     |     | .   |     |     |     |     |     |     |     | .   |
| เม.ย.-58 | 20,636                    | 769,887.06 | 1.80          | 0   | 0   | 1   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   |
| พ.ค.-58  | 26,626                    | 752,317.02 | 1.51          | 0   | 0   | 0   | 1   | 0   | 0   | 0   | 0   | 0   | 0   | 0   |
| มิ.ย.-58 | 22,234                    | 739,919.12 | 1.53          | 0   | 0   | 0   | 0   | 1   | 0   | 0   | 0   | 0   | 0   | 0   |
| ก.ค.-58  | 25,225                    | 724,452.65 | 1.50          | 0   | 0   | 0   | 0   | 0   | 1   | 0   | 0   | 0   | 0   | 0   |
| ส.ค.-58  | 22,543                    | 722,422.10 | 1.64          | 0   | 0   | 0   | 0   | 0   | 0   | 1   | 0   | 0   | 0   | 0   |
| ก.ย.-58  | 24,173                    | 725,505.52 | 1.85          | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 1   | 0   | 0   | 0   |

และเมื่อทำการวิเคราะห์ข้อมูลยอดขายรถยนต์ กับ GDP และอัตราดอกเบี้ยเงินกู้ 12 เดือน เมื่อเพิ่มปัจจัยฤดูกาล ซึ่งเป็นตัวแปรอิสระเชิงคุณภาพ โดยใช้คำสั่งใน Data Analysis ของโปรแกรม Microsoft Excel จะได้ค่าความสัมพันธ์ดังตาราง 4.7

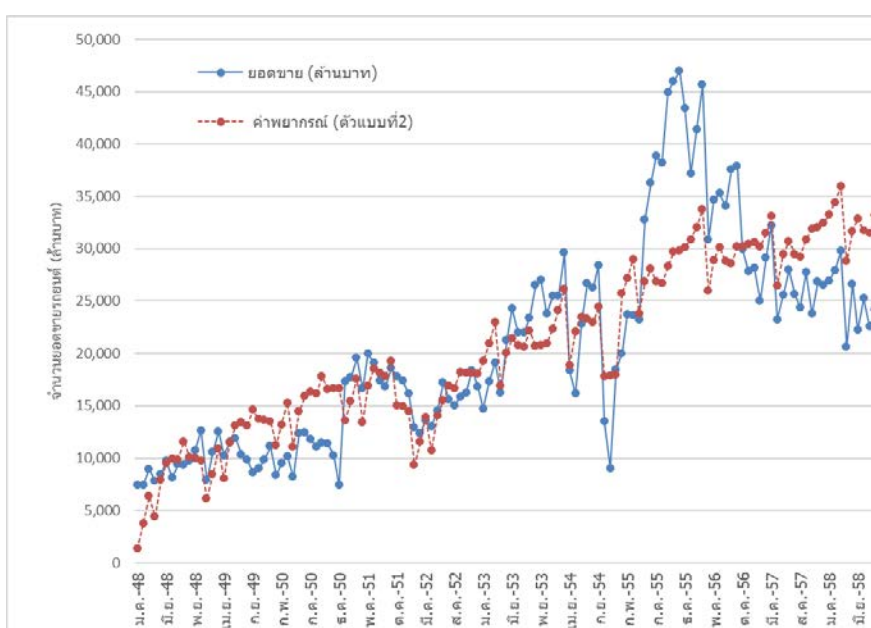
ตาราง 4.7 แสดงค่าสัมประสิทธิ์การถดถอยของตัวแบบพยากรณ์ถดถอยเชิงเส้นแบบพหุ  
ตัวแบบที่ 2

| ตัวแปร        | สัมประสิทธิ์ | ตัวแปร | สัมประสิทธิ์ |
|---------------|--------------|--------|--------------|
| จุดตัดแกน Y   | -74045.81    | Jun    | 19,930.63    |
| GDP           | 0.12         | July   | 20,565.18    |
| อัตราดอกเบี้ย | 315.07       | Aug    | 20,534.25    |
| Feb           | 5,290.07     | Sep    | 21,798.43    |
| Mar           | 10,309.01    | Oct    | 14,704.91    |
| Apr           | 12,265.16    | Nov    | 13,332.29    |
| May           | 17,295.46    | Dec    | 11,415.12    |

จากตาราง 4.7 จะได้ตัวแบบที่ 2 ซึ่งเป็นตัวแบบการถดถอยเชิงเส้นแบบพหุที่แสดงความสัมพันธ์ระหว่างยอดขายกับ GDP และอัตราดอกเบี้ยเงินกู้ 12 เดือน เมื่อเพิ่มปัจจัยฤดูกาลคือ

$$\begin{aligned}
 Y = & (-74,045.81) + (0.1169) \times (\text{GDP}) + (315.0701) \times (\text{อัตราดอกเบี้ย}) \\
 & + (5,290.0704) \times (\text{Feb}) + (10,309.0051) \times (\text{Mar}) + (12,265.1558) \times (\text{Apr}) \\
 & + (17,295.4577) \times (\text{May}) + (19,930.6251) \times (\text{Jun}) + (20,565.1789) \times (\text{Jul}) \\
 & + (20,534.2476) \times (\text{Aug}) + (21,798.4298) \times (\text{Sep}) + (14,704.9074) \times (\text{Oct}) \\
 & + (13,332.2902) \times (\text{Nov}) + (11,415.1192) \times (\text{Dec}) \quad (4.8)
 \end{aligned}$$

เมื่อนำค่าพยากรณ์ที่ได้จากตัวแบบที่ใช้ GDP อัตราดอกเบี้ยเงินกู้ 12 เดือน และปัจจัยฤดูกาล ดังในสมการ (4.8) มาเปรียบเทียบกับค่าจริงของยอดขายรถยนต์ แสดงได้ดังรูปที่ 4.11



รูปที่ 4.11 กราฟเปรียบเทียบยอดขายรถยนต์กับค่าที่ได้จากตัวแบบการพยากรณ์ที่ใช้ GDP และอัตราดอกเบี้ย เมื่อเพิ่มปัจจัยฤดูกาล

ค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์ (MAPE) และรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของตัวแบบที่ 2 ที่เป็นตัวแบบการถดถอยเชิงเส้นแบบพหุ ที่แสดงความสัมพันธ์ระหว่างยอดขายกับ GDP และอัตราดอกเบี้ยเงินกู้ เมื่อเพิ่มปัจจัยฤดูกาลตามสมการ (4.8) คือ

$$\begin{aligned} \text{MAPE} &= \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \times 100 \\ &= \frac{1}{129} \sum_{i=1}^{129} \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \times 100 \\ &= \frac{1}{129} (2,842.8789) \\ &= 22.0378\% \end{aligned} \quad (4.9)$$

$$\begin{aligned} \text{RMSE} &= \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}} \\ &= \sqrt{\frac{3,719,671,485.25208}{129}} \\ &= \sqrt{28,834,662.6763727} \\ &= 5,369.79167 \end{aligned} \quad (4.10)$$

จากข้อมูลยอดขายรถยนต์ที่แสดงในรูปที่ 4.3 จะสังเกตได้ว่าข้อมูลยอดขายรถยนต์มีความผันผวนในช่วงเดือนกันยายน 2554 จนถึงเดือนธันวาคม 2556 ซึ่งในช่วงเวลาดังกล่าวประเทศไทยมีเหตุการณ์ที่ทำให้เกิดผลกระทบต่ออุตสาหกรรมยานยนต์และอุตสาหกรรมอื่น ๆ อันส่งผลต่อยอดขายรถยนต์อย่างเห็นได้ชัด นั่นคือสถานการณ์น้ำท่วมครั้งใหญ่ในประเทศช่วงเดือนกันยายนถึงเดือนธันวาคม 2554 ทำให้ยอดขายรถยนต์ลดลง และลดลงมากในช่วงเดือนตุลาคม 2554 ด้วยเหตุที่ยอดขายรถยนต์ลดลงจึงส่งผลกระทบต่ออุตสาหกรรมยานยนต์อย่างยิ่ง ทำให้รัฐบาลได้ประกาศใช้นโยบายกระตุ้นเศรษฐกิจโดยการสนับสนุนให้ผู้ไม่มีรถยนต์เป็นของตนเอง สามารถใช้สิทธิในการซื้อรถยนต์คันแรก และออกมาตรการคืนเงินภาษีรถยนต์คันแรก สำหรับผู้ถือครองรถยนต์ครบ 1 ปี รายละไม่เกิน 100,000 บาท ทำให้ยอดขายรถยนต์รวมถึงยอดจองรถยนต์มีจำนวนเพิ่มขึ้นอย่างเห็นได้ชัดในช่วงเดือนธันวาคม 2554 ถึงเดือนมิถุนายน 2555 และสถานการณ์ยอดขายรถยนต์กลับมาเป็นปกติภายหลังหมดระยะเวลาของโครงการดังกล่าวนี้ ด้วยเหตุนี้ตัวแบบการพยากรณ์ยอดขายรถยนต์แบบที่ 3 จึงได้เพิ่มปัจจัยเกี่ยวกับสถานการณ์พิเศษให้เป็นตัวแปรอิสระเชิงคุณภาพอีกกลุ่มหนึ่ง โดยกำหนดเป็นตัวแปรหุ่นตามช่วงเวลาที่เกิดสถานการณ์นั้น ๆ ดังนี้

$e_1$  คือ ช่วงเวลาที่เกิดเหตุน้ำท่วมใหญ่ (กันยายน 2554 – พฤศจิกายน 2554)

$e_2$  คือ ช่วงเวลาที่เกิดน้ำท่วม และรัฐบาลได้ออกมาตรการคืนเงินภาษีรถยนต์คันแรก (ธันวาคม 2554 – มิถุนายน 2555)

$e_3$  คือ ช่วงเวลาที่มีมาตรการคืนเงินภาษีรถยนต์คันแรก

(กรกฎาคม 2555 – ธันวาคม 2555)

$e_4$  คือ ช่วงเวลาหลังจากหมดสถานการณ์พิเศษนี้แล้ว

(มกราคม 2556 – มกราคม 2557)

และกำหนดรหัสตัวเลขให้กับตัวแปรสถานการณ์พิเศษ ดังตาราง 4.8

ตาราง 4.8 ตารางกำหนดรหัสตัวเลขสำหรับตัวแปรปัจจัยสถานการณ์พิเศษ

| เดือน    | e1 | e2 | e3 | e4 | เดือน    | e1 | e2 | e3 | e4 |
|----------|----|----|----|----|----------|----|----|----|----|
| ก.ย.-54  | 1  | 0  | 0  | 0  | ธ.ค.-55  | 0  | 0  | 1  | 0  |
| ต.ค.-54  | 1  | 0  | 0  | 0  | ม.ค.-56  | 0  | 0  | 0  | 1  |
| พ.ย.-54  | 1  | 0  | 0  | 0  | ก.พ.-56  | 0  | 0  | 0  | 1  |
| ธ.ค.-54  | 0  | 1  | 0  | 0  | มี.ค.-56 | 0  | 0  | 0  | 1  |
| ม.ค.-55  | 0  | 1  | 0  | 0  | เม.ย.-56 | 0  | 0  | 0  | 1  |
| ก.พ.-55  | 0  | 1  | 0  | 0  | พ.ค.-56  | 0  | 0  | 0  | 1  |
| มี.ค.-55 | 0  | 1  | 0  | 0  | มิ.ย.-56 | 0  | 0  | 0  | 1  |
| เม.ย.-55 | 0  | 1  | 0  | 0  | ก.ค.-56  | 0  | 0  | 0  | 1  |
| พ.ค.-55  | 0  | 1  | 0  | 0  | ส.ค.-56  | 0  | 0  | 0  | 1  |
| มิ.ย.-55 | 0  | 1  | 0  | 0  | ก.ย.-56  | 0  | 0  | 0  | 1  |
| ก.ค.-55  | 0  | 0  | 1  | 0  | ต.ค.-56  | 0  | 0  | 0  | 1  |
| ส.ค.-55  | 0  | 0  | 1  | 0  | พ.ย.-56  | 0  | 0  | 0  | 1  |
| ก.ย.-55  | 0  | 0  | 1  | 0  | ธ.ค.-56  | 0  | 0  | 0  | 1  |
| ต.ค.-55  | 0  | 0  | 1  | 0  | ม.ค.-57  | 0  | 0  | 0  | 1  |
| พ.ย.-55  | 0  | 0  | 1  | 0  | ก.พ.-57  | 0  | 0  | 0  | 0  |

จากนั้นทำการสร้างตัวแบบที่ 3 ที่เป็นตัวแบบการถดถอยเชิงเส้นแบบพหุแสดงความสัมพันธ์ระหว่างยอดขายรถยนต์กับ GDP และอัตราดอกเบี้ยเงินกู้ 12 เดือน โดยเพิ่มปัจจัยฤดูกาลและสถานการณ์พิเศษ โดยใช้ข้อมูลดังตาราง 4.9



เมื่อทำการวิเคราะห์ข้อมูลด้วยโปรแกรมสำเร็จรูป จะได้ค่าความสัมพันธ์ดังตาราง 4.10

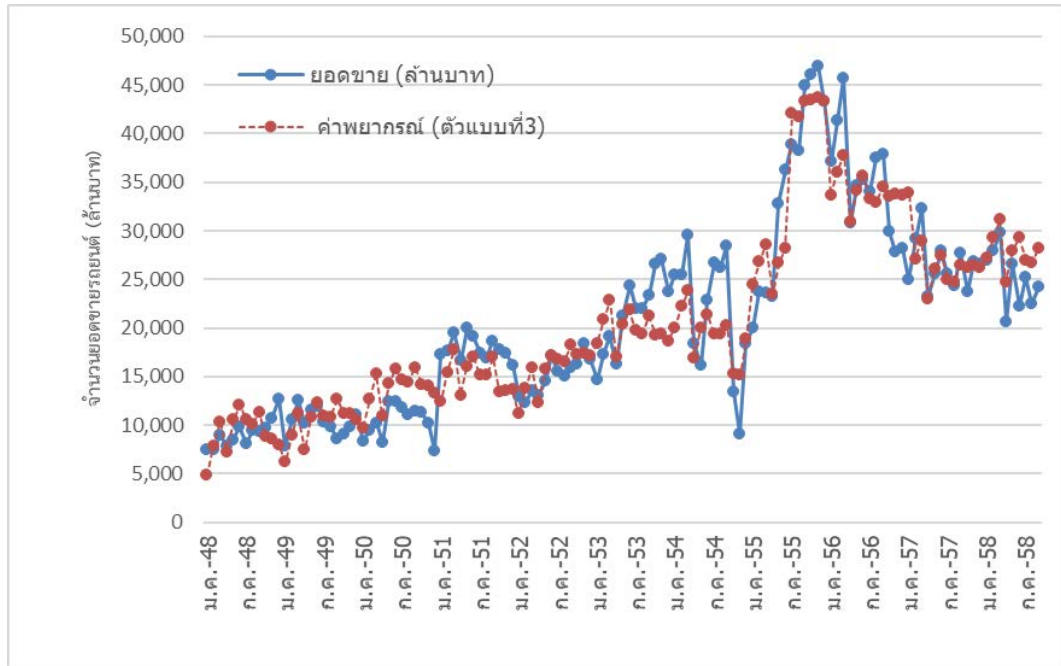
ตาราง 4.10 แสดงค่าสัมประสิทธิ์การถดถอยของตัวแบบพยากรณ์ถดถอยเชิงเส้นแบบพหุ  
ตัวแบบที่ 3

| ตัวแปร        | สัมประสิทธิ์ | ตัวแปร | สัมประสิทธิ์ |
|---------------|--------------|--------|--------------|
| จุดตัดแกน Y   | -42,685.207  | Aug    | 14,136.974   |
| GDP           | 0.078        | Sep    | 15,612.827   |
| อัตราดอกเบี้ย | -698.495     | Oct    | 9,911.320    |
| Feb           | 4,966.321    | Nov    | 8,956.426    |
| Mar           | 9,035.212    | Dec    | 7,121.731    |
| Apr           | 8,674.537    | e1     | -1,098.794   |
| May           | 13,013.761   | e2     | 3,282.933    |
| Jun           | 15,300.052   | e3     | 19,562.705   |
| Jul           | 14,222.899   | e4     | 8,988.843    |

จากตาราง 4.10 จะได้ตัวแบบที่ 3 ซึ่งเป็นตัวแบบการถดถอยเชิงเส้นแบบพหุที่แสดงความสัมพันธ์ระหว่างยอดขายกับ GDP อัตราดอกเบี้ยเงินกู้ 12 เดือน เมื่อเพิ่มปัจจัยฤดูกาลและสถานการณ์พิเศษ คือ

$$\begin{aligned}
 Y = & (-42,685.207) + (0.078) \times (\text{GDP}) + (-698.495) \times (\text{อัตราดอกเบี้ย}) \\
 & + (4,966.321) \times (\text{Feb}) + (9,035.212) \times (\text{Mar}) + (8,674.537) \times (\text{Apr}) \\
 & + (13,013.761) \times (\text{May}) + (15,300.052) \times (\text{Jun}) + (14,222.899) \times (\text{Jul}) \\
 & + (14,136.974) \times (\text{Aug}) + (15,612.827) \times (\text{Sep}) + (9,911.320) \times (\text{Oct}) \\
 & + (8,956.426) \times (\text{Nov}) + (7,121.731) \times (\text{Dec}) + (-1,098.794) \times (e_1) \\
 & + (3,282.933) \times (e_2) + (19,562.705) \times (e_3) + (8,988.843) \times (e_4) \quad (4.11)
 \end{aligned}$$

เมื่อนำค่าพยากรณ์ที่ได้จากตัวแบบการถดถอยเชิงเส้นแบบพหุที่ใช้ GDP อัตราดอกเบี้ยเงินกู้ 12 เดือน ปัจจัยฤดูกาลและสถานการณ์พิเศษ ตามสมการ (4.11) มาเปรียบเทียบกับค่าจริงของยอดขายรถยนต์ จะได้ดังรูปที่ 4.12



รูปที่ 4.12 กราฟเปรียบเทียบยอดขายรถยนต์กับค่าที่ได้จากตัวแบบการพยากรณ์ที่ใช้ GDP และอัตราดอกเบี้ย เมื่อเพิ่มปัจจัยฤดูกาลและสถานการณ์พิเศษ

ค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์ (MAPE) และรากที่สองของความคลาดเคลื่อนกำลังสอง (RMSE) ของตัวแบบที่ 3 ที่เป็นตัวแบบการถดถอยเชิงเส้นแบบพหุ ที่แสดงความสัมพันธ์ระหว่างยอดขายกับ GDP และอัตราดอกเบี้ยเงินกู้ 12 เดือน เมื่อเพิ่มปัจจัยฤดูกาลและสถานการณ์พิเศษตามตัวแบบในสมการ (4.11) คือ

$$\begin{aligned}
 \text{MAPE} &= \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \times 100 \\
 &= \frac{1}{129} \sum_{i=1}^{129} \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \times 100 \\
 &= \frac{1}{129} (2,025.262) \\
 &= 15.6997\%
 \end{aligned} \tag{4.12}$$

$$\begin{aligned}
 \text{RMSE} &= \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}} \\
 &= \sqrt{\frac{1,520,363,166.2809}{129}} \\
 &= \sqrt{11,785,760.98} \\
 &= 3,433.0396
 \end{aligned} \tag{4.13}$$

เปรียบเทียบค่าทางสถิติของตัวแบบการถดถอยเชิงเส้นแบบพหุทั้ง 3 แบบ ดังตาราง 4.11

ตาราง 4.11 แสดงค่าทางสถิติของตัวแบบทั้ง 3 ตัวแบบ

|             | $R^2$  | MAPE   | RMSE     |
|-------------|--------|--------|----------|
| ตัวแบบที่ 1 | 42.84% | 29.82% | 7,358.70 |
| ตัวแบบที่ 2 | 66.65% | 22.04% | 5,369.79 |
| ตัวแบบที่ 3 | 85.90% | 15.70% | 3,433.04 |

จากตาราง 4.11 จะพบว่า ตัวแบบที่ 1 เป็นตัวแบบการถดถอยเชิงเส้นแบบพหุเพื่อพยากรณ์ยอดขายรถยนต์ที่ไม่มีปัจจัยฤดูกาลและสถานการณ์พิเศษ จะมีค่า  $R^2$  เพียง 42.84% และ MAPE สูงถึง 29.82% ตัวแบบที่ 2 เป็นตัวแบบการถดถอยเชิงเส้นแบบพหุที่นำปัจจัยฤดูกาลมาพิจารณาด้วย จะพบว่ามีค่า  $R^2$  ที่สูงขึ้นเป็น 66.65% และค่า MAPE ลดลงเป็น 22.04% และในตัวแบบที่ 3 ที่เป็นตัวแบบการถดถอยเชิงเส้นแบบพหุที่มีทั้งปัจจัยฤดูกาลและสถานการณ์พิเศษ แสดงให้เห็นว่าเป็นตัวแบบที่มีแนวโน้มการพยากรณ์ได้ดีที่สุด โดยมีค่า  $R^2$  ที่ปรับค่าสูงสุดถึง 85.90% และ MAPE ต่ำสุดที่ 15.70%

ดังนั้นปัจจัยสถานการณ์พิเศษจึงเป็นปัจจัยที่เหมาะสมสำหรับตัวแบบการถดถอยเชิงเส้นแบบพหุร่วมกับปัจจัยฤดูกาล GDP และอัตราดอกเบี้ยเงินกู้ 12 เดือน เนื่องจากเป็นปัจจัยที่ทำให้ตัวแบบที่ใช้ในการพยากรณ์ยอดขายรถยนต์มีประสิทธิภาพ และมีความแม่นยำมากขึ้น

พิจารณาตัวแบบการถดถอยเชิงเส้นแบบพหุตัวแบบที่ 3 ตามสมการ (4.11) จะพบว่าค่าสัมประสิทธิ์การถดถอย เป็นดังตาราง 4.10 ซึ่งจะพบว่าค่าสัมประสิทธิ์ของตัวแปรอิสระ 2 ตัวคือ GDP และอัตราดอกเบี้ยเงินกู้ 12 เดือน กับสัมประสิทธิ์ของตัวแปรปัจจัยฤดูกาล 11 ตัวแปร และสัมประสิทธิ์ของตัวแปรปัจจัยสถานการณ์พิเศษ 4 ตัวแปร มีทั้งค่าลบและค่าบวก ซึ่งอาจอธิบายได้ว่า การที่สัมประสิทธิ์ของอัตราดอกเบี้ยเงินกู้ 12 เดือน มีค่าเป็นลบ แสดงให้เห็นว่าแนวโน้มที่ประชาชนจะซื้อรถยนต์มีน้อยลง นั่นคือยอดขายรถยนต์มีแนวโน้มลดลงเมื่ออัตราดอกเบี้ยเงินกู้สูงขึ้น ในทางตรงข้ามค่า GDP ที่มีค่าเป็นบวกนั้นหมายถึงยอดขายรถยนต์จะมีแนวโน้มเพิ่มขึ้นเมื่อสถานะเศรษฐกิจดีขึ้นนั่นเอง ส่วนตัวแปรปัจจัยฤดูกาลทั้ง 11 ตัวแปร มีค่าเป็นบวก หากพิจารณาตามช่วงเวลาของเดือนก็จะพบว่าเดือนมกราคมถือเป็นฤดูกาลหรือช่วงเวลาที่แนวโน้มยอดขายรถยนต์ต่ำที่สุดของประเทศไทย นั่นอาจเป็นเพราะช่วงเวลานี้ประชาชนมีการใช้จ่ายเงินกับกิจกรรมอื่น ๆ ในช่วงเทศกาลคริสต์มาสและปีใหม่ นั่นเอง เช่นเดียวกับปัจจัยสถานการณ์ทั้ง 4 สถานการณ์ จะพบว่าสถานการณ์  $e_1$  ซึ่งเป็นช่วงที่เกิดสถานการณ์น้ำท่วม ก็มีแนวโน้มที่จะส่งผลให้ยอดขายรถยนต์ลดลงเช่นกัน แต่สำหรับสถานการณ์  $e_2$  และ  $e_3$  ซึ่งเป็นช่วงเวลาที่รัฐบาลได้ออกมาตรการกระตุ้นเศรษฐกิจซึ่งรวมถึงกระตุ้นยอดขายรถยนต์โดยการคืนเงินภาษีรถยนต์สำหรับผู้ซื้อรถยนต์คันแรกส่งผลให้ยอดขายรถยนต์มีแนวโน้มสูงขึ้น โดยเฉพาะอย่างยิ่งช่วงสถานการณ์  $e_3$  ที่นโยบายของรัฐบาลมีผลเต็มที่ ทำให้ยอดขายรถยนต์ถึงจุดสูงสุด ส่วนช่วงสถานการณ์  $e_4$  ซึ่งเป็นช่วงหลังจากสิ้นสุดนโยบาย

คืนเงินภาษีรถยนต์คันแรกแล้วนั้น สัมประสิทธิ์ของเหตุการณ์ยังมีค่าเป็นบวกอาจเนื่องมาจาก ประชาชนบางส่วนที่ยังไม่ได้ซื้อรถยนต์ในช่วงที่มีนโยบายคืนเงินภาษีรถยนต์คันแรกนั้น อาจจะมีการสำรองเงินไว้ซื้อรถยนต์ในภายหลังซึ่งจะส่งผลให้แนวโน้มยอดขายรถยนต์เพิ่มขึ้นและกลับเข้าสู่รูปแบบปกติเหมือนก่อนหน้าที่จะมีเหตุการณ์น้ำท่วมใหญ่และนโยบายกระตุ้นยอดขายรถยนต์นั่นเอง

ในการพยากรณ์รวมถึงการสร้างตัวแบบเพื่อพยากรณ์หรืออธิบายความสัมพันธ์ของตัวแปรต่าง ๆ นั้น ส่วนสำคัญอย่างหนึ่งที่จะเป็นการเพิ่มความน่าเชื่อถือและความแม่นยำของตัวแบบที่สร้างขึ้นได้ก็คือการพิจารณาค่าคลาดเคลื่อนของตัวแบบ ซึ่งถ้าตัวแบบมีความคลาดเคลื่อนน้อยนั้นก็หมายถึงตัวแบบที่ได้มีความแม่นยำและน่าเชื่อถือมากนั่นเอง

## บทที่ 5

# สรุปผลการวิจัยและข้อเสนอแนะ

### 5.1 สรุปผลการวิจัย

จากการพิจารณาตัวแบบพยากรณ์อนุกรมเวลาโดยใช้วิธีแยกส่วนประกอบ และตัวแบบการพยากรณ์ถดถอยเชิงเส้นแบบพหุโดยมีตัวแปรหลักคือ GDP และอัตราดอกเบี้ยเงินกู้ 12 เดือน เพื่ออธิบายความสัมพันธ์ของยอดขายรถยนต์ขนาดเล็กมากในประเทศไทย ในช่วงที่มีวิกฤติเหตุการณ์น้ำท่วมและนโยบายพิเศษของรัฐบาลเพื่อกระตุ้นอุตสาหกรรมยานยนต์ พบว่าเมื่อนำปัจจัยฤดูกาลและสถานการณ์พิเศษมาพิจารณาในการสร้างตัวแบบอนุกรมเวลาโดยวิธีแยกส่วนประกอบ มีค่า MAPE เท่ากับ 8.27% ซึ่งถือได้ว่ามีระดับความแม่นยำสูงกว่าตัวแบบพยากรณ์ถดถอยเชิงเส้นแบบพหุ ส่วนตัวแบบพยากรณ์ถดถอยเชิงเส้นแบบพหุนั้น ตัวแบบที่ 3 มีความแม่นยำในระดับสูงเมื่อพิจารณาเปรียบเทียบกับค่าทางสถิติของตัวแบบพยากรณ์ถดถอยเชิงเส้นทั้ง 3 ตัวแบบ ดังแสดงในตารางที่ 4.11 นั่นคือ มีค่า MAPE ต่ำสุดที่ 15.70% จากการศึกษาและการสร้างตัวแบบการพยากรณ์ทั้งสองวิธี อาจกล่าวได้ว่าปัจจัยฤดูกาลและสถานการณ์พิเศษนั้นส่งผลต่อยอดขายรถยนต์อย่างเห็นได้ชัด นั่นคือมียอดขายรถยนต์รวมถึงยอดการจองรถยนต์เพิ่มมากขึ้น และยังส่งผลต่อเนื่องจนถึงภายหลังจากที่สิ้นสุดสถานการณ์พิเศษนั้นด้วย นอกจากนี้ตัวแบบที่ได้จากการวิเคราะห์นี้อาจเป็นประโยชน์ต่อการเตรียมการรับกับสถานการณ์ในอนาคต เช่นหากมีการรายงานแจ้งเตือนจากกรมอุตุนิยมวิทยาถึงปริมาณน้ำฝนหรือปริมาณน้ำหลากในประเทศไทยซึ่งอาจส่งผลกระทบต่ออุตสาหกรรมยานยนต์ หรืออุตสาหกรรมอื่น ๆ ที่เกี่ยวข้อง รัฐบาลจะได้เตรียมการวางนโยบายรวมทั้งองค์กรบริษัทจะได้วางแผนการดำเนินงานในอนาคตต่อไปได้นั่นเอง

### 5.2 ข้อเสนอแนะ

ในการวิเคราะห์ตัวแบบเพื่ออธิบายความสัมพันธ์ของยอดขายรถยนต์ มีการใช้ปัจจัยหุ่นในการวิเคราะห์หลายตัวแปรซึ่งอาจจะส่งผลให้เกิดความยุ่งยากและเกิดข้อผิดพลาดในตัวแบบที่สร้างขึ้นได้ จึงควรระมัดระวังในกรณีที่มิปัจจัยหลาย ๆ ตัวแปรในตัวแบบนั้นด้วย และในการวิเคราะห์ตัวแบบพยากรณ์ อาจพิจารณาปัจจัยแปรผันเชิงคุณภาพอื่น ๆ ที่อาจแฝงอยู่ในข้อมูลอนุกรมเวลา นอกเหนือจาก GDP และอัตราดอกเบี้ยเงินกู้ 12 เดือน เช่น รายได้ครัวเรือน สถานการณ์ราคาน้ำมัน สถานการณ์การเมือง นโยบายของรัฐ เป็นต้น

## บรรณานุกรม

- กรมสรรพสามิต. 2555. **รถยนต์ใหม่คันแรก ตามนโยบายรัฐบาล.** Thailand e-Government. [Online]. Available : <http://www.egov.go.th/e-government-service/190/>
- กัลยา วานิชย์บัญชา. 2544. **การวิเคราะห์ตัวแปรหลายตัวด้วย SPSS.** พิมพ์ครั้งที่ 2. กรุงเทพฯ : โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- กัลยา วานิชย์บัญชา. 2545. **การวิเคราะห์สถิติ : สถิติสำหรับการบริหารและวิจัย.** พิมพ์ครั้งที่ 6. กรุงเทพฯ : โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- กลุ่มอุตสาหกรรมยานยนต์แห่งประเทศไทย (The Federation of Thailand Industries Automotive Industry Club : F.T.I). 2016. [Online]. Available : <http://www.aic.or.th>
- ณัฐกานต์ อติเรกสันต์. 2559. “การพยากรณ์ปริมาณความต้องการสินค้าสำหรับการหาปริมาณการสั่งซื้อที่เหมาะสม.” สหกิจศึกษา ปริญญาวิทยาศาสตรบัณฑิต (คณิตศาสตร์ประยุกต์) คณะวิทยาศาสตร์, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.
- ทรงศิริ แต่สมบัติ. 2548. **การวิเคราะห์การถดถอย.** พิมพ์ครั้งที่ 3. กรุงเทพฯ : สำนักพิมพ์มหาวิทยาลัยเกษตรศาสตร์.
- ทรงศิริ แต่สมบัติ. 2549. **การพยากรณ์เชิงปริมาณ.** กรุงเทพฯ : สำนักพิมพ์มหาวิทยาลัยเกษตรศาสตร์.
- ธิดาเดียว มยุรีสุวรรณค์. 2559. **การวิเคราะห์การถดถอย.** ขอนแก่น : บริษัทเพ็ญพรินดี จำกัด.
- มุกดา แม้นมินทร์. 2549. **อนุกรมเวลาและการพยากรณ์.** กรุงเทพฯ : สำนักพิมพ์ประกายพริก.
- สุนทรตรา จันทบุรี. 2016. **อุตสาหกรรมยานยนต์ : ไทยมีศักยภาพสูงสุดในอาเซียน.** ส่วนอาเซียน สำนักการประชาสัมพันธ์ต่างประเทศ กรมประชาสัมพันธ์. 24 พฤษภาคม 2559. [Online]. Available : <http://www.aseanhai.net>
- CIVIC ES Group, General talk. 2555. **การแบ่งประเภทรถยนต์ในประเทศไทย ที่คนไทยหลายคนยังไม่ทราบ.** CIVIC ES Group. 16 พฤษภาคม 2555. [Online]. Available : <http://www.civicesgroup.com/forum/topic42946>
- Bank of Thailand, "Thailand's Macro Economic Indicators," Bank of Thailand, 2015. [Online]. Available : <http://www2.bot.or.th/> [Accessed 15 December 2015].
- Brad M. Barber, Reid W. Click and Masako N. Darrough. 1999. “The impact of shocks to exchange rate and oil prices on U.S. sales of American and Japanese automakers.” *Japan and World Economy.* 11 : 57-93.

- Chartchai Leenawong and Ponrudee Netisopakul. 2016. "Forecasting Car Sales in Thailand With the Impact of the National Flood and the First-car Buyer Program." *International Conference on Business and Social Sciences (ICBASS)*. Conference Proceedings, March 29-31, Kyoto Japan : 1006-1007.
- Dean Fantazzini and Zhamal Toktamysova. 2015. "Forecasting German Car Sales Using Google Data and Multivariate Models." *International Journal of Production Economics*. 170 (Part A, December 2015) : 97-135.
- Fidlizan Muhammad, Mohd Yahya Mohd Hussin and Azila Ab Razak. 2012. "Automobile Sales and Macroeconomic Variables: A Pooled Mean Group Analysis for ASEAN Countries." *Journal of Business and Management (IOSRJBM)*. 2(1) : 15-21.
- Hamilton, J.D. 1988. "A neoclassical model of unemployment and the business cycle." *Journal of Political Economy*. 96 : 593-617.
- Joyce M. Dargay and Detmot Gately. 1999. "Income's effect on car and vehicle ownership, worldwide: 1960-2015." *Transportation Research Part A*. 33 : 101-138.
- Joyce M. Dargay. 2001. "The effect of income an car ownership: evidence of asymmetry." *Transportation Research Part A*. 35 : 807-821.
- Multiple Linear Regression Analysis.** Education to Empower The Reliability Profession. [Online]. Available : [http://reliawiki.org/index.php/Multiple\\_Linear\\_Regression\\_Analysis](http://reliawiki.org/index.php/Multiple_Linear_Regression_Analysis)
- Office of Industrial Economics, "INDUSTRIAL INDICES," Office of Industrial Economics, 2015. [Online]. Available : <http://www.oie.go.th/> [Accessed 15 December 2015].
- Pindyck, R.S. and Rotemberg, J.J. 1983. "Dynamic factor demands and the effects of energy price shocks." *American Economic Review*. 73 : 1066-1079.
- STAT 501 Regression Method.** PennState Eberly College of Science. The Pennsylvania State University. [Online]. Available : <https://onlinecourses.science.psu.edu/stat501/>
- Syed Shahabudin. 2009. 'Forecasting automobile sales.' *Management Research News*. 32(7) : 670-682.

- Sydney Ludvigson. 1998. "The Channel of Monetary Transmission to Demand: Evidence from the Market for Automobile Credit." *Journal of Money, Credit and Banking*. 30(3) : 365-383.
- Taweessin Apiwattanachai and Juta Pichitlamken. 2008. "Time-series Forecasting Model for Automobile Sales in Thailand." 298-303. In Thailand. **National Conference on Operations Research Network in Thailand 2008, July 24-25, 2008.**
- Watchareeporn Chaimongkol and Chutatip Tansathit. 2010. "Forecasting Model for Automobile Sales in Thailand." **Proceedings of The 2<sup>nd</sup> International Conference on Logistics (ICLT 2010), and Transport & The 1<sup>st</sup> International Conference on Business and Economics (ICBE 2010) "Managing Finance and Risk in Global Supply Chain.** New Zealand : Lincoln University. " [Online]. Available : <http://preet.sesolution.com/ict2010/Full%20Papers/Inventory%20and%20Warehouse%20Management/0179-Watchareeporn.pdf>
- Wilson, J.H. and B.Keating. 2009. **Business Forecasting with Forecast X.** McGraw-Hill Education.
- Wirotcheewan P., Kengpol A., Ishii K. and Shimada Y. 2011. "Modelling and Forecasting for Automotive Parts Demand of Foreign Markets on Thailand." *Asian International Journal of Science and Technology in Production and Manufacturing Engineering*. 4(1) : 1-13.
- Younger, Mary Sue. 1979. **A First Course in Linear Regression.** USA. PWS Publisher. Wadsworth, Inc.

ภาคผนวก

## THE EFFECTS OF SPECIAL EVENTS ON REGRESSION FOR SUBCOMPACT CAR SALES IN THAILAND

Witchaya Rattanametawee<sup>a</sup>, Chartchai Leenawong<sup>a\*</sup>,  
Ponrudee Netisopakul<sup>b</sup>

<sup>a</sup>Department of Mathematics, Faculty of Science,  
King Mongkut's Institute of Technology Ladkrabang,  
Bangkok, 10520 Thailand

<sup>b</sup>Knowledge Management and Knowledge Engineering Lab.,  
Faculty of Information Technology, King Mongkut's Institute of  
Technology Ladkrabang, Bangkok, 10520, Thailand

### Article history

Received

17 June 2016

Received in revised form

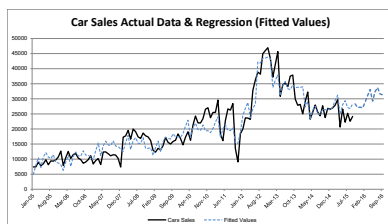
20 September 2016

Accepted

18 October 2016

\*Corresponding author  
chartchai.le@kmitl.ac.th

### Graphical abstract



### Abstract

This research proposes a method to dealing with multiple linear regression that integrates the seasonality as well as the effects of some special or unanticipated events for sales figures. The method is then applied to the car sales figures in Thailand after having been through the 2011 national big flood and the 2011-2012 government's initiative tax-incentive program for boosting the automobile industry. Besides Thailand's Gross Domestic Products (GDP) and the 12-month Loan's Interest Rate as explanatory variables, seasonal dummy variables along with the proposed special event variables and appropriate event tagging are incorporated. The statistical results obtained from the proposed regression model with seasons and events, compared to the models with neither seasons nor both yields highest adjusted coefficient of determination (R-square) and accuracy (MAPE).

Keywords: Multiple Regression, Forecasting Event Modelling, Car Sales, Thailand

© 2016 Penerbit UTM Press. All rights reserved

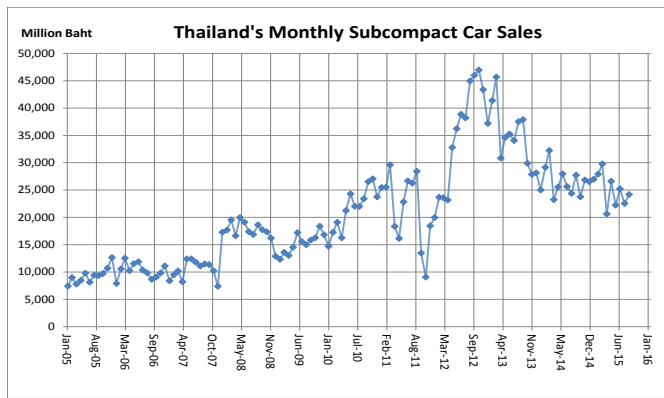
## 1.0 INTRODUCTION

In the past several years, subcompact car sales figures in Thailand have been fluctuating considerably due mostly to two crucial events. One is the severe flooding across the country that occurred approximately during the later months of the year 2011 and the other is the tax-incentive first-car buyer program [1] that followed just slightly later for another 1-plus year.

The program was supposed to support the almost devastated automobile industry by reimbursing car taxes to Thai people who reserved for buying their first-owned subcompact cars with not over 1,500-litre engines during the time period 16 September 2011 to 31 December 2012. The reimbursement limit was 100,000 baht per car per person.

The monthly subcompact car sales figures (in Million Baht) in Thailand from January 2005 to September 2015 obtained from the Office of Industrial Economics, Ministry of Industry, Thailand [2], [3] are shown in Figure 1. All four components of the time series [4], namely, trend, seasonal, cyclical, and irregular components are obviously revealed in the figure as well.

Multiple regression [5] has been used in model fitting and forecasting for over a century in various areas such as water resources planning [6], Korean box-office revenue [7], gold prices [8], or even automobile sales in the US [9] and Taiwan [10]. Recently, Thailand car sales forecasting [11] has been conducted using the time-series decomposition method with special event flags. As the name of the method suggested, it considered just one contributing factor, that is, time. Other possible economic causes have been discarded.



**Figure 1** Thailand's monthly car sales (in Million Baht) from January 2005 to September 2015 [2]

In this research, it is, therefore, aimed to find the economic causal factors that influence the dependent variable, i.e., subcompact car sales in Thailand. In addition, other dummy variables are added to the regression so as to account for the seasonality. Besides, the effects of unanticipated events such as the case of Thailand's big flood and the effects of some short-noticed promotional programs such as the case of the tax-incentive program will be accounted for via the proposed special (or unanticipated) event variables. The big flood and the tax-incentive program caused the subcompact car sales data to fluctuate extremely during the period from September 2011 up until January 2014 as seen also in Figure 1 Thailand's monthly car sales (in Million Baht) from January 2005 to September 2015 [2]

The rest of this paper is organized as followed. In Section 2, the methodology to use in this research including multiple linear regression, how to deal with seasonality, and how the special event variables are constructed, is described. In Section 3, the experimental design is explained in detail followed by the experimental results in Section 4. Finally, the conclusion of this research is provided in Section 5.

## 2.0 METHODOLOGY

In this section, the methodology to be used in this research as well as the experimental design are presented. The methodology includes multiple linear regression, seasonality in regression, and the proposed method for special events. Then, applying the above methodology to the real data, namely, Thailand's car sales figures is explained.

### 2.1 Multiple Linear Regression

Linear regression is a modeling approach for examining the linear relationship between a variable of interest called the dependent variable  $y$  and an independent variable  $x$  (simple regression) or a set of independent variables  $x_1, x_2, \dots, x_k$  (multiple regression)

that can explain the amount of variation in  $y$ . Equation (1) displays the probabilistic linear equation of multiple regression.

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_kx_k + \varepsilon \tag{1}$$

where

$k$  is the number of independent variables,  
 $\beta_i$  is the contribution of the independent variable  $x_i$ ,  
 and  $\varepsilon$  is the error or residual term.

The Ordinary Least Squares (OLS) method used for the parameter estimation minimizes the sum of squared errors for the sample. Subsequently,  $b_i$  is used in place of  $\beta_i$  and the deterministic prediction equation becomes Equation (2) as follows.

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k. \tag{2}$$

In our case here, the  $b_i$  coefficients are obtained through a statistical package.

Turning now to the choices of the independent variables, since GDP is a good indicator of a nation's overall economic condition [12], a better economy positively affects the overall consumption and definitely stimulates more car sales. In contrast, since many car buyers do not pay in cash, the interest rate of the loan to pay for the car negatively affects the buying decision [13]. The higher the loan's interest rate is, the lower the car sales will be. Confirmed by the correlation coefficients between the variables with the correct signs shown in Table 1, Thailand's GDP and the loan's interest rate are considered good predictors for the car sales.

Revealed also by Table 1, the two independent variables, GDP and the loan rate, are not highly linearly correlated to each other. Therefore, the two predictors can be included in the multiple regression models.

**Table 1** Correlation coefficient table

| Correlation      | Car Sales | GDP   | Loan Rate |
|------------------|-----------|-------|-----------|
| <b>Car Sales</b> | 1.00      | 0.64  | -0.46     |
| <b>GDP</b>       | 0.64      | 1.00  | -0.51     |
| <b>Loan Rate</b> | -0.46     | -0.51 | 1.00      |

### 2.2 Seasonality in Regression

For seasonal data such as our car sales figures here, seasonality [14] must be taken into account for. One could compute a seasonal index for each and every season and use these indices to deseasonalize all of the original data before undertaking regression.

However, in this research, another method of dealing with seasonality in regression [15] is employed. It is done so by adding to the regression  $n-1$  dummy variables, probably named  $s_2, s_3, \dots, s_n$ , for the  $n$  seasons. Each dummy variable  $s_i$  is treated as another

independent variable except that its values can only be either 1 if season  $i$  is observed or 0 otherwise. Thus, the regression equation for the data with  $k$  predictors and  $n$  seasons simply becomes Equation (3) shown here.

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k + a_2s_2 + a_3s_3 + \dots + a_ns_n \quad (3)$$

where  $a_i$  represents the coefficient of each seasonal variable.

Since we have monthly data here, 11 dummy variables are required; for example,  $s_2, s_3, \dots, s_{12}$ , with  $s_2$  equal to 1 if it is February,  $s_3$  equal to 1 if it is March. Apparently, when all the 11 seasonal dummy variables are equal to 0, it refers to January.

### 2.3 The Proposed Method for Special Events

During some sales periods, unexpected or short-noticed events that have enormous impacts on sales figures can happen. National disasters such as floods, Tsunami floods, earthquakes, and windstorms are extreme examples that can cause unguarded businesses millions of dollars. Other examples of special events are some short-noticed gigantic promotional campaigns having also massive impacts on some industries. It is easy just to remove the data affected by these special events from the regression model fitting. Nevertheless, keeping those data intact and finding a way to tag them appropriately will be more useful, especially if in the future, these events can somehow be predicted to happen again. Businesses then can prepare for what it is worth, either favorable or unfavorable to them.

Analogous to the dummy variables for seasonality, special event variables are created so that one variable refers to one specific event that can last for more than one period. For example,  $e_1$  refers to the flooding period that lasts from September through November 2011, resulting in  $e_1$  taking the values of 1's in these flooding months and 0's in all others. However, for one real gigantic event that can cause different impacts for different periods, that one event could be split into more than one sub-event and thus more than one event variable.

Consequently, the multiple regression equation having  $k$  predictors,  $n$  seasons, and  $m$  special events eventually becomes

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k + a_2s_2 + a_3s_3 + \dots + a_ns_n + \gamma_1e_1 + \gamma_2e_2 + \dots + \gamma_me_m \quad (4)$$

where  $\gamma_i$  represents the coefficient of each special event variable.

In this research, two real gigantic events that affect the car sales tremendously are the 2011 Thailand big flood and the government's tax-incentive first-car buyer program in 2011-2012 [1]. First came the big flood in the third quarter of 2011, then a few months later, the incentive program started. When the big

flood was starting to emerge, most people had not been impacted yet and therefore the car sales remained roughly normal. The panic commenced later after it was certain that the capital and financial city, Bangkok, would be flooded too. Even though the tax incentive program was already begun during the flood period, the car sales had not turned back high up as expected. Until the flood ended and the people turned calm, the incentive program worked its way up to the top before the promotional campaign ended. With all the situations explained above, four event variables are constructed, each of which represents a sub-event as follows.

- $e_1$  refers to the flood period alone.
- $e_2$  refers to the flood with the incentive program period.
- $e_3$  refers to the incentive program period alone.
- $e_4$  refers to the aftermath period.

### 2.4 The Experiments

The monthly data used in our multiple regression are collected [2], [3] from the period January 2005 to September 2015, totaling to 129 data. The original quarterly GDPs were modestly transformed into monthly by first extracting out their quarterly seasonal indices and then non-linear curve fitting them to obtain the monthly seasonal indices which eventually lead to the final monthly GDPs.

To summarize all the gathered data for multiple regression, we have the following.

- $y$  = Thailand's subcompact car sales
- $x_1$  = GDP
- $x_2$  = the loan's interest rate
- $feb, mar, apr, \dots, dec$  = eleven seasonal dummy variables for the observed months February through December
- $e_1, e_2, e_3, e_4$  = the four dummy event variables,

and so the regression equation can be constructed as shown in Equation (5).

$$y = b_0 + b_1x_1 + b_2x_2 + a_2feb + a_3mar + \dots + a_{12}dec + \gamma_1e_1 + \gamma_2e_2 + \gamma_3e_3 + \gamma_4e_4 \quad (5)$$

where  $\gamma_i$  represents the coefficient of each event variable.

The four special events are assigned to the relevant monthly data according to the values in Table 2. When  $e_i$  is 1, the event  $e_i$  is in effect; 0 otherwise. For example,  $e_3$  for the month July 2012 equals 1 simply referring to the period where the tax-incentive program was in full effect and the flood was already gone. Most car buyers just started to realize the benefits of this incentive program and thus kept pouring in car reservations for purchase before the program ended. Note also that besides these months in the table, the values of the four event variables in

all other months are equal to 0 meaning that normal situations are observed.

**Table 2** Event assignment to monthly data (either 0 or 1)

| Month  | e <sub>1</sub> | e <sub>2</sub> | e <sub>3</sub> | e <sub>4</sub> | Month  | e <sub>1</sub> | e <sub>2</sub> | e <sub>3</sub> | e <sub>4</sub> |
|--------|----------------|----------------|----------------|----------------|--------|----------------|----------------|----------------|----------------|
| Sep-11 | 1              | 0              | 0              | 0              | Nov-12 | 0              | 0              | 1              | 0              |
| Oct-11 | 1              | 0              | 0              | 0              | Dec-12 | 0              | 0              | 1              | 0              |
| Nov-11 | 1              | 0              | 0              | 0              | Jan-13 | 0              | 0              | 0              | 1              |
| Dec-11 | 0              | 1              | 0              | 0              | Feb-13 | 0              | 0              | 0              | 1              |
| Jan-12 | 0              | 1              | 0              | 0              | Mar-13 | 0              | 0              | 0              | 1              |
| Feb-12 | 0              | 1              | 0              | 0              | Apr-13 | 0              | 0              | 0              | 1              |
| Mar-12 | 0              | 1              | 0              | 0              | May-13 | 0              | 0              | 0              | 1              |
| Apr-12 | 0              | 1              | 0              | 0              | Jun-13 | 0              | 0              | 0              | 1              |
| May-12 | 0              | 1              | 0              | 0              | Jul-13 | 0              | 0              | 0              | 1              |
| Jun-12 | 0              | 1              | 0              | 0              | Aug-13 | 0              | 0              | 0              | 1              |
| Jul-12 | 0              | 0              | 1              | 0              | Sep-13 | 0              | 0              | 0              | 1              |
| Aug-12 | 0              | 0              | 1              | 0              | Oct-13 | 0              | 0              | 0              | 1              |
| Sep-12 | 0              | 0              | 1              | 0              | Nov-13 | 0              | 0              | 0              | 1              |
| Oct-12 | 0              | 0              | 1              | 0              | Dec-13 | 0              | 0              | 0              | 1              |
|        |                |                |                |                | Jan-14 | 0              | 0              | 0              | 1              |

In the next section, results from different multiple regression models regarding Thailand's car sales figures are compared and discussed.

### 3.0 RESULTS AND DISCUSSION

In this section, various multiple regression models are examined. First, the regression model with the two independent variables only (Model 1), the regression model with added seasonality (Model 2), and regression model with added seasonality and special events (Model 3) are compared for their adjusted R<sup>2</sup>'s and Mean Absolute Percentage Errors (MAPEs) to see which regression model is best fitted and best accurate among these three. Then, the seasonal regression model with events or Model 3 is explored in detail.

#### 3.1 Comparison among the Three Multiple Regression Models

According to Table 3, the multiple regression Model 1 that does not take into account the seasonal and the special event effects has a low adjusted R<sup>2</sup> of just 42.84% and a high MAPE of 29.82%. The multiple regression Model 2 that incorporates the seasonal effect has a higher adjusted R<sup>2</sup> of 66.65% and a lower MAPE of 22.04%. The last regression Model 3 that includes both seasons and special events is shown to be the most promising model with the highest adjusted R<sup>2</sup> of 85.89% and the lowest MAPE of 15.80%.

**Table 3** Three multiple regression models' statistics

|         | Adjusted R <sup>2</sup> | MAPE   |
|---------|-------------------------|--------|
| Model 1 | 42.84%                  | 29.82% |
| Model 2 | 66.65%                  | 22.04% |
| Model 3 | 85.89%                  | 15.80% |

As a result, the proposed special event variables coupled with the appropriate tagging used in our seasonal regression model with events are proved effective and also make the model more explainable and more accurate. This Model 3 is hence worth a closer look in the next section.

#### 3.2 The Seasonal Model with Events (Model 3)

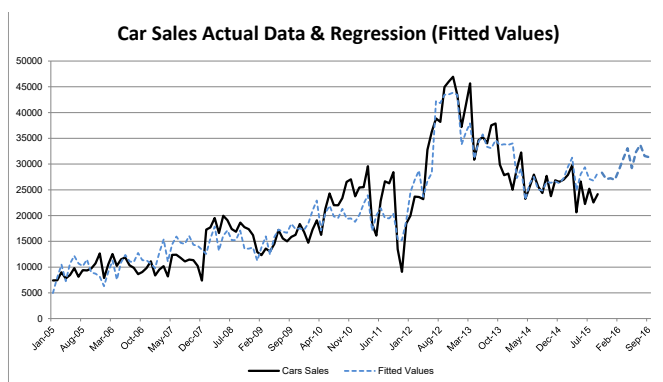
By any statistical package, the regression results obtained are equivalent. The coefficients for the two independent variables, eleven seasonal dummy variables, and four event variables of Model 3 are shown in Table 4 with corrected signs as expected. For example, the sign of the loan's interest rate is negative indicating intuitively that people tend to buy less cars when the loan rate is high. In contrast, the GDP's sign is positive as one would expect the car sales to go together with the general economic conditions. Furthermore, the 698.48 absolute coefficient value for the loan rate is much more than the 0.08 value for the GDP indicating also that the interest rate of the loan to pay for the car is a more influential factor on Thailand subcompact car sales.

As for the signs of the monthly seasonal dummy variables, it appears that the eleven variables are all positive. This means that the based month of January is the lowest season of the compact car sales in Thailand, possibly due to the money spending elsewhere during the Christmas and New Year celebrations. The high seasons, on the contrary, for the car sales, according also to Table 4, are from the months of May through September.

**Table 4** Coefficient table for Model 3

| Description        | Coefficient | Description          | Coefficient |
|--------------------|-------------|----------------------|-------------|
| <i>Y-Intercept</i> | -42,685.58  | <i>aug</i>           | 14,137.11   |
| <i>GDP</i>         | 0.08        | <i>sep</i>           | 15,612.88   |
| <i>Loan Rate</i>   | -698.48     | <i>oct</i>           | 9,911.42    |
| <i>feb</i>         | 4,966.20    | <i>nov</i>           | 8,956.51    |
| <i>mar</i>         | 9,035.47    | <i>dec</i>           | 7,121.69    |
| <i>apr</i>         | 8,674.60    | <i>e<sub>1</sub></i> | -1,098.98   |
| <i>may</i>         | 13,013.80   | <i>e<sub>2</sub></i> | 3,282.95    |
| <i>jun</i>         | 15,300.08   | <i>e<sub>3</sub></i> | 19,562.81   |
| <i>jul</i>         | 14,222.91   | <i>e<sub>4</sub></i> | 8,988.80    |

Let us now consider the last set of variables, namely, the special event variables, the sign for the flood period alone, i.e., event  $e_1$ , is negative confirming that the big flood really affected the car sales inversely. The signs for all other event variables, even the aftermath period,  $e_4$ , are positive. For events  $e_2$ ,  $e_3$  where the tax incentive campaign was active, it is quite normal to see positive coefficients. Especially in the period  $e_3$  when the campaign was in full effect without the flood, the  $e_3$  coefficient reached the peak. As for the reason why the sign of the coefficient in the aftermath event  $e_4$  (where the tax-incentive campaign already ended) is still positive, it is because some people did not actually buy the cars during the campaign period but what they did was just to reserve the cars for later buy. Consequently, the purchase could really occur sometime later, or more specifically, in the event  $e_4$  period. Afterwards, the sales seasons resume their normal patterns as before the two gigantic events occurred and thus all the event variables  $e_i$  turn back to 0s. Figure 2 plots the actual and fitted data of the Thailand's monthly car sales figures from January 2005 to September 2015.



**Figure 2** The actual data and its regression of Thailand's monthly compact car sales

## 4.0 CONCLUSION

In this research, by looking at the correlation coefficients between the subcompact car sales in Thailand as the dependent variable and the economic causal factors examined, it is evidenced that GDP and the loan's interest rate are good explanatory variables. However, these two predictors even better explain the variation of the car sales when seasonal and special event effects are added into the multiple regression model. In our case here, great fluctuations in car sales for some certain periods are due to these two particular events, namely, Thailand's

2011 big flood and the 2011-2012 tax-incentive first-car buyer program. An approach to handling such special events can be done through constructing the proposed special event variables together with proper event tagging. By the high adjusted  $R^2$  and high MAPE obtained, this multiple regression model with seasonality and event variables is proved to be suitably fitted and accurate.

To make the multiple regression model even more fitted and more accurate, other explanatory variables that influence the car industry either positively or negatively should be explored in more detail.

## References

- [1] Bank of Thailand. 2013. *Thailand's Economic Conditions in 2012*. Bank of Thailand, Bangkok.
- [2] Office of Industrial Economics. 2015. "Industrial Indices" [Online]. Available: <http://www.oie.go.th/>. [Accessed 15 December 2015].
- [3] Bank of Thailand. 2015. Thailand's Macro Economic Indicators [Online]. Available: <http://www2.bot.or.th/>. [Accessed 15 December 2015].
- [4] Dagum, E. B. and S. Bianconcini. 2016. *Seasonal Adjustment Methods and Real Time Trend-Cycle Estimation*. London: Springer-Verlag.
- [5] Pearson, K. 1908. On the Generalized Probable Error in Multiple Normal Correlation. *Biometrika*. 6: 59-68.
- [6] Billings, R. and D. Agthe. 1998. State-space Versus Multiple Regression for Forecasting Urban Water Demand. *Journal of Water Resources Planning and Management*. 124(2): 113-117.
- [7] Kim, T., J. Hong, and H. Koo. 2013. Forecasting Box-Office Revenue by Considering Social Network Services in the Korean Market. *Jurnal Teknologi*. 64(2): 97-101.
- [8] Ismail, Z., A. Yahya, and A. Shabri. 2009. Forecasting Gold Prices Using Multiple Linear Regression Method. *American Journal of Applied Sciences*. 6(8): 1509-1514.
- [9] Shahabuddin, S. 2009. Forecasting Automobile Sales. *Management Research News*. 32(7): 670-682.
- [10] Wang, F., K. Chang, and C. Tzeng. 2011. Using Adaptive Network-Based Fuzzy Inference System To Forecast Automobile Sales. *Expert Systems with Applications*. 38: 10587-10593.
- [11] Leenawong, C. and P. Netisopakul. 2016. Forecasting Car Sales in Thailand with the Impact of the National Flood and the First-Car Buyer Program. *Proceedings of the 4th International Conference on Business and Social Sciences*. Kyoto, Japan. 29-31 March 2016. 1006-1008.
- [12] Haugh, D., A. Mourougane, and O. Chatal. 2010. The Automobile Industry In and Beyond the Crisis. *OECD Economics Department, Working Papers*745.
- [13] G. Riley. 2011. "tutor2u" [Online]. Available: <http://www.tutor2u.net/economics/blog/economics-qa-what-economic-factors-affect-the-demand-for-new-cars>. [Accessed 15 December 2015].
- [14] Lovell, M. C. 1963. Seasonal Adjustment of Economic Time Series and Multiple Regression. *Journal of the American Statistical Association*. 58(304): 993-1010.
- [15] Wilson, J. H. and B. Keating. 2009. *Business Forecasting with ForecastX*. McGraw-Hill Education.

---

---

Reprint submitted to Thai Journal of Mathematics

---

---

<http://thaijmath.in.cmu.ac.th>

Online ISSN 1686-0209

## Event Index Computation for Forecasting Case Study: Car Sales in Thailand

Witchaya Rattanametawee and Chartchai Leenawong <sup>1</sup>

<sup>†</sup>Department of Mathematics, Faculty of Science,

<sup>‡</sup>King Mongkut's Institute of Technology Ladkrabang,  
Bangkok, 10520 THAILAND

e-mail : [witchaya232@gmail.com](mailto:witchaya232@gmail.com) (W.Rattanametawee)

[cleenawong@gmail.com](mailto:cleenawong@gmail.com) (C.Leenawong)

**Abstract :** Due to the impact of special events, both positive and negative, on the sales data, the ordinary Time-series Decomposition (TSD) forecasting model cannot merely capture these effects, even with the added seasonality and trends. Therefore, in this research, a new method for computing the event indices, representing the unusual fluctuations for a certain period in the time series, is proposed in order for it to be incorporated into TSD, alongside the conventional trend, seasonal, and cyclical components. A case study of subcompact car sales monthly data in Thailand during the years 2011-2018 is examined as for that time period contains the 2011 nationwide big flood reflecting the negative impact, as well as the nation's tax-incentive first-car buyer scheme reflecting the positive impact on the dataset. The mean absolute percentage error (MAPE) is used as an accuracy measure of the proposed forecasting model and it illustrates the promising results in the end.

**Keywords :** event index; time series forecasting; decomposition method; car sales.

**2000 Mathematics Subject Classification :** 47H09; 47H10 (2000 MSC )

---

---

<sup>1</sup>Corresponding author email: [cleenawong@gmail.com](mailto:cleenawong@gmail.com) (C.Leenawong)

## 1 Introduction

Thailand's automotive industry has been an important foundation for the nation's economic boost for many years. The government policies are always the key mechanisms for the industry to increase and improve the vehicle production capacities to replace the import volumes [1]. Since the government's liberalization policy on drawing in oversea investment for the industry development in 1991, Thailand has continuously expanded her production sizes for engine part making, automotive assembly, and other associated industries. As a result, automotive-related figures continued to grow in every segment including manufacturing, marketing, employment, and technology development, among others. Thanks also to the state's successful push on stimulating the domestic demand, Thailand became one of the global automotive manufacturing bases as well as the world's leading manufacturer of pickups and motorcycles [2]. Moreover, in the years 2014 and 2015, Thailand has achieved the highest production capacities in ASEAN [3] as shown in Table 1. From 2014 to 2015, ASEAN's total car production volume went down from 3.98 million to 3.89 million units. However, not only did Thailand manage to remain ASEAN's top car producer in both years, but also the production figure went up from 1.88 million units in 2014 to 1.91 million in 2015. Among these 1.91 million, 1.15 million cars were for commercial use while 0.76 million cars were for private use. Lagged behind Thailand in the ASEAN's car producer list, Indonesia and Malaysia were the numbers 2 and 3, each of which had roughly half the predecessor's production size.

Table 1: ASEAN's car production capacities from 2014 to 2015. (units: cars)

| Country      | Private Car      | Commercial vehicles | 2015             | 2014             | Percent Change (%) |
|--------------|------------------|---------------------|------------------|------------------|--------------------|
| Thailand     | 760,688          | 1,152,314           | 1,913,002        | 1,880,007        | 2                  |
| Indonesia    | 824,445          | 274,335             | 1,098,780        | 1,298,523        | -15                |
| Malaysia     | 563,883          | 50,781              | 614,664          | 596,418          | 3                  |
| Vietnam      | 99,052           | 72,701              | 171,753          | 121,084          | 42                 |
| Philippines  | 36,395           | 62,373              | 98,768           | 88,854           | 11                 |
| <b>Total</b> | <b>2,284,463</b> | <b>1,612,504</b>    | <b>3,896,967</b> | <b>3,984,877</b> | <b>-2</b>          |

Ref. : ASEAN Automotive Federation [3]

The monthly car sales figures in Thailand [4] for 8 years, ranged from January 2011 to December 2018 are shown in Figure 1. Based on this historical sales record, when looked closely, it revealed that the sales figures were very fluctuate, especially during the years 2011-2012. In that time, two major incidents occurred. First, the national big flood in the second half of the year 2011 severely damaged the country, the people, and obviously, many businesses. Even though, the flood lasted for about 5-6 months but the aftermath carried on for at least another year. The second coincident was the government policy on the intensive tax refund for

the first car buyers who bought or reserved to buy their first subcompact car during the period of September 2011-December 2012 [5].

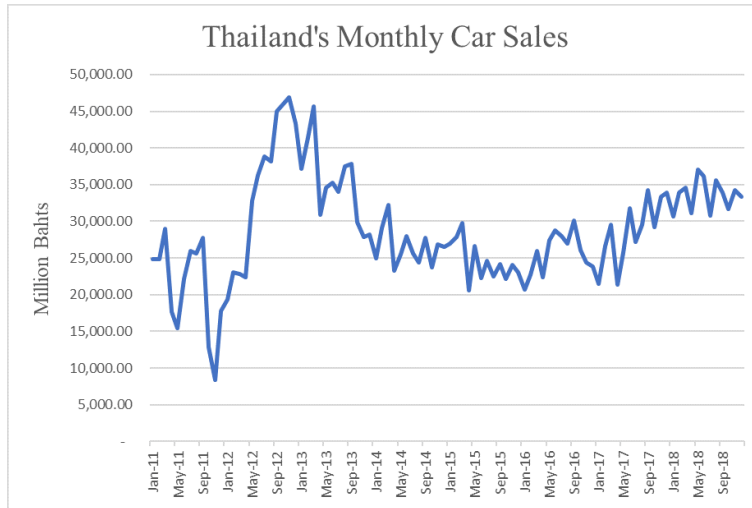


Figure 1: Thailand's monthly car sales (in Million Baht) from January 2011 to December 2018.

Previous work regarding this research in various aspects has been reviewed and concluded here as follows.

Ludvigson (1998), showed that the government financial policies concerning bank loans for car purchase affected car sales figures. Inversely, increasing interest rates had a negative impact on car sales figures [6]. The proportion of car ownership from 26 countries in the Organization for Economic Co-operation and Development (OECD) group, between the year 1960 and 1992 was relatively high in the industrialized member countries [7] including Portugal, Greece, and Ireland, based on their own expectation that they would be able seek higher income growth in the future. Vice versa, in China, India, and Pakistan, the growth of car ownership followed the growth of per-capita income of the countries.

As a result of the household expenses survey from 1970 to 1995, Dargay (2001) found that car ownership tended to increase with increasing revenue and negative correlation when revenue decreased [8]. Moreover, the research by Shahabudin (2009) showed a regression model incorporating all important predictive variables such as income levels, interest rates, financial aggregate and unemployment rates, that affected domestic and foreign car sales figures [9]. Wirotcheewan P. et al. (2011), used linear programming to find an accurate forecasting model for predicting advanced demand in exported automotive to foreign market and calculate the optimal quantity for export [10]. Muhammad et al. (2012) revealed that the GDP variables were positively correlated with car sales figures in five ASEAN countries, namely, Malaysia, Indonesia, Singapore, Philippine, and Thailand [11].

This indicated that national income was really one of the important factors to the automotive industry, presumably as well as many other industries. On the other hand, inflation unemployment rates and interest rates had a negative impact on car sales figures.

Moreover, some researches revealed that significant increase in fuel price could also negatively impact the car sales [12],[13],[14]. This factor was found to induce uncertainty in operation costs that could affect the total car production costs. This, in turn, impacted the investment value and the consumer demand. The same finding was shown by Lee and Ni (2002) [15] in their research on 14 industrial countries. The hike in fuel prices was found to disrupt supply by the industrial sector, while at the same time, influenced consumer demand for industrial products, especially motor vehicles.

Rattanametawee W., Leenawong C. and Netisopakul P. (2016) [16] proposed multiple linear regression that incorporated seasonality and special events into the regression model. Their dependent variable was the car sales figures in Thailand while the independent variables were the nation's GDP and the car loan rates with dummy variables for handling the seasonality and the irregular events. Among all their three regression models, it turned out, as expected, that the model with the seasonal variables and the dummy event variables proved to be most useful according to the relevant high adjusted R-square.

Other research focusing on the time series forecasting methodology is presented in the following. The research by Apiwattanachai T. and Pichitlamken J. (2008) [17] used the three exponential smoothing technique or the Holt-Winter's forecasting model to estimate the customer demand in automobile. Both the multiplicative and the additive models with all three constants,  $\alpha$ ,  $\beta$ , and  $\gamma$  were used due to the presence of the trend and seasonality components. The result showed in their work that the additive model had a lower MAPE and hence a better model.

Chaimongkol W. and Tansathit C. (2010) [18] proposed a modification on the Decomposition and the Holt-Winter's forecasting models. The new model combining the additive decomposition with the additive Holt-Winters was applied to the passenger car demand data. The results was then compared to those from the original Decomposition and the Holt- Winter's models by their mean absolute percentage error (MAPE). The combined model exhibited better accuracy.

This paper proposes a new computation method for event indices to be use in the time series forecasting with a case study of the subcompact car sales data in Thailand due to the two special incidents, the national big flood and the tax-incentive car buyer program. In the next section, the proposed methodology for computing the event indices is explained, followed by the experiments and their results in Section 3. Then, this research is summarized and concluded in Section 4.

## 2 The Methodology

The Time-Series Decomposition (TSD) Method is a widely used model for forecasting when the forecast is based only on one factor, namely, time [19]. Typically, the four components of a time series are composed of the Trend ( $T$ ), Seasonal ( $S$ ), Cyclical ( $C$ ), and Irregular ( $I$ ) components. In this research, a multiplicative decomposition model is assumed and expressed as

$$y = T * S * C * I.$$

In addition to the four typical components above, one additional aperiodic component is added into the above formula, which is subsequently called the special event component ( $E$ ), and thus the multiplicative model becomes

$$y = T * S * C * I * E.$$

The special event component accounts for some untypical occurrence that really affects, for some period, the time series, either positively or negatively. For example, a marketing campaign could contribute to a sharp peak of the sales figure, thus the positive effect whereas a tsunami could cause a sharp drop in the sales for some particular product, thus the negative effect. In the proposed methodology, both marketing campaign and the tsunami will be called special events.

To describe the proposed computation for event indices more understandably, a case study of Thailand's subcompact car sales figure over an 8-year period starting from year 2011 up until 2018 was examined. Accordingly, two special events were found to have occurred during this 8-year monthly time-series sales data. One was the national big flood in Thailand in the late 2011, supposedly causing negative effects on many industries including this passenger car industry. While the other event was the government's tax-incentive first-car buyer program being officially in effect for those who purchased or made a reservation to purchase their first car from September 2011 to December 2012, supposedly producing positive effects on this particular car industry.

However, as the effective periods of the two special events overlapped, they classify the sales data into four distinctive events to reflect the effects of both special events. Furthermore, these four distinctive events to be flagged in the monthly sales periods do not have to be in compliance with the official time frame mentioned above. It is focused more on the real actions of the buyers whether or not to buy their first cars. The four classified events along with their event flag values as well as related parameters and notations are defined as follows.

Let  $X_{y,m}^k$  refers to the original monthly car sales data of year  $y$ , months  $m$ , with the assigned event flag  $k$ , where  $y = 1, 2, \dots$  refers to the year 2011, 2012, and so on,  $m = 1, 2, \dots, 12$  refers to the month January, February, ..., December. Finally, the event flag  $k$  takes one of the values 0, 1, 2, 3 where flag  $k = 0$  (the base flag) refers to the regular sales period when there is no effect from neither

the big flood nor the incentive program, flag  $k = 1$  refers to the period where buyers' decisions are affected mainly by the flood, flag  $k = 2$  refers the period where buyers' decisions are affected by both the flood and the incentive program, and flag  $k = 3$  refers to the sale period after the incentive program ends.

The time-series decomposition algorithm incorporating special events is proposed and explained here. The steps for computing the event indices and removing them from the original data resulting in the de-evented data are described in Steps 1 through 5. Then, from Step 6 onward, the conventional decomposition of the seasonal and cyclical components is incorporated into the de-evented data so that the subsequent data obtained in Step 11 is left with the trend component only, and hence called the de-all data. Finally, the forecasted data can be predicted in last two steps. First of all, let us start with finding the centered moving averages of the original monthly data.

### The Event Decomposition Algorithm

**Step 1:** Finding the Centered Moving Average (CMAs).

Let  $\bar{X}_{y,m}^k$  refer to the CMA of  $X_{y,m}^k$  when the number of periods to average is

12. Obviously,  $\bar{X}_{y,m}^k$  starts at  $\bar{X}_{1,7}^k$  and ends at  $\bar{X}_{8,6}^k$ . For example,

$$\bar{X}_{1,7}^k = \text{CMA}(X_{1,7}^k) = (X_{1,1}^k + X_{1,2}^k + \dots + X_{1,12}^k)/12,$$

$$\bar{X}_{1,8}^k = \text{CMA}(X_{1,8}^k) = (X_{1,2}^k + X_{1,3}^k + \dots + X_{1,12}^k + X_{2,1}^k)/12,$$

$$\bar{X}_{1,9}^k = \text{CMA}(X_{1,9}^k) = (X_{1,3}^k + X_{1,4}^k + \dots + X_{1,12}^k + X_{2,1}^k + X_{2,2}^k)/12, \dots,$$

$$\bar{X}_{8,6}^k = \text{CMA}(X_{8,6}^k) = (X_{8,1}^k + X_{8,2}^k + \dots + X_{8,11}^k + X_{8,12}^k)/12.$$

Then, each  $X_{y,m}^k$  as well as its corresponding  $\bar{X}_{y,m}^k$  is assigned a flag  $k$  where  $k = 0, 1, 2$ , or 3 represents each of the four special events previously mentioned above.

**Step 2:** Computing the Mean Base-flagged CMAs.

For each month  $m$ , take only the  $\bar{X}_{y,m}^k$  having flag  $k = 0$  or the base flag, and compute the mean of the average of these data. In terms of notation, it can be expressed as

$$\bar{\bar{X}}_m^0 = \text{avg} \bar{X}_{y,m}^0 \quad (2.1)$$

**Step 3:** Finding the Event Factors.

Let  $EF_{y,m}^k$  represent the event factor of year  $y$ , month  $m$ , which also contains the same flag  $k$ , as in  $\bar{X}_{y,m}^k$ . The event factor is obtained by dividing its corresponding CMA by the Mean Base-flagged CMA,  $\bar{\bar{X}}_m^0$ . Therefore,

$$EF_{y,m}^k = \frac{\bar{X}_{y,m}^k}{\bar{\bar{X}}_m^0} \quad (2.2)$$

**Step 4:** Computing the Event Indices.

Let  $E^k$  represent the event index for flag  $k$ , where  $k = 0, 1, 2, 3$ . This index can be obtained by, for each  $k$ , taking the average of all the event factors  $EF_{y,m}^k$

having the same flag  $k$  across all the data. That is,

$$E^k = \text{avg}EF_{y,m}^k, \quad (2.3)$$

**Step 5:** Removing the Event Effect from the Data (De-eventing).

Let  $D_{y,m}^k$  along with the common three parameters year  $y$ , month  $m$ , and flag  $k$ , represent the data after the event effect is removed or de-evented. Since this model is multiplicative,  $D_{y,m}^k$  is obtained by taking the initial data,  $X_{y,m}^k$ , and dividing it by the event index,  $E^k$ , of the same flag  $k$ , as illustrated below

$$D_{y,m}^k = \frac{X_{y,m}^k}{E^k} \quad (2.4)$$

Because the event indices have been taken out, for less complicated notations, the superscript  $k$  can be discarded. That is, from here on,  $D_{y,m} = D_{y,m}^k$  and the data is now left only with the seasonal and cyclical components. The remaining steps are to remove the effects of these two components from the current de-evented data. More precisely, seasonal indices are obtained through Steps 6 to 8 while cyclical factors are obtained through Steps 9 to 10.

**Step 6:** Finding the CMAs of the De-evented Data.

Let  $\bar{D}_{y,m}$  refer to the CMA of  $D_{y,m}$  when the number of period to average is 12. Therefore  $\bar{D}_{y,m}$  starts at  $\bar{D}_{1,7}$  and ends at  $\bar{D}_{8,6}$ , for instance

$$\begin{aligned} \bar{D}_{1,7} &= \text{CMA}(D_{1,7}) = (D_{1,1} + D_{1,2} + \dots + D_{1,11} + D_{1,12})/12, \\ \bar{D}_{1,8} &= \text{CMA}(D_{1,8}) = (D_{1,2} + D_{1,3} + \dots + D_{1,11} + D_{1,12} + D_{2,1})/12, \dots, \\ \bar{D}_{8,6} &= \text{CMA}(D_{8,6}) = (D_{8,1} + D_{8,2} + \dots + D_{8,11} + D_{8,12})/12. \end{aligned}$$

**Step 7:** Computing the Seasonal Factors.

Let  $SF_{y,m}$  represent the seasonal factor of year  $y$  and month  $m$ . Therefore, the seasonal factors are obtained from dividing the de-evented data by its CMA counterparts as follows

$$SF_{y,m} = \frac{D_{y,m}}{\bar{D}_{y,m}}. \quad (2.5)$$

**Step 8:** Computing the Seasonal Indices.

Let  $S_m$  represent the seasonal index for month  $m$ . Then, this index is found by taking the average of the same month- $m$  seasonal factors data across the years. That is,

$$S_m = \text{avg}SF_{y,m}. \quad (2.6)$$

Before removing the seasonal effects from the de-evented data, steps for obtaining the cyclical factors are carried out in the following Steps 9 to 10. Subsequently, both seasonal and cyclical factors will be removed simultaneously in Step 11.

**Step 9:** Obtaining the Trend Line Equation of the De-evented Data.

Let  $TD_{y,m}$  refer to the year  $y$ , month  $m$  data obtained from the linear trend equation of the CMAs of the de-evented data,  $\bar{D}_{y,m}$ , from Step 6. The general formula for this trend line depending on time  $t$  is as follows

$$TD_{y,m} = b_0 + b_1 t. \quad (2.7)$$

where  $b_0$  is the  $y$ -intercept and  $b_1$  is the slope of the trend line. After getting all the data from the trend line equation, the next step is to find the cycle factors.

**Step 10:** Finding the Cycle Factors.

A cycle factor is just a division of the CMA of the de-evented data and its trend-line counterpart and when  $CF_{y,m}$  represents the cycle factor of year  $y$  and month  $m$ ,

$$CF_{y,m} = \frac{\bar{D}_{y,m}}{TD_{y,m}}. \quad (2.8)$$

**Step 11:** Removing the Seasonal and Cyclical Components from the De-evented Data.

The de-all data,  $DA_{y,m}$ , refers to the data whose all three main components, namely, event, seasonal, and cyclical components, have been removed; therefore, divide the de-evented data from Step 5 by its corresponding seasonal index from Step 8 and also the cycle factor from Step 10 as follows

$$DA_{y,m} = \frac{D_{y,m}}{S_m * CF_{y,m}}. \quad (2.9)$$

**Step 12:** Obtaining the Trend Line Equation of the De-all Data.

Now, the only component left in the de-all data is the trend component. Let  $TDA_{y,m}$  refer to the year  $y$ , month  $m$  data obtained from the linear trend equation of the de-all data in the previous step. The general formula for this trend line depending on time  $t$  is as follows

$$TDA_{y,m} = c_0 + c_1 t. \quad (2.10)$$

where  $c_0$  is the  $y$ -intercept and  $c_1$  is the slope of the trend line.

**Step 13:** Incorporating the Three Main Components Back to the Forecasted Data.

Using the trend line equation of the de-all data in (2.10), forecasted data in year  $y$ , month  $m$ , or  $TDA_{y,m}$ , can all be obtained. Then, by the multiplicative approach, the final forecasted data, denoted by  $TF_{y,m}$ , becomes

$$TF_{y,m} = TDA_{y,m} * S_m * CF_{y,m} * E^k. \quad (2.11)$$

This proposed algorithm will be applied on the Thailand's car sales data in the next section. In the meantime, during the development of this algorithm, it is found that the event index of the base flag case or the normal case is always 1 as shown in the following Theorem.

**Theorem 2.1.** Let  $E^k$  represent the event index for flag  $k, k = 0, 1, 2, 3$ , and be defined as in (2.3), i.e.,  $E^k = \text{avg}EF_{y,m}^k$ , then  $E^0 = 1$ .

*Proof.* Let  $n_m^0$  be the total number of data of month  $m$  across all years having flag 0;  $m = 1, 2, \dots, 12$ .

From Step 4, we know that the event index for the base flag is

$$E^0 = \text{avg}EF_{y,m}^0,$$

which can be expressed as

$$E^0 = \text{avg}EF_{y,m}^0 = \frac{\sum EF_{y,m}^0}{n_1^0 + n_2^0 + \dots + n_{12}^0}.$$

By the formula (2.2) in Step 3, we obtain

$$E^0 = \frac{\sum \left( \frac{\bar{X}_{y,m}^0}{\bar{X}_m^0} \right)}{n_1^0 + n_2^0 + \dots + n_{12}^0},$$

which can be decomposed into

$$E^0 = \frac{\sum \left( \frac{\bar{X}_{y,1}^0}{\bar{X}_1^0} \right) + \sum \left( \frac{\bar{X}_{y,2}^0}{\bar{X}_2^0} \right) + \dots + \sum \left( \frac{\bar{X}_{y,12}^0}{\bar{X}_{12}^0} \right)}{n_1^0 + n_2^0 + \dots + n_{12}^0}.$$

By formula (2.1) in Step 2, we have

$$\begin{aligned} E^0 &= \frac{\sum \left( \frac{\bar{X}_{y,1}^0}{\text{avg}\bar{X}_{y,1}^0} \right) + \sum \left( \frac{\bar{X}_{y,2}^0}{\text{avg}\bar{X}_{y,2}^0} \right) + \dots + \sum \left( \frac{\bar{X}_{y,12}^0}{\text{avg}\bar{X}_{y,12}^0} \right)}{n_1^0 + n_2^0 + \dots + n_{12}^0} \\ &= \frac{\left( \frac{\sum \bar{X}_{y,1}^0}{\left( \frac{\sum \bar{X}_{y,1}^0}{n_1^0} \right)} \right) + \left( \frac{\sum \bar{X}_{y,2}^0}{\left( \frac{\sum \bar{X}_{y,2}^0}{n_2^0} \right)} \right) + \dots + \left( \frac{\sum \bar{X}_{y,12}^0}{\left( \frac{\sum \bar{X}_{y,12}^0}{n_{12}^0} \right)} \right)}{n_1^0 + n_2^0 + \dots + n_{12}^0}. \end{aligned}$$

Then,

$$E^0 = \frac{\left( \frac{\sum \bar{X}_{y,1}^0}{\sum \bar{X}_{y,1}^0} \right) n_1^0 + \left( \frac{\sum \bar{X}_{y,2}^0}{\sum \bar{X}_{y,2}^0} \right) n_2^0 + \dots + \left( \frac{\sum \bar{X}_{y,12}^0}{\sum \bar{X}_{y,12}^0} \right) n_{12}^0}{n_1^0 + n_2^0 + \dots + n_{12}^0}.$$

Hence,  $E^0 = 1$ .

□

### 3 The Experimentals and the Results

In this section, the proposed time-series decomposition algorithm incorporating special events from the previous section is applied to the Thailand's monthly car sales figures ranged from January 2011 to December 2018. Due to the two major incidents being the national big flood and the tax incentive program and the definitions of the four flags previously stated in Section 2, the following flag assignment to the monthly car sales data is obtained:

Flag 1 referring to the sales period where car buyers' decisions are affected mainly by the flood, is assigned to months Oct 2011 - Jan 2012,

Flag 2 referring the period where car buyers' decisions are affected by both the flood and the incentive program, is assigned to months Feb 2012 - Sep 2013,

Flag 3 referring to the aftermath period, is assigned to Oct 2013 - Sep 2014,

Flag 0 referring to the normal sales period, is assigned to the remaining months.

Table 2: Results from the event decomposition algorithm Steps 1 to 5.

| Month  | Time | Car sales | flag | $\bar{X}_{y,m}^k$ | $\bar{X}_m^0$ | $EF_{y,m}^k$ | $E^k$ | $D_{y,m}$ |
|--------|------|-----------|------|-------------------|---------------|--------------|-------|-----------|
| Jan-11 | 1    | 24,812.31 | 0    |                   | 27,239.87     |              | 1.00  | 24,812.31 |
| Feb-11 | 2    | 24,856.65 | 0    |                   | 27,410.99     |              | 1.00  | 24,856.65 |
| Mar-11 | 3    | 28,944.01 | 0    |                   | 27,592.34     |              | 1.00  | 28,944.01 |
| Apr-11 | 4    | 17,675.21 | 0    |                   | 27,739.47     |              | 1.00  | 17,675.21 |
| May-11 | 5    | 15,471.06 | 0    |                   | 27,899.45     |              | 1.00  | 15,471.06 |
| Jun-11 | 6    | 22,160.50 | 0    |                   | 28,048.28     |              | 1.00  | 22,160.50 |
| Jul-11 | 7    | 26,003.30 | 0    | 20,795.62         | 24,954.96     | 0.83         | 1.00  | 26,003.30 |
| Aug-11 | 8    | 25,601.39 | 0    | 20,490.56         | 24,979.90     | 0.82         | 1.00  | 25,601.39 |
| Sep-11 | 9    | 27,763.89 | 0    | 20,159.89         | 25,010.45     | 0.81         | 1.00  | 27,763.89 |
| Oct-11 | 10   | 12,813.77 | 1    | 20,102.19         | 26,840.77     | 0.75         | 0.80  | 16,030.14 |
| Nov-11 | 11   | 8,403.67  | 1    | 21,020.13         | 27,131.00     | 0.77         | 0.80  | 10,513.07 |
| Dec-11 | 12   | 17,787.90 | 1    | 22,328.01         | 27,469.30     | 0.81         | 0.80  | 22,252.84 |
| Jan-12 | 13   | 19,319.96 | 1    | 23,450.14         | 27,239.87     | 0.86         | 0.80  | 24,169.46 |
| Feb-12 | 14   | 23,027.61 | 2    | 24,510.31         | 27,410.99     | 0.89         | 1.32  | 17,464.63 |
| Mar-12 | 15   | 22,836.78 | 2    | 25,751.72         | 27,592.34     | 0.93         | 1.32  | 17,319.89 |
| Apr-12 | 16   | 22,397.63 | 2    | 27,851.60         | 27,739.47     | 1.00         | 1.32  | 16,986.83 |
| .      | .    | .         | .    | .                 | .             | .            | .     | .         |
| .      | .    | .         | .    | .                 | .             | .            | .     | .         |
| Jul-13 | 31   | 34,043.82 | 2    | 34,521.43         | 24,954.96     | 1.38         | 1.32  | 25,819.55 |
| Aug-13 | 32   | 37,534.01 | 2    | 33,504.01         | 24,979.90     | 1.34         | 1.32  | 28,466.59 |
| Sep-13 | 33   | 37,879.10 | 2    | 32,434.55         | 25,010.45     | 1.30         | 1.32  | 28,728.30 |
| Oct-13 | 34   | 29,906.23 | 3    | 31,558.23         | 26,840.77     | 1.18         | 1.05  | 28,464.52 |
| Nov-13 | 35   | 27,857.49 | 3    | 30,862.94         | 27,131.00     | 1.14         | 1.05  | 26,514.55 |
| Dec-13 | 30   | 28,153.73 | 3    | 30,181.07         | 27,469.30     | 1.10         | 1.05  | 26,796.52 |
| .      | .    | .         | .    | .                 | .             | .            | .     | .         |
| .      | .    | .         | .    | .                 | .             | .            | .     | .         |
| Jul-14 | 43   | 25,633.96 | 3    | 26,585.74         | 24,954.96     | 1.07         | 1.05  | 24,398.21 |
| Aug-14 | 44   | 24,377.96 | 3    | 26,616.19         | 24,979.90     | 1.07         | 1.05  | 23,202.76 |
| Sep-14 | 45   | 27,720.11 | 3    | 26,461.53         | 25,010.45     | 1.06         | 1.05  | 26,383.80 |
| Oct-14 | 46   | 23,767.18 | 0    | 26,250.17         | 26,840.77     | 0.98         | 1.00  | 23,767.18 |
| Nov-14 | 47   | 26,852.59 | 0    | 26,186.28         | 27,131.00     | 0.97         | 1.00  | 26,852.59 |
| .      | .    | .         | .    | .                 | .             | .            | .     | .         |
| .      | .    | .         | .    | .                 | .             | .            | .     | .         |
| Sep-18 | 93   | 33,901.57 | 0    |                   | 25,010.45     |              | 1.00  | 33,901.57 |
| Oct-18 | 94   | 31,710.03 | 0    |                   | 26,840.77     |              | 1.00  | 31,710.03 |
| Nov-18 | 95   | 34,267.79 | 0    |                   | 27,131.00     |              | 1.00  | 34,267.79 |
| Dec-18 | 96   | 33,409.57 | 0    |                   | 27,469.30     |              | 1.00  | 33,409.57 |

By following the steps of the proposed algorithm, Table 2 shows the results of undertaking Steps 1 through 5 for computing the event indices and de-event

them from the sales data. The event indices for the four flags are shown in the second right-most column,  $E^k$ , more particularly,  $E^0 = 1.00$ ,  $E^1 = 0.80$ ,  $E^2 = 1.32$  and  $E^3 = 1.05$ . These event indices really capture the essence of each of the four flags. The index for flag 0 or the base flag being exactly equal to 1.00 can truly represent the the normal sales period while the index for flag 1 is lower than that of the normal period because, during the flood period, people are naturally not in the mood for buying cars. Later, when the flood is going away and, to some people, with their cars, the index for flag 2 therefore gets higher than that of the normal period showing that the tax-incentive program is working effectively in people's minds. The event index for the last period flag 3 is roughly around 1.00 showing everything is coming back to normal. Note also that the base flag event index  $E^0$  at 1.00 satisfies Theorem 2.1. After obtaining all the event indices, the de-vented data are shown in the right-most column,  $D_{y,m}^k$  of Table 2.

Next, to obtain the seasonal indices, Steps 6 through 8 are carried out on the de-vented data  $D_{y,m}^k$  from Table 2. The right-most column of Table 3 shows the resulted seasonal indices for each month, as also reported here as follows,  $S_1 = 0.95$ ,  $S_2 = 0.99$ ,  $S_3 = 1.07$ ,  $S_4 = 0.83$ ,  $S_5 = 1.00$ ,  $S_6 = 1.04$ ,  $S_7 = 1.03$ ,  $S_8 = 1.02$ ,  $S_9 = 1.13$ ,  $S_{10} = 0.97$ ,  $S_{11} = 0.96$  and  $S_{12} = 1.01$ .

Table 3: Results from the event decomposition algorithm Steps 6 to 8.

| Month  | Time | Car sales | $D_{y,m}^k$ | $\bar{D}_{y,m}$ | $SF_{y,m}^k$ | $S_m$ |
|--------|------|-----------|-------------|-----------------|--------------|-------|
| Jan-11 | 1    | 24,812.31 | 24,812.31   |                 |              | 0.95  |
| Feb-11 | 2    | 24,856.65 | 24,856.65   |                 |              | 0.99  |
| Mar-11 | 3    | 28,944.01 | 28,944.01   |                 |              | 1.07  |
| Apr-11 | 4    | 17,675.21 | 17,675.21   |                 |              | 0.83  |
| May-11 | 5    | 15,471.06 | 15,471.06   |                 |              | 1.00  |
| Jun-11 | 6    | 22,160.50 | 22,160.50   |                 |              | 1.04  |
| Jul-11 | 7    | 26,003.30 | 26,003.30   | 21,813.58       | 1.19         | 1.03  |
| Aug-11 | 8    | 25,601.39 | 25,601.39   | 21,478.79       | 1.19         | 1.02  |
| Sep-11 | 9    | 27,763.89 | 27,763.89   | 20,686.45       | 1.34         | 1.13  |
| Oct-11 | 10   | 12,813.77 | 16,030.14   | 20,173.43       | 0.79         | 0.97  |
| Nov-11 | 11   | 8,403.67  | 10,513.07   | 20,535.97       | 0.51         | 0.96  |
| Dec-11 | 12   | 17,787.90 | 22,252.84   | 21,149.11       | 1.05         | 1.01  |
| Jan-12 | 13   | 19,319.96 | 24,169.46   | 21,515.35       | 1.12         | 0.95  |
| Feb-12 | 14   | 23,027.61 | 17,464.63   | 21,799.96       | 0.80         | 0.99  |
| Mar-12 | 15   | 22,836.78 | 17,319.89   | 22,204.31       | 0.78         | 1.07  |
| Apr-12 | 16   | 22,397.63 | 16,986.83   | 23,254.44       | 0.73         | 0.83  |
| .      | .    | .         | .           | .               | .            | .     |
| .      | .    | .         | .           | .               | .            | .     |
| Jul-13 | 31   | 34,043.82 | 25,819.55   | 27,767.76       | 0.93         | 1.03  |
| Aug-13 | 32   | 37,534.01 | 28,466.59   | 27,432.41       | 1.04         | 1.02  |
| Sep-13 | 33   | 37,879.10 | 28,728.30   | 27,115.93       | 1.06         | 1.13  |
| Oct-13 | 34   | 29,906.23 | 28,464.52   | 26,898.34       | 1.06         | 0.97  |
| Nov-13 | 35   | 27,857.49 | 26,514.55   | 26,764.16       | 0.99         | 0.96  |
| Dec-13 | 30   | 28,153.73 | 26,796.52   | 26,678.26       | 1.00         | 1.01  |
| .      | .    | .         | .           | .               | .            | .     |
| .      | .    | .         | .           | .               | .            | .     |
| Jul-14 | 43   | 25,633.96 | 24,398.21   | 25,668.26       | 0.95         | 1.03  |
| Aug-14 | 44   | 24,377.96 | 23,202.76   | 25,807.48       | 0.90         | 1.02  |
| Sep-14 | 45   | 27,720.11 | 26,383.80   | 25,776.12       | 1.02         | 1.13  |
| Oct-14 | 46   | 23,767.18 | 23,767.18   | 25,676.21       | 0.93         | 0.97  |
| Nov-14 | 47   | 26,852.59 | 26,852.59   | 25,710.33       | 1.04         | 0.96  |
| .      | .    | .         | .           | .               | .            | .     |
| .      | .    | .         | .           | .               | .            | .     |
| Sep-18 | 93   | 33,901.57 | 33,901.57   |                 |              | 1.13  |
| Oct-18 | 94   | 31,710.03 | 31,710.03   |                 |              | 0.97  |
| Nov-18 | 95   | 34,267.79 | 34,267.79   |                 |              | 0.96  |
| Dec-18 | 96   | 33,409.57 | 33,409.57   |                 |              | 1.01  |

Table 4: Results from the event decomposition algorithm Steps 9 to 13.

| Month  | Time | Car sales | $TD_{y,m}$ | $CF_{y,m}$ | $DA_{y,m}$ | $TDA_{y,m}$ | $TF_{y,m}$ |
|--------|------|-----------|------------|------------|------------|-------------|------------|
| Jan-11 | 1    | 24,812.31 | 23,517.25  |            |            | 23,240.68   |            |
| Feb-11 | 2    | 24,856.65 | 23,582.98  |            |            | 23,312.41   |            |
| Mar-11 | 3    | 28,944.01 | 23,648.71  |            |            | 23,384.14   |            |
| Apr-11 | 4    | 17,675.21 | 23,714.44  |            |            | 23,455.86   |            |
| May-11 | 5    | 15,471.06 | 23,780.17  |            |            | 23,527.59   |            |
| Jun-11 | 6    | 22,160.50 | 23,845.91  |            |            | 23,599.32   |            |
| Jul-11 | 7    | 26,003.30 | 23,911.64  | 0.91       | 27,725.66  | 23,671.05   | 22,149.41  |
| Aug-11 | 8    | 25,601.39 | 23,977.37  | 0.90       | 27,978.55  | 23,742.78   | 21,675.44  |
| Sep-11 | 9    | 27,763.89 | 24,043.10  | 0.86       | 28,583.80  | 23,814.51   | 23,078.10  |
| Oct-11 | 10   | 12,813.77 | 24,108.83  | 0.84       | 19,662.31  | 23,886.23   | 15,566.47  |
| Nov-11 | 11   | 8,403.67  | 24,174.56  | 0.85       | 12,883.34  | 23,957.96   | 15,627.53  |
| Dec-11 | 12   | 17,787.90 | 24,240.30  | 0.87       | 25,163.53  | 24,029.69   | 16,986.40  |
| Jan-12 | 13   | 19,319.96 | 24,306.03  | 0.89       | 28,877.34  | 24,101.42   | 16,124.71  |
| Feb-12 | 14   | 23,027.61 | 24,371.76  | 0.89       | 19,664.13  | 24,173.15   | 28,307.88  |
| Mar-12 | 15   | 22,836.78 | 24,437.49  | 0.91       | 17,844.66  | 24,244.88   | 31,027.48  |
| Apr-12 | 16   | 22,397.63 | 24,503.22  | 0.95       | 21,570.40  | 24,316.60   | 25,249.14  |
| .      | .    | .         | .          | .          | .          | .           | .          |
| Jul-13 | 31   | 34,043.82 | 25,489.20  | 1.09       | 23,053.40  | 25,392.53   | 37,498.09  |
| Aug-13 | 32   | 37,534.01 | 25,554.93  | 1.07       | 25,960.68  | 25,464.26   | 36,816.28  |
| Sep-13 | 33   | 37,879.10 | 25,620.66  | 1.06       | 24,044.24  | 25,535.99   | 40,229.18  |
| Oct-13 | 34   | 29,906.23 | 25,686.39  | 1.05       | 27,898.58  | 25,607.71   | 27,450.50  |
| Nov-13 | 35   | 27,857.49 | 25,752.12  | 1.04       | 26,558.24  | 25,679.44   | 26,935.70  |
| Dec-13 | 30   | 28,153.73 | 25,817.85  | 1.03       | 25,584.76  | 25,751.17   | 28,336.86  |
| .      | .    | .         | .          | .          | .          | .           | .          |
| Jul-14 | 43   | 25,633.96 | 26,277.98  | 0.98       | 24,295.43  | 26,253.27   | 27,699.66  |
| Aug-14 | 44   | 24,377.96 | 26,343.71  | 0.98       | 23,186.81  | 26,325.00   | 27,677.36  |
| Sep-14 | 45   | 27,720.11 | 26,409.44  | 0.98       | 23,944.96  | 26,396.73   | 30,558.42  |
| Oct-14 | 46   | 23,767.18 | 26,475.17  | 0.97       | 25,152.79  | 26,468.45   | 24,952.74  |
| Nov-14 | 47   | 26,852.59 | 26,540.90  | 0.97       | 28,856.90  | 26,540.18   | 24,639.87  |
| .      | .    | .         | .          | .          | .          | .           | .          |
| Sep-18 | 93   | 33,901.57 | 29,564.56  | 1.14       | 26,233.91  | 29,839.68   | 38,472.40  |
| Oct-18 | 94   | 31,710.03 | 29,630.29  | 1.15       | 28,419.23  | 29,911.41   | 33,298.10  |
| Nov-18 | 95   | 34,267.79 | 29,696.02  | 1.15       | 31,089.95  | 29,983.14   | 32,971.70  |
| Dec-18 | 96   | 33,409.57 | 29,761.75  | 1.15       | 28,780.89  | 30,054.87   | 34,808.05  |

Then, Steps 9 and 10 for obtaining the cycle factors are performed. All of the event factors are revealed in column  $CF_{y,m}$  of Table 4. Also, in Table 4, the de-all data whose all three time-series components, i.e., the event, the seasonal, and the cyclical components have been removed in Step 11 are shown in column  $DA_{y,m}$ . Furthermore, the final fitted car sales data obtained from Steps 12 and 13 having included back the three components are displayed in the last column,  $TF_{y,m}$ , of Table 4.

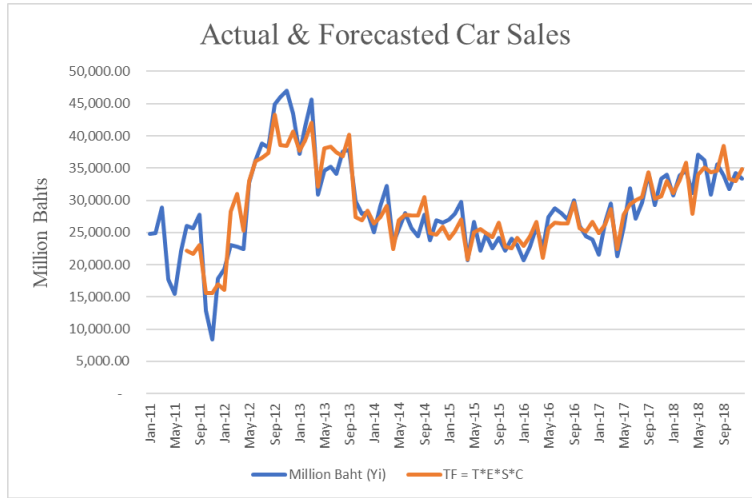


Figure 2: Actual Thailand’s monthly car sales and their forecasts by the event decomposition.

In Figure 2, the original Thailand’s monthly car sales as well as their final forecasted car sales data obtained from the proposed time-series event decomposition algorithm as shown in Table 4 are plotted together in Figure 2. The two curves appear closely next to each other as the calculated mean absolute percentage error (MAPE) of this forecast is relatively good at 8.17%.

## 4 Conclusion

In this article, a new time-series decomposition to incorporate the effects of special events is proposed. Thailand’s monthly car sales figures from the years 2011-2018 demonstrates a good application due to the large affective two incidents, namely, the national big flood and the tax-incentive program for first car buyers. These two incidents create four distinctive events or flags for the proposed method. Each flag’s event index obtained from the proposed computation method agrees well with the flag definition. Coupled with the seasonal, cyclical, and the trend components obtained, the final forecasts of the original car sales data prove very promising results with a low MAPE of just 8.17%. Ultimately, the event index computation proposed in this work can be beneficial to time-series decomposition forecasting when incidents having huge positive/negative impacts are present in the data.

## References

- [1] F.T.I Automotive Industry Club, The Federation of Thai Industries Automotive Industry Club. <http://www.aic.or.th>, 2016.
- [2] S. Chantaburee. 2016. [Online]. Available: [http://www.asean thai.net/ewt\\_news.php?nid=5720 & filename=index](http://www.asean thai.net/ewt_news.php?nid=5720 & filename=index). [Accessed 4 November 2018].
- [3] ASEAN Automotive Federation. [Online]. Available: <http://www.aseanautofed.com/statistics.html>. [Accessed December 2017].
- [4] Office of Industrial Economics 2018. [Online] Available:<http://www.oie.go.th/>. [Accessed 31 May 2018].
- [5] Thailand e-Government. [Online] Available: <https://www.egov.go.th/th/e-government-service/190/>. [Accessed 7 September 2018].
- [6] S. Ludvigson, The channel of Monetary Transmission to Demand: Evidence from the Market for Automobile Credit, *Journal of Money, Credit, and Banking*, 30(1998) 365-383.
- [7] J.M. Dargay and D.Gately, Income's effect on car and vehicle ownership, worldwide: 1960-2015, *Transportation Research Part A: Policy and Practice*, 33(1999) 101-138.
- [8] J.M. Dargay, The effect of income on car ownership: evidence of asymmetry, *Transportation Research Part A: Policy and Practice*, 35(2001) 807-821.
- [9] S. Shahabudin, Forecasting automobile sales, *Management Research News*, 32(2009) 670-682.
- [10] P. Wirotcheewan, A. Kengpol, K. Ishii, and Y. Shimada, Modelling and Forecasting for Automotive Parts Demand of Foreign Markets on Thailand, *Asian International Journal of Science And Technology in Production and Manufacturing Engineering: AIJSTPME* 4(2011) 1-13.
- [11] F. Muhammad, M. Yahya, M. Hussin and A. Ab Razak, Automobile Sales and Macroeconomic Variables: A Pooled Mean Group Analysis for ASEAN Countries, *Journal of Business and Management (IOSRJBM)*, 2(2012) 15-21.
- [12] J.D. Hamilton, A neoclassical model of unemployment and the business cycle, *Journal of Political Economy*. 96(1988) 593-617.
- [13] R.S. Pindyck, and J.J. Rotemberg, Dynamic factor demands and the effects of energy price shocks, *American Economic Review*, 73(1983) 1066-1079.
- [14] B.M. Barber, W. Reid, Click and M.N. Darrough, The impact of shocks to exchange rate and oil prices on U.S. sales of American and Japanese automakers, *Japan and World Economy*, 11(1999) 57-93.
- [15] K. Lee, and S. Ni, On the dynamic effects of oil price shocks: a study using industry level data, *Journal of Monetary Economics*, 49(2002) 823-852.

- [16] W. Rattanametawee, C. Leenawong, and P. Netisopakul, The effects of Special Events on Regression for Subcompact Car Sales in Thailand, *Jurnal Teknologi (Sciences & Engineering)* 78(2016) 161165.
- [17] T. Apiwattanachai and J. Pichitlamken, Time-Series Forecasting Model for Automobile Sales in Thailand, *Operations Research Network of Thailand*, (2008) 298-303.
- [18] W. Chaimongkol and C. Tansathit, Forecasting Model for Automobile Sale in Thailand, *The 2nd International Conference on Logistics and Transport & The 1st International Conference on Business and Economics Ridges Lakeland Resort Queenstown, Queenstown, New Zealand Theme: Managing Risk and Finance in Global Supply Chain*, (2010) 16-18.
- [19] B.L. Bowerman and R.T. O'Connell, *Forecasting and Time Series : An Applied Approach*, Wadsworth, Inc., USA, 1993.

(Received xx xx xx)

(Accepted xx xx xx)

## ประวัติผู้เขียน

ชื่อ : นางสาววิชญา รัตนเมธาวิ  
รหัสนักศึกษา : 60605092  
เชื้อชาติ : ไทย  
สัญชาติ : ไทย  
ศาสนา : พุทธ  
ที่อยู่ : 125 หมู่ 11 ถนนมหาสารคาม-วาปีปทุม อำเภอเมือง  
จังหวัดมหาสารคาม 44000  
e-mail : witchaya232@hotmail.com