ประสิทธิภาพของการพยากรณ์โดยตัวแบบเชิงเส้นนัยทั่วไปเมื่อข้อมูลมีการแจกแจงแบบเบ้ขวากับการประยุกต์สำหรับปริมาณการใช้น้ำมันสำเร็จรูปของประเทศไทย

วิทยานิพนธ์

ของ

อัมพิกา ก้อนแก้ว

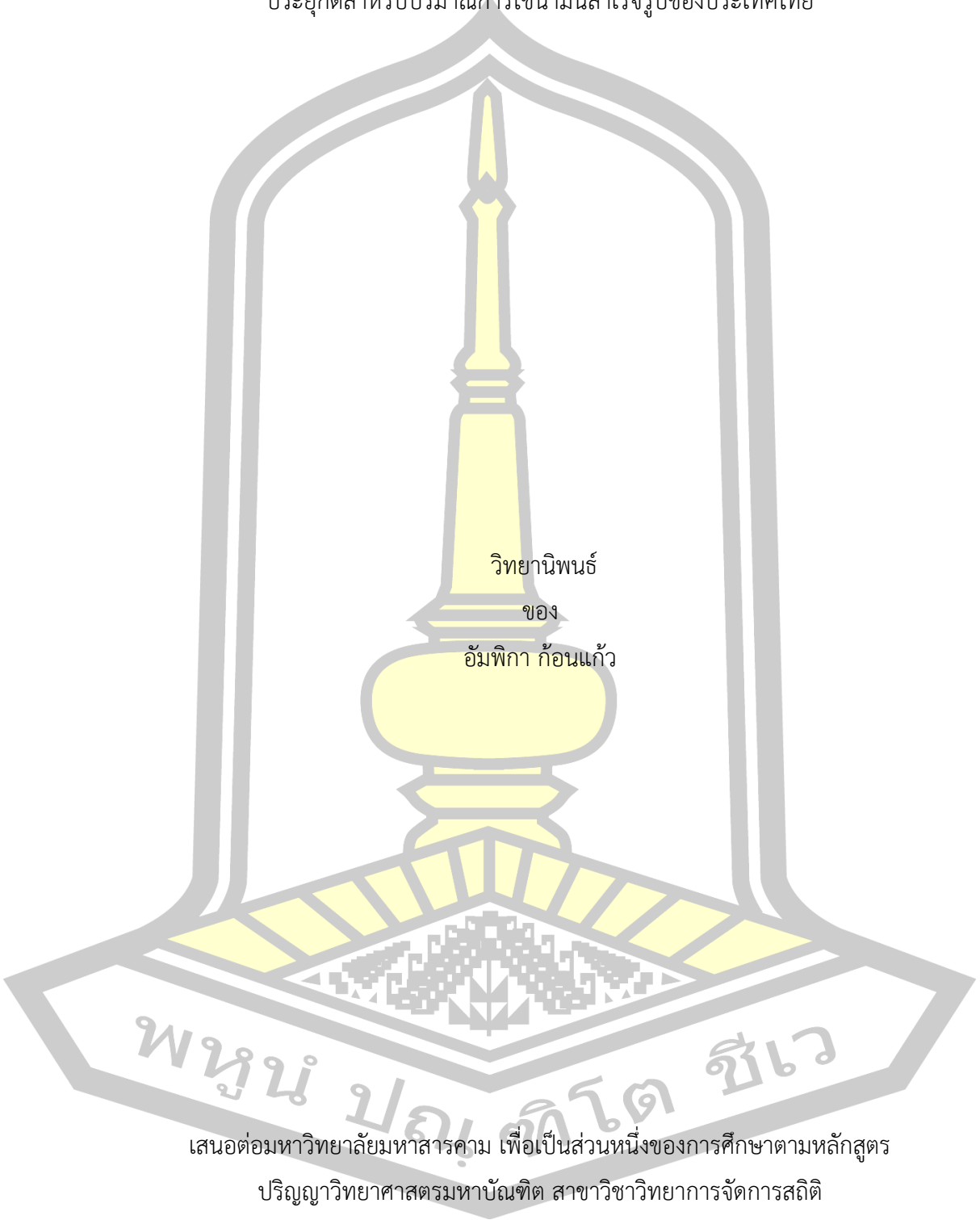
เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการจัดการสถิติ

มิถุนายน 2562

สงวนลิขสิทธิ์เป็นของมหาวิทยาลัยมหาสารคาม

ประสิทธิภาพของการพยากรณ์โดยตัวแบบเชิงเส้นนัยทั่วไปเมื่อข้อมูลมีการแจกแจงแบบเบ้ขวากับการประยุกต์สำหรับปริมาณการใช้น้ำมันสำเร็จรูปของประเทศไทย



วิทยานิพนธ์

ของ

อัมพิกา ก้อนแก้ว

เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

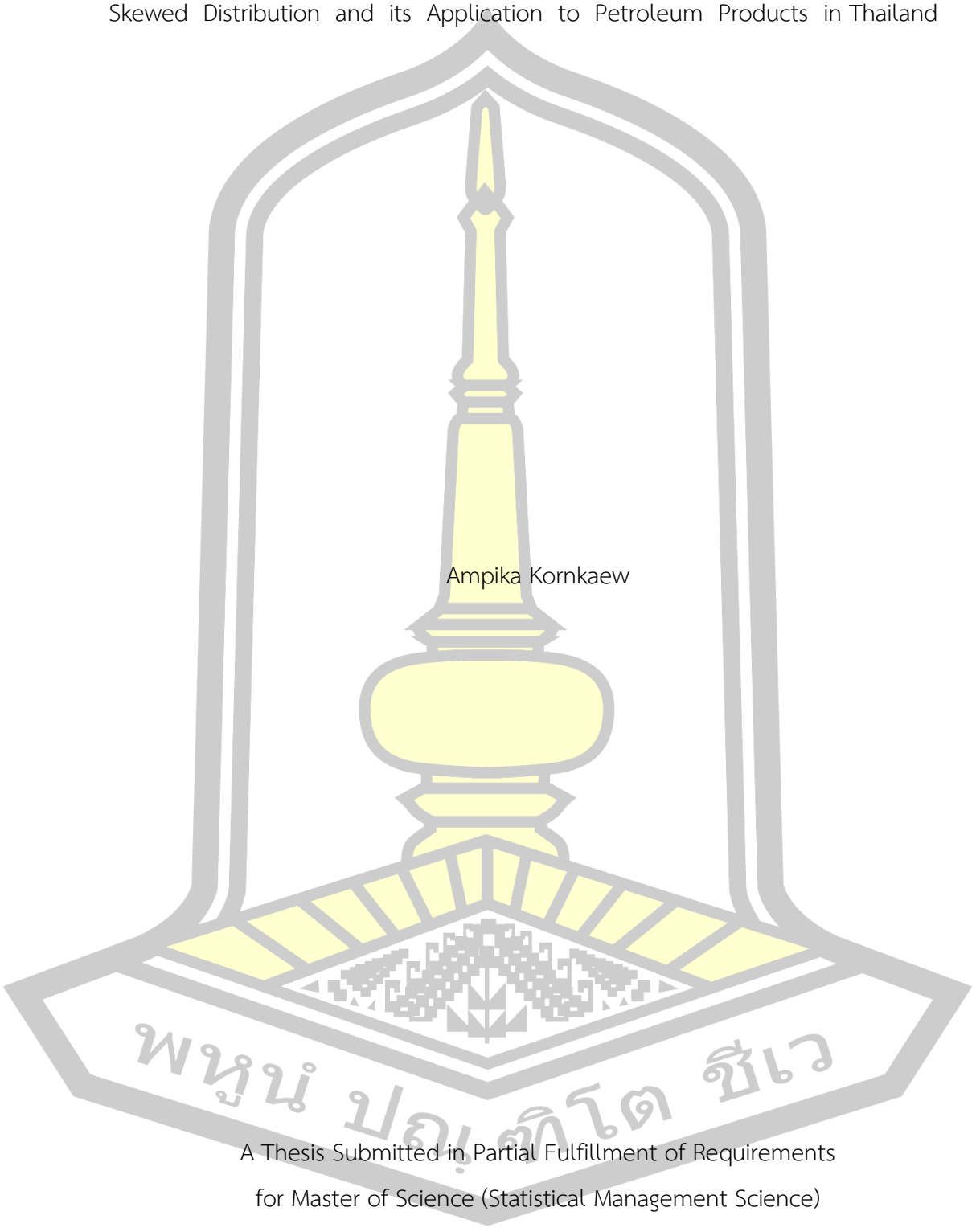
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการจัดการสถิติ

มิถุนายน 2562

สงวนลิขสิทธิ์เป็นของมหาวิทยาลัยมหาสารคาม

Efficiency of Forecasting using Generalized Linear Model with Positively Skewed Distribution and its Application to Petroleum Products in Thailand

Ampika Kornkaew



A Thesis Submitted in Partial Fulfillment of Requirements

for Master of Science (Statistical Management Science)

June 2019

Copyright of Mahasarakham University

คณะกรรมการสอบวิทยานิพนธ์ ได้พิจารณาวิทยานิพนธ์ของนางสาวอัมพิกา ก้อนแก้ว แล้วเห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการจัดการสถิติ ของมหาวิทยาลัยมหาสารคาม

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ
(ผศ. ดร. มานัดถุ์ คำกอง)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(ผศ. ดร. สุจิตตา สุระภี)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม
(ดร. โรจนี หอมชาลี)

.....กรรมการ
(รศ. ดร. นิภาพร ชุติมันต์)

.....กรรมการ
(ดร. มนชยา เจียงประดิษฐ์)

มหาวิทยาลัยอนุมัติให้รับวิทยานิพนธ์ฉบับนี้ เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการจัดการสถิติ ของมหาวิทยาลัยมหาสารคาม

.....
(ศ. ดร. ไพโรจน์ ประมวล)
คณบดีคณะวิทยาศาสตร์

.....
(ผศ. ดร. กริสน์ ชัยมูล)
คณบดีบัณฑิตวิทยาลัย

พหุน ปณฺ ฑิโต ชีเว

พหุน ปณฺ ฑิโต ชีเว

ชื่อเรื่อง ประสิทธิภาพของการพยากรณ์โดยตัวแบบเชิงเส้นนัยทั่วไปเมื่อข้อมูลมีการแจกแจงแบบเบ้ขวากับการประยุกต์สำหรับปริมาณการใช้น้ำมันสำเร็จรูปของประเทศไทย

ผู้วิจัย อัมพิกา ก้อนแก้ว

อาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร. สุจิตตา สุระภี

ดร. โรจนี หอมชาลี

ปริญญา วิทยาศาสตรมหาบัณฑิต **สาขาวิชา** วิทยาการจัดการสถิติ

มหาวิทยาลัย มหาวิทยาลัยมหาสารคาม **ปีที่พิมพ์** 2562

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของการพยากรณ์เมื่อข้อมูลมีการแจกแจงแบบเบ้ขวา โดยการใช้การวิเคราะห์ตัวแบบเชิงเส้นนัยทั่วไป (GLMs) มีองค์ประกอบเชิงสุ่มที่มีการแจกแจงแกมมา และมีองค์ประกอบเชิงสุ่มที่มีการแจกแจงอินเวอร์สเกาส์เซียน และตัวแบบเชิงเส้นทั่วไป (GLM) โดยงานวิจัยประกอบไปด้วย 2 ส่วนคือ การจำลองข้อมูลและการประยุกต์ใช้กับข้อมูล ในส่วนของการจำลองข้อมูลจะกำหนดการแจกแจงของ Y ให้มี 2 การแจกแจงดังนี้ ให้ Y มีการแจกแจงแกมมาที่มีพารามิเตอร์รูปร่างมีค่าเท่ากับ 2 และพารามิเตอร์ขนาดมีค่าเท่ากับ 1 และ Y มีการแจกแจงอินเวอร์สเกาส์เซียนที่มีพารามิเตอร์แสดงตำแหน่งมีค่าเท่ากับ 1 และพารามิเตอร์รูปร่างมีค่าเท่ากับ 3 และกำหนดให้ตัวแปรอิสระให้มีการแจกแจงปรกติมาตรฐานที่มีพารามิเตอร์แสดงตำแหน่งหรือค่าเฉลี่ยมีค่าเท่ากับ 0 และพารามิเตอร์รูปร่างหรือความแปรปรวนมีค่าเท่ากับ 1 แล้วกำหนดขนาดตัวอย่างในการศึกษาเท่ากับ 30 50 และ 100 ตามลำดับ เปรียบเทียบตัวแบบทั้งสอง โดยพิจารณาค่า MAPE และ RMSE ที่ต่ำกว่าจะเป็นตัวแบบที่เหมาะสมมากกว่า ผลการศึกษาพบว่าเมื่อเปรียบเทียบประสิทธิภาพของการพยากรณ์ ทั้ง 2 ตัวแบบ ผลที่ได้ไม่ค่อยแตกต่างกันมากนักเมื่อพิจารณาที่ค่า MAPE ตัวแบบส่วนใหญ่ที่เหมาะสมจะเป็นตัวแบบเชิงเส้นทั่วไปที่มีการแปลงค่า y แต่เมื่อพิจารณาที่ค่า RMSE กลับพบว่าตัวแบบเชิงเส้นนัยทั่วไปจะให้ค่าที่เหมาะสมในการพยากรณ์มากกว่า

นอกจากนี้ได้นำข้อมูลมาประยุกต์ใช้กับปริมาณการใช้น้ำมันสำเร็จรูปในประเทศไทย เพื่อหาสมการที่เหมาะสมและหาปัจจัยที่ส่งผลต่อปริมาณการใช้น้ำมันสำเร็จรูป ทำการเก็บรวบรวมข้อมูลจากสำนักงานนโยบายและแผนพลังงานตั้งแต่เดือนมกราคม 2553 ถึง เดือนธันวาคม 2558 โดยการใช้การวิเคราะห์ตัวแบบเชิงเส้นนัยทั่วไปและตัวแบบเชิงเส้นทั่วไป ผลการศึกษาพบว่าปริมาณ

การใช้น้ำมันสำเร็จรูปโดยใช้ตัวแบบทั้ง 2 ตัวแบบ เมื่อพิจารณาค่า MAPE และ RMSE สามารถสรุปได้ดังนี้

- ตัวแบบที่เหมาะสมสำหรับปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91, ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95, ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20 ข้อมูลตั้งแต่มกราคม พ.ศ.2553 ถึง กุมภาพันธ์ พ.ศ.2556 , ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E85, ปริมาณการใช้น้ำมันเบนซิน 91, ปริมาณการใช้น้ำมันเบนซิน 95 และปริมาณการใช้แก๊ส NGV คือ ตัวแบบเชิงเส้นทั่วไป

- ตัวแบบที่เหมาะสมสำหรับปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20 ข้อมูลตั้งแต่มีนาคม พ.ศ.2556 ถึง ธันวาคม พ.ศ.2558 และปริมาณการใช้แก๊ส LPG คือ ตัวแบบเชิงเส้นนัยทั่วไป

คำสำคัญ : ตัวแบบเชิงเส้นนัยทั่วไป, ตัวแบบเชิงเส้นทั่วไป, ฟังก์ชันเชื่อมโยง, การแจกแจงแบบแกมมา , การแจกแจงแบบอินเวอร์เกาส์เซียน, ปริมาณการใช้น้ำมัน

พหุน ปณ ฑิโต ชีเว

TITLE Efficiency of Forecasting using Generalized Linear Model with Positively Skewed Distribution and its Application to Petroleum Products in Thailand

AUTHOR Ampika Kornkaew

ADVISORS Assistant Professor Sujitta Surapee , Ph.D.
Rojanee Homchalee , Ph.D.

DEGREE Master of Science **MAJOR** Statistical Management Science

UNIVERSITY Mahasarakham University **YEAR** 2019

ABSTRACT

The objective of this research is comparing the efficiency of forecasting with positively skewed using Generalized Linear Model (GLMs) and General Linear Model (GLM). For GLMs, random component is defined to be Gamma and Inverse Gaussian distribution. The research consists of 2 parts of data which are simulation and application to forecast amount of petroleum products. For the first part, Y are defined in two distributions that are Gamma distribution with shape and scale parameters equal to be 2 and 1 respectively and Y is the Inverse Gaussian distribution with location and shape parameters equal to be 1 and 3 respectively. Let X be an independent variable which is Standard Normal distribution with mean and variance parameters are 0 and 1. From both cases, size of sample is 30, 50 and 100, respectively. The result of comparison models using minimum MAPE and RMSE showed that both models are not significantly different in forecasting efficiency. When we consider the MAPE values, General Linear Model is more suitable than Generalized linear mode, but when we consider the RMSE value, Generalized Linear Model (GLMs) is more efficient than General Linear Mode.

Moreover, the application of this model is finding the appropriate

equations and factors that affected to the amount of petroleum products in Thailand. Data were collected from the Energy Policy and Planning Office from January 2553 to December 2558. The results showed that both models by considering MAPE and RMSE values can be summarized as follows.

- The models that are suitable for forecasting consumption of gasohol 91, consumption of gasohol 95, consumption of gasohol E20 with data from January 2010 to February 2013, consumption of gasohol E85, consumption of gasoline 91, consumption of gasoline 91 and consumption of NGV are a General Linear Model.

- The models that are suitable for forecasting consumption of gasohol E20 with data from March 2013 to December 2015 and consumption of LPG are a Generalized Linear Model.

Keyword : Generalized Linear Model, General linear model, Link function, Gamma distribution, Inverse gaussian distribution, Consumption petroleum products

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จสมบูรณ์ได้ด้วยความกรุณาและความช่วยเหลืออย่างสูงยิ่งจาก ผู้ช่วยศาสตราจารย์ ดร. สุจิตตา สุระภี ประธานกรรมการควบคุมวิทยานิพนธ์ อาจารย์ ดร. โรจนี หอมชาลี กรรมการควบคุมวิทยานิพนธ์ ผู้ช่วยศาสตราจารย์ ดร. มานัดถุ์ คำกอง ประธานกรรมการสอบ และรองศาสตราจารย์ ดร. นิภาพร ชุติมันต์ และอาจารย์ ดร. มนชยา เจียงประดิษฐ์ กรรมการสอบวิทยานิพนธ์

ขอขอบพระคุณอาจารย์นิภาดา พาภักดี สาขาสถิติประยุกต์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตขอนแก่น นางสาวประภาวรรณ เสนาเพ็ง และพี่ๆ เพื่อนๆน้องๆ ร่วมสาขา ซึ่งคอยสนับสนุนและเป็นกำลังใจในการศึกษามาโดยตลอด จนทำให้วิทยานิพนธ์ฉบับนี้สำเร็จผู้วิจัยขอขอบพระคุณไว้ ณ โอกาสนี้ด้วย

อัมพิกา ก้อนแก้ว

พหูน ปณฺฑิโต ชีเว

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... ฉ

กิตติกรรมประกาศ ..... ช

สารบัญ ..... ฌ

สารบัญตาราง ..... ฎ

สารบัญภาพ ..... ฑ

บทที่ 1 บทนำ 1

1.1 หลักการและเหตุผล 1

1.2 วัตถุประสงค์ของการวิจัย 4

1.3 ขอบเขตของการวิจัย 5

1.4 ประโยชน์ของการวิจัย 6

1.5 นิยามศัพท์ 6

บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง 7

2.1 ตัวแบบเชิงเส้นทั่วไป (General Linear Model : GLM) 7

2.1.1 ตัวประมาณค่าพารามิเตอร์ของตัวแบบเชิงเส้นทั่วไป 8

2.1.2 เกณฑ์ที่ใช้พิจารณาความเหมาะสมของตัวแบบเชิงเส้นทั่วไป 9

2.1.3 สถิติที่ใช้ทดสอบความมีนัยสำคัญของพารามิเตอร์ในตัวแบบเชิงเส้นทั่วไป 9

2.2 ตัวแบบเชิงเส้นนัยทั่วไป (Generalized Linear Model : GLMs) 10

2.2.1 ตัวประมาณค่าพารามิเตอร์ของตัวแบบเชิงเส้นนัยทั่วไป 15

2.2.2 เกณฑ์ที่ใช้พิจารณาความเหมาะสมของตัวแบบเชิงเส้นนัยทั่วไป 17

2.2.3 สถิติที่ใช้ทดสอบความมีนัยสำคัญของพารามิเตอร์ในตัวแบบเชิงเส้นนัยทั่วไป 17

2.3 ลักษณะการแจกแจงที่ศึกษา 18
2.4 ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation coefficient) 20
2.5 การคัดเลือกตัวแปรอิสระด้วยวิธีการถดถอยแบบขั้นบันได 20
2.6 เกณฑ์ที่ใช้ในการคัดเลือกตัวแบบ 21
2.7 ข้อมูลน้ำมันสำเร็จรูปในประเทศไทย 22
2.8 งานวิจัยที่เกี่ยวข้อง 26
บทที่ 3 วิธีดำเนินการวิจัย 29
3.1 แผนการดำเนินงานวิจัยในการสร้างแบบจำลอง 29
3.2 ขั้นตอนในการประยุกต์ใช้ตัวแบบเชิงเส้นทั่วไปและตัวแบบเชิงเส้นนัยทั่วไป 34
บทที่ 4 ผลการวิจัย 38
4.1 ผลการจำลองข้อมูลเพื่อหาตัวแบบที่มีประสิทธิภาพในการพยากรณ์ เมื่อข้อมูลมีการแจกแจงแบบเบ้ขวา 38
4.2 ผลการวิเคราะห์หาสมการที่เหมาะสมในการพยากรณ์ปริมาณการใช้น้ำมันสำเร็จรูปและ หาปัจจัยที่ส่งผลต่อปริมาณการใช้น้ำมันสำเร็จรูปของประเทศไทย 54
บทที่ 5 สรุปผล อภิปรายผล และข้อเสนอแนะ 89
5.1 สรุปผล 89
5.2 อภิปรายผล 107
5.3 ข้อเสนอแนะ 108
บรรณานุกรม 110
ภาคผนวก 113
ภาคผนวก ก พยากรณ์ปริมาณการใช้น้ำมันสำเร็จรูปของประเทศไทยจำนวน 72 เดือน ตั้งแต่ปี พ.ศ.2553 ถึง พ.ศ.2558 114
ประวัติผู้เขียน 142

สารบัญตาราง

หน้า

ตาราง 2.1 ชนิดของตัวแบบสำหรับการวิเคราะห์ทางสถิติ 14

ตาราง 2.2 เกณฑ์การประเมินผลการพยากรณ์โดยค่าเปอร์เซ็นต์ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย 21

ตาราง 3.3 ตารางการกำหนดสถานการณ์ต่างๆ ในการจำลองข้อมูล 32

ตาราง 4.4 ค่าเฉลี่ยของค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อน (MAPE) และรากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE) จำแนกตามขนาดตัวอย่าง ชนิดของตัวแบบและรูปแบบการแปลงค่าของ y เมื่อกำหนดให้ y มีการแจกแจงแกมมา (2,1) กรณีความสัมพันธ์ระหว่าง x และ y อยู่ในระดับต่ำ 39

ตาราง 4.5 ค่าเฉลี่ยของค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อน (MAPE) และรากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE) จำแนกตามขนาดตัวอย่าง ชนิดของตัวแบบและรูปแบบการแปลงค่าของ y เมื่อกำหนดให้ y มีการแจกแจงแกมมา (2,1) กรณีความสัมพันธ์ระหว่าง x และ y อยู่ในระดับสูง 41

ตาราง 4.6 ค่าเฉลี่ยของค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อน (MAPE) และรากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE) จำแนกตามขนาดตัวอย่าง ชนิดของตัวแบบและรูปแบบการแปลงค่าของ y เมื่อกำหนดให้ y มีการแจกแจงแกมมา (2,1) กรณีมีตัวแปรอิสระ 2 ตัวที่มีระดับความสัมพันธ์อยู่ในระดับต่ำ 43

ตาราง 4.7 ค่าเฉลี่ยของค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อน (MAPE) และรากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE) จำแนกตามขนาดตัวอย่าง ชนิดของตัวแบบและรูปแบบการแปลงค่าของ y เมื่อกำหนดให้ y มีการแจกแจงแกมมา (2,1) กรณีมีตัวแปรอิสระ (x) 2 ตัวที่มีระดับความสัมพันธ์อยู่ในระดับสูง 45

ตาราง 4.8 ค่าเฉลี่ยของค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อน (MAPE) และรากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE) จำแนกตามขนาดตัวอย่าง ชนิดของตัวแบบและรูปแบบการแปลงค่าของ y เมื่อกำหนดให้ y มีการแจกแจงอินเวอร์เกาส์เซียน (1,3) กรณีความสัมพันธ์ของ x และ y อยู่ในระดับต่ำ 47

ตาราง 4.9 ค่าเฉลี่ยของค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อน (MAPE) และรากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE) จำแนกตามขนาดตัวอย่าง ชนิดของตัวแบบและรูปแบบการ

แปลงค่าของ y เมื่อกำหนดให้ y มีการแจกแจงอินเวอร์เกาส์เซียน (1,3) กรณีความสัมพันธ์ของ x และ y อยู่ในระดับสูง 49

ตาราง 4.10 ค่าเฉลี่ยของค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อน (MAPE) และรากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE) จำแนกตามขนาดตัวอย่าง ชนิดของตัวแบบและรูปแบบการแปลงค่าของ y เมื่อกำหนดให้ y มีการแจกแจงอินเวอร์เกาส์เซียน (1,3) กรณีมีตัวแปรอิสระ (x) 2 ตัวที่มีระดับความสัมพันธ์อยู่ในระดับต่ำ 51

ตาราง 4.11 ค่าเฉลี่ยของค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อน (MAPE) และรากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE) จำแนกตามขนาดตัวอย่าง ชนิดของตัวแบบและรูปแบบการแปลงค่าของ y เมื่อกำหนดให้ y มีการแจกแจงอินเวอร์เกาส์เซียน (1,3) กรณีมีตัวแปรอิสระ (x) 2 ตัวที่มีระดับความสัมพันธ์อยู่ในระดับสูง 53

ตาราง 4.12 สมการพยากรณ์ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91 และค่าสถิติแสดงความคลาดเคลื่อนและความเหมาะสมของตัวแบบพยากรณ์ 55

ตาราง 4.13 สมการพยากรณ์ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95 และค่าสถิติแสดงความคลาดเคลื่อนและความเหมาะสมของตัวแบบพยากรณ์ 58

ตาราง 4.14 สมการพยากรณ์*ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20* ข้อมูลตั้งแต่มกราคม พ.ศ.2553 ถึง กุมภาพันธ์ พ.ศ.2556 และค่าสถิติแสดงความคลาดเคลื่อนและความ 62

ตาราง 4.15 สมการพยากรณ์*ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20* ข้อมูลตั้งแต่มีนาคม พ.ศ.2556 ถึง ธันวาคม พ.ศ.2558 และค่าสถิติแสดงความคลาดเคลื่อนและความ 65

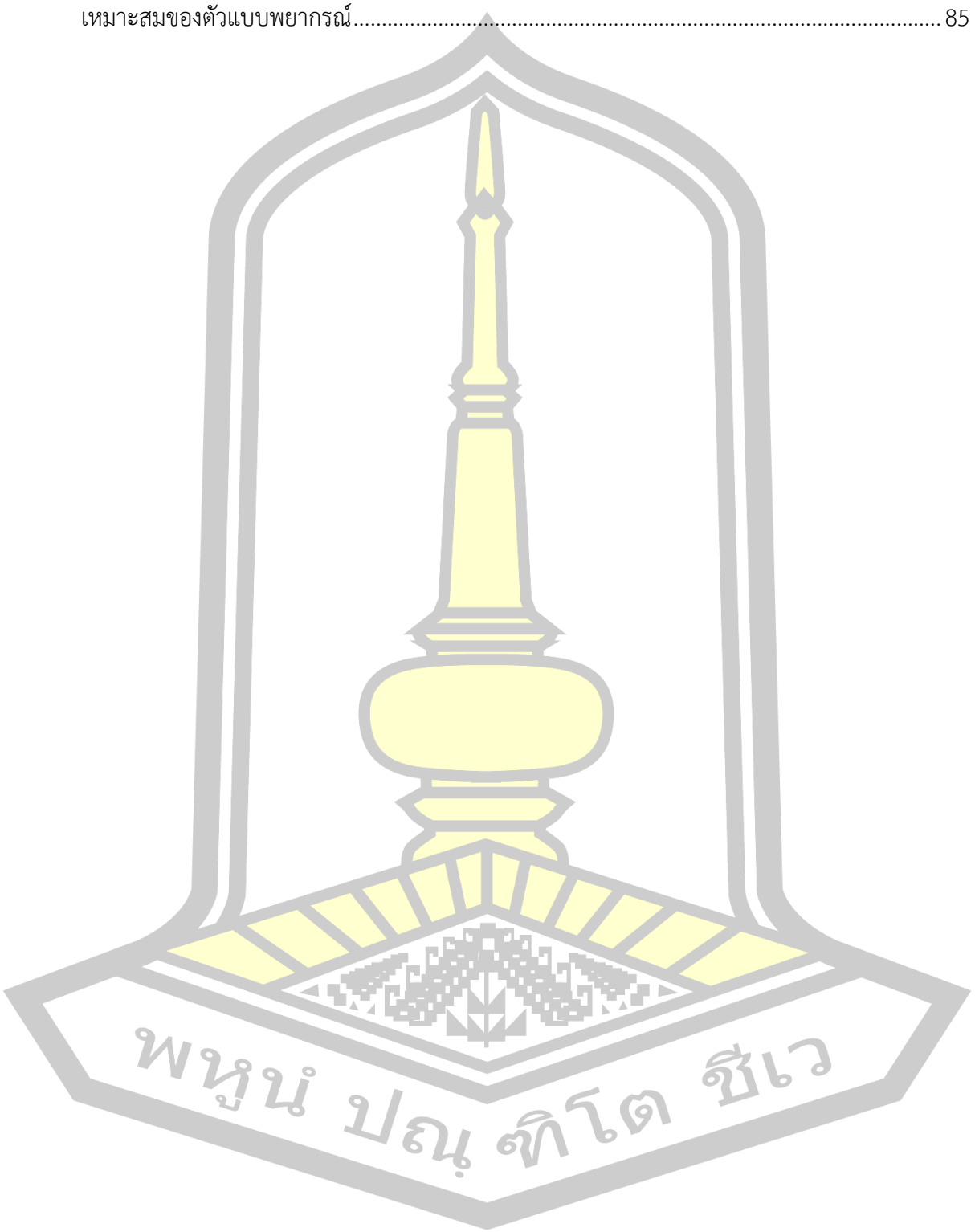
ตาราง 4.16 สมการพยากรณ์ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E85 และค่าสถิติแสดงความคลาดเคลื่อนและความเหมาะสมของตัวแบบพยากรณ์ 69

ตาราง 4.17 สมการพยากรณ์ปริมาณการใช้น้ำมันเบนซิน 91 และค่าสถิติแสดงความคลาดเคลื่อนและความเหมาะสมของตัวแบบพยากรณ์ 72

ตาราง 4.18 สมการพยากรณ์ปริมาณการใช้น้ำมันเบนซิน 95 และค่าสถิติแสดงความคลาดเคลื่อนและความเหมาะสมของตัวแบบพยากรณ์ 76

ตาราง 4.19 สมการพยากรณ์ปริมาณการใช้แก๊ส LPG และค่าสถิติแสดงความคลาดเคลื่อนและความเหมาะสมของตัวแบบพยากรณ์ 81

ตาราง 4.20 สมการพยากรณ์ปริมาณการใช้แก๊ส NGV และค่าสถิติแสดงความคลาดเคลื่อนและความเหมาะสมของตัวแบบพยากรณ์..... 85



สารบัญภาพ

หน้า

รูปภาพที่ 2.1 กราฟแสดงรูปร่างของฟังก์ชันความน่าจะเป็นของการแจกแจงแกมมา 19

รูปภาพที่ 2.2 กราฟแสดงรูปร่างของฟังก์ชันความน่าจะเป็นของการแจกแจงอินเวอร์สเกาส์เซียน... 19

รูปภาพที่ 2.3 ปริมาณการใช้น้ำมันสำเร็จรูปในประเทศไทย (พันบาร์เรล/วัน) 26

รูปภาพที่ 3.4 ขั้นตอนการจำลองข้อมูลเพื่อเปรียบเทียบประสิทธิภาพของตัวแบบเชิงเส้นทั่วไป 31

รูปภาพที่ 3.5 ขั้นตอนการหาสมการที่เหมาะสมในการพยากรณ์ปริมาณการใช้น้ำมันสำเร็จรูป 37

พหุน ปณฺฑิโต ชีเว

บทที่ 1
บทนำ

1.1 หลักการและเหตุผล

ในปัจจุบันเทคนิคการพยากรณ์โดยใช้หลักการทางสถิติถูกพัฒนามาอย่างต่อเนื่อง เนื่องจากเทคนิคการพยากรณ์มีประโยชน์อย่างมากในการวางแผนและตัดสินใจในการดำเนินงานทั้งระยะสั้นและระยะยาว และการพยากรณ์ยังเป็นเครื่องมืออย่างหนึ่งที่จะช่วยให้ได้ข้อมูลในอนาคตเพื่อประกอบการวางแผนจึงเป็นที่ยอมรับกันว่าการพยากรณ์มีบทบาทสำคัญทั้งในงานของภาครัฐและภาคเอกชน วิธีการพยากรณ์แต่ละวิธีที่มีประสิทธิภาพจะพิจาณาจากค่าความคลาดเคลื่อนน้อยที่สุดของการพยากรณ์ โดยทั่วไปเทคนิคหรือตัวแบบที่ใช้ในการพยากรณ์จะแบ่งเป็น 2 ประเภท ซึ่งจะขึ้นอยู่กับลักษณะของข้อมูล ถ้าข้อมูลเป็นข้อมูลอนุกรมเวลาก็จะใช้เทคนิคการวิเคราะห์ข้อมูลอนุกรมเวลา โดยใช้ตัวแบบอนุกรมเวลา (Time Series Model) แต่ถ้าข้อมูลเป็นข้อมูลภาคตัดขวาง (Cross Sectional Data) ซึ่งเป็นการศึกษาความสัมพันธ์ของตัวแปรต่างๆ ที่เกิดขึ้นในช่วงเวลาเดียวกันจะใช้เทคนิคการวิเคราะห์ตัวแบบเชิงเส้นทั่วไป เช่น ตัวแบบการถดถอยเชิงเส้น ซึ่งการเลือกวิธีการพยากรณ์ที่เหมาะสมควรพิจารณาจากหลายๆ ปัจจัย เช่น ถ้าใช้ตัวแบบอนุกรมเวลาจะต้องพิจารณาช่วงการพยากรณ์ที่ต้องการ เวลาที่ใช้ในการพยากรณ์ ลักษณะข้อมูลอนุกรมเวลาและขนาดอนุกรมเวลา (มุกดา แม้นมินทร์, 2549) สำหรับวิธีการพยากรณ์อนุกรมเวลาที่นิยมใช้มีหลายวิธี ได้แก่ วิธีการปรับให้เรียบ การวิเคราะห์การถดถอย และวิธีของบอกซ์-เจนกินส์ เป็นต้น (ทรงศิริ แต้สมบัติ, 2549) แต่เนื่องจากการเก็บรวบรวมข้อมูลอนุกรมเวลาซึ่งเป็นข้อมูลที่เกิดขึ้นในอดีต อาจจะไม่มีความครบถ้วน ข้อมูลไม่แน่นอนหรือข้อมูลมีความคลุมเครือ อาจทำให้เกิดความคลาดเคลื่อนสูงในการพยากรณ์ ส่วนการวิเคราะห์การถดถอยเชิงเส้นตรงซึ่งสามารถแบ่งได้เป็น 2 ประเภท ได้แก่ การวิเคราะห์การถดถอยเชิงเส้นตรงแบบง่าย ซึ่งมีตัวแปรอิสระเพียงตัวเดียวกับการวิเคราะห์การถดถอยพหุคูณเชิงเส้นตรง ซึ่งมีตัวแปรอิสระตั้งแต่ 2 ตัวขึ้นไป ทั้ง 2 ประเภทต่างก็เป็นการหาความสัมพันธ์และอิทธิพลระหว่างตัวแปรอิสระกับตัวแปรตามที่เกิดขึ้นในช่วงเวลาเดียวกันอันจะนำไปสู่การพยากรณ์การเปลี่ยนแปลงของตัวแปรตามที่เป็นผลจากตัวแปรอิสระ โดยการวิเคราะห์การถดถอยนี้จะต้องพิจารณาด้วยว่าข้อมูลเป็นไปตามข้อตกลงเบื้องต้นหรือไม่

ตัวแบบเชิงเส้นนัยทั่วไป (Generalized Linear Model : GLMs) เป็นตัวแบบที่ขยายมาจากตัวแบบเชิงเส้นทั่วไป (Classical Linear Model) ซึ่งเป็นตัวแบบพื้นฐานของการสร้างตัวแบบเชิงสถิติ (Statistical Modeling) ที่ใช้ในการอธิบายโครงสร้างความสัมพันธ์ของตัวแปรต่างๆ ทั้งที่เป็นตัวแปรต่อเนื่องและตัวแปรไม่ต่อเนื่อง โดยมีส่วนประกอบด้วยกัน 3 ส่วน คือ ส่วนที่หนึ่ง คือ

ส่วนประกอบเชิงสุ่ม (Random Component) ซึ่งเป็นส่วนที่ใช้ในการแสดงการแจกแจงความน่าจะเป็นของตัวแปรตอบสนอง (Response Variable) ที่นอกจากจะมีการแจกแจงปรกติแล้วยังสามารถขยายไปสู่การแจกแจงอื่นๆ ที่อยู่ในวงศ์เลขชี้กำลัง (Exponential Family) เช่น การแจกแจงทวินาม (Binomial Distribution) การแจกแจงปัวซง (Poisson Distribution) การแจกแจงแกมมา (Gamma Distribution) และการแจกแจงอินเวอร์สเกาส์เซียน (Inverse – Gaussian Distribution) เป็นต้น ส่วนที่สองเป็นส่วนประกอบเชิงระบบ (Systematic Component) ส่วนนี้จะแสดงฟังก์ชันเชิงเส้นของตัวแปรอิสระที่ใช้เป็นตัวแปรพยากรณ์ในตัวแบบ และส่วนที่สาม คือส่วนฟังก์ชันเชื่อมโยง (Link Function) เป็นส่วนที่ใช้ในการอธิบายฟังก์ชันความสัมพันธ์ระหว่างส่วนประกอบแบบมีระบบกับค่าเฉลี่ยของส่วนประกอบเชิงสุ่ม ที่นอกจากจะใช้ฟังก์ชันเอกลักษณ์ (Identical Link) เป็นฟังก์ชันเชื่อมโยงแล้วยังสามารถขยายไปใช้กับฟังก์ชันเชื่อมโยงอื่นๆ อีกหลายแบบที่เป็นฟังก์ชันทางเดียวและหาอนุพันธ์ได้ (Monotonic Differentiable Function) ตัวแบบเชิงเส้นนัยทั่วไปนี้สามารถขจัดปัญหาที่เกิดจากสหสัมพันธ์ระหว่างตัวแปรอิสระที่อาจมีความสัมพันธ์ซึ่งกันและกัน และนอกจากนั้นยังสามารถใช้ตรวจสอบผลกระทบจากความสัมพันธ์ซึ่งกันและกันของตัวแปรอิสระได้ โดยตัวแบบเชิงเส้นนัยทั่วไป ถ้าตัวแปรตามที่เป็นตัวแปรเชิงปริมาณแบบต่อเนื่องไม่ได้มีการแจกแจงปรกติจะตรวจสอบก่อนว่ามีการแจกแจงแบบอื่นๆ ที่อยู่ในวงศ์เลขชี้กำลังหรือไม่ เช่น การแจกแจงแกมมาและการแจกแจงอินเวอร์สเกาส์เซียน ซึ่งเป็นการแจกแจงที่มีลักษณะเบ้ขวา และนอกจากนี้ตัวแบบเชิงเส้นนัยทั่วไปยังสามารถใช้ได้กับตัวแปรตอบสนองที่เป็นข้อมูลเชิงกลุ่ม (Categorical Data) และข้อมูลที่เป็นจำนวนนับ (Count Data)

ด้วยเหตุนี้ผู้วิจัยจึงต้องการศึกษาการเปรียบเทียบประสิทธิภาพของการพยากรณ์เมื่อตัวแปรตามหรือตัวแปรที่เราสนใจศึกษาไม่ได้มีการแจกแจงปรกติแต่มีการแจกแจงแบบเบ้ขวา และใช้วิธีการวิเคราะห์ตัวแบบเชิงเส้นทั่วไป (General Linear Model : GLM) เปรียบเทียบกับวิธีการใช้การวิเคราะห์ตัวแบบเชิงเส้นนัยทั่วไป (Generalized Linear Model) แล้วเปรียบเทียบประสิทธิภาพของทั้งสองวิธี รวมทั้งจะประยุกต์ใช้กับข้อมูลจริงในการเลือกตัวแบบที่เหมาะสมในการพยากรณ์ปริมาณการใช้น้ำมันสำเร็จรูปของประเทศไทยซึ่งเป็นข้อมูลเชิงปริมาณแบบต่อเนื่อง

เนื่องจากในยุคปัจจุบันน้ำมันถือได้ว่าเป็นพลังงานเชื้อเพลิงที่เป็นปัจจัยที่สำคัญมากกว่าพลังงานชนิดอื่นๆ ที่ทุกประเทศต้องการ โดยน้ำมันมีบทบาทต่อการดำเนินชีวิตประจำวันเป็นอย่างมากทั้งในภาคอุตสาหกรรม ภาคครัวเรือน ภาคขนส่ง ที่มีความต้องการเพิ่มขึ้นอย่างต่อเนื่อง เห็นได้จากปริมาณความต้องการของน้ำมันสำเร็จรูป ซึ่งได้แก่ น้ำมันเบนซิน น้ำมันดีเซล และน้ำมันแก๊สโซฮอล์ จนเริ่มตระหนักถึงพลังงานที่แปรรูปจากธรรมชาติ (น้ำมันดิบ) กำลังจะหมดไป เนื่องจากน้ำมันดิบเป็นทรัพยากรเชื้อเพลิงที่มีอยู่อย่างจำกัดและเป็นชนิดใช้แล้วหมดไป มนุษย์ไม่สามารถผลิตขึ้นมาทดแทนได้ เพราะการเกิดขึ้นของน้ำมันดิบเกิดจากการทับถมของทรัพยากรต่างๆ

ทางธรรมชาติต้องใช้ระยะเวลายาวนานเป็นหลายพันปีหรือหลายหมื่นปี (กมลวรรณ สารพานิช, 2555) จึงได้มีการคิดหาวิธีอื่นๆ ในการสร้างพลังงานทดแทนขึ้นมา ซึ่งประเทศไทยก็เป็นประเทศหนึ่งที่มีความต้องการใช้น้ำมันไม่ต่างจากประเทศต่างๆ จึงทำให้ประเทศไทยเป็นประเทศที่มีความต้องการน้ำมันทั้งการนำเข้าและการส่งออกของผลิตภัณฑ์น้ำมันสำเร็จรูป

จากเหตุผลข้างต้นทำให้ผู้วิจัยศึกษาปัจจัยที่มีผลต่อการใช้ปริมาณน้ำมันสำเร็จรูปในประเทศไทย จึงได้มีการทบทวนงานวิจัยที่เกี่ยวข้องของ (รามนรี ภูดีบุตร, 2550) ได้หาตัวแบบการพยากรณ์สำหรับการใช้พลังงานของประเทศไทย โดยวิธีการปรับให้เรียบ (Smoothing Method) และวิธีการของบอกซ์-เจนกินส์ (Box-Jenkins's Method) โดยใช้ข้อมูลการใช้พลังงานของประเทศไทยซึ่งเก็บรวบรวมเป็นรายเดือนตั้งแต่เดือนมกราคม พ.ศ.2544 ถึงเดือนกันยายน พ.ศ.2550 ผลการศึกษาพบว่า สำหรับข้อมูลการใช้พลังงานขั้นปฐมภูมิ 4 ชนิด วิธีการที่เหมาะสมคือ บอกซ์-เจนกินส์ ตัวแบบการพยากรณ์เป็นดังนี้ ตัวแบบสำหรับข้อมูลถ่านหิน คือ AR(1), ลิกไนต์ คือ AR(1), ปิโตรเลียม คือ ARIMA(0,1,1), ก๊าซธรรมชาติ คือ AR(1) และข้อมูลการใช้พลังงานขั้นสุดท้าย 10 ชนิด วิธีการที่เหมาะสม คือ บอกซ์-เจนกินส์ ตัวแบบการพยากรณ์เป็นดังนี้ ตัวแบบสำหรับข้อมูลถ่านหิน คือ ARIMA(0,1,1), ลิกไนต์ คือ ARIMA(1,1,0), ก๊าซปิโตรเลียมเหลว คือ ARIMA(0,1,1), น้ำมันเบนซิน (พิเศษ) 95 รวมแก๊สโซฮอล์ 95 คือ ARIMA(1,1,0), น้ำมันเบนซิน (ธรรมดา) 91 รวมแก๊สโซฮอล์ รวมแก๊สโซฮอล์ 91 คือ ARIMA(1,1,0), น้ำมันเครื่องบิน คือ AR(1), น้ำมันก๊าด คือ AR(1), น้ำมันดีเซล คือ AR(1), น้ำมันเตา คือ ARIMA(0,1,1) และก๊าซธรรมชาติ คือ ARIMA(1,1,1) เมื่อทราบตัวแบบที่เหมาะสมแล้วสามารถใช้ค่าพยากรณ์มาประกอบการวางแผน เพื่อลดการใช้พลังงานของประเทศไทยในอนาคต

ต่อมา (สัณฐิติ ทองช่วง และคณะ, 2553) ได้ศึกษาการวิเคราะห์ปริมาณการใช้แก๊สโซฮอล์ E20 ในประเทศไทย โดยใช้ข้อมูลทุติยภูมิในช่วงปี พ.ศ. 2551 - 2552 จากแหล่งต่างๆ ของหน่วยงานที่เกี่ยวข้องและใช้วิธีการวิเคราะห์ข้อมูลเชิงปริมาณในรูปแบบความสัมพันธ์เชิงเส้นโดยสมการความถดถอยเชิงพหุ เพื่อวิเคราะห์ปัจจัยที่มีผลต่อปริมาณการใช้แก๊สโซฮอล์ E20 ในประเทศไทย ผลการศึกษาพบว่าปัจจัยที่มีผลต่อปริมาณการใช้แก๊สโซฮอล์ E20 โดยมีความสัมพันธ์ ไปในทิศทางเดียวกันกับปริมาณการใช้แก๊สโซฮอล์ E20 อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 คือ ราคาขายปลีกเฉลี่ยของน้ำมันเบนซิน 95 ส่วนปัจจัยที่มีผลต่อปริมาณการใช้แก๊สโซฮอล์ E20 โดยมีความสัมพันธ์ไปในทิศทางเดียวกันกับปริมาณการใช้แก๊สโซฮอล์ E20 อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 คือ ราคาขายปลีกเฉลี่ยของน้ำมันเบนซิน 91 ส่วนราคาขายปลีกเฉลี่ยของแก๊สโซฮอล์ 91 มีความสัมพันธ์ไปในทางทิศทางเดียวกันกับปริมาณการใช้แก๊สโซฮอล์ E20 อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.20 ซึ่งปัจจัยดังกล่าวถือได้ว่าเป็นสินค้าทดแทนของน้ำมันแก๊สโซฮอล์ E20

และนอกจากนี้ยังมีงานวิจัยที่เปรียบเทียบผลลัพธ์จากการใช้การวิเคราะห์การถดถอยเชิงพหุและโครงข่ายประสาทเทียมในการพยากรณ์ เช่นการประมาณความต้องการใช้พลังงานของประเทศเกาหลีใต้ เพื่อนำไปกำหนดนโยบายด้านพลังงาน โดยใช้โครงข่ายประสาทเทียมแบบ Feed-forward Multilayer และ Back-propagation Algorithm ตัวแบบที่ได้สามารถพยากรณ์ได้ดีกว่าตัวแบบการถดถอยเชิงเส้นและตัวแบบเอ็กโปเนนเชียล เมื่อพิจารณาจากค่า RMSE อีกทั้งยังไม่มีปัญหา Over-fitting (Zong W.G. and William E., 2010) อีกงานวิจัยหนึ่งทำการพยากรณ์ค่าการใช้น้ำมันเชื้อเพลิงของแทรกเตอร์โดยวิธีโครงข่ายประสาทเทียมแบบ Back-propagation และใช้อัลกอริทึมในการเรียนรู้ 6 แบบ อัลกอริทึมการเรียนรู้ที่ให้ประสิทธิภาพสูงสุด คือ Levenberg–Marquardt ซึ่งใช้ชั้นซ่อน 2 ชั้น แต่ละชั้นมี 10 โหนด ผลลัพธ์ของวิธีโครงข่ายประสาทเทียมเมื่อเทียบกับการวิเคราะห์การถดถอยที่เลือกตัวแปรอิสระแบบ Stepwise (Stepwise Regression) พบว่าให้ค่าใกล้เคียงกัน แต่วิธีโครงข่ายประสาทเทียมให้ค่าความแม่นยำในการพยากรณ์สูงกว่า (Fatemeh R.A. and Youset A.G., 2011)

จะเห็นได้จากเหตุผลและงานวิจัยข้างต้น ทำให้ผู้วิจัยสนใจศึกษาปริมาณการใช้น้ำมันสำเร็จรูปในประเทศไทย เนื่องจากข้อมูลปริมาณการใช้น้ำมันสำเร็จรูปมีการแจกแจงแบบเบ้ขวา โดยในงานวิจัยนี้จะศึกษาเปรียบเทียบประสิทธิภาพของการพยากรณ์ โดยตัวแบบเชิงเส้นนัยทั่วไปและตัวแบบเชิงเส้นทั่วไป เพื่อประโยชน์ในการนำผลที่ได้จากการพยากรณ์ไปใช้ในการวางแผนทั้งด้านนโยบายและการจัดหาน้ำมันสำเร็จรูปให้เพียงพอต่อปริมาณการใช้งานในอนาคตภายในประเทศให้เกิดความสมดุลกับพลังงานน้ำมันที่มีอยู่รวมทั้งหาแหล่งพลังงานทดแทนต่อไปได้

1.2 วัตถุประสงค์ของการวิจัย

1.2.1 เพื่อศึกษาเปรียบเทียบประสิทธิภาพของการพยากรณ์เมื่อข้อมูลมีการแจกแจงแบบเบ้ขวา ด้วยวิธีการวิเคราะห์ตัวแบบเชิงเส้นนัยทั่วไป (Generalized Linear Model) กับวิธีการวิเคราะห์ตัวแบบเชิงเส้นทั่วไป (General Linear Model) โดยการแปลงข้อมูล

1.2.2 เพื่อหาสมการที่เหมาะสมในการพยากรณ์ปริมาณการใช้น้ำมันสำเร็จรูปของประเทศไทย

1.2.3 เพื่อหาปัจจัยที่ส่งผลต่อปริมาณการใช้น้ำมันสำเร็จรูปของประเทศไทย

1.3 ขอบเขตของการวิจัย

1.3.1 งานวิจัยนี้เป็นลักษณะการสร้างแบบจำลอง เพื่อหาตัวแบบที่มีประสิทธิภาพในการพยากรณ์ เมื่อข้อมูลมีการแจกแจงแบบเบ้ขวา โดยกำหนดการแจกแจงเป็นการแจกแจงแกมมาที่มีพารามิเตอร์รูปร่าง (Shape Parameter : α) มีค่าเท่ากับ 2 และพารามิเตอร์ขนาด (Scale Parameter : β) มีค่าเท่ากับ 1 และการแจกแจงอินเวอร์สเกาส์เซียน ที่มีพารามิเตอร์แสดงตำแหน่ง (Location Parameter : μ) มีค่าเท่ากับ 1 และพารามิเตอร์รูปร่าง (Shape Parameter : λ) มีค่าเท่ากับ 3 และกำหนดตัวแปรอิสระเป็นการแจกแจงปรกติมาตรฐาน ที่มีพารามิเตอร์แสดงตำแหน่งหรือค่าเฉลี่ย (Location Parameter ; μ) มีค่าเท่ากับ 0 และพารามิเตอร์รูปร่างหรือความแปรปรวน (Shape Parameter : σ^2) มีค่าเท่ากับ 1

1.3.2 นำข้อมูลปริมาณการใช้น้ำมันสำเร็จรูปของประเทศไทย โดยเป็นข้อมูลทุติยภูมิรายเดือน จำนวน 72 เดือน ตั้งแต่ปี พ.ศ. 2553 ถึง พ.ศ. 2558 โดยใช้วิธีการวิเคราะห์ตัวแบบเชิงเส้นนัยทั่วไปกับวิธีการวิเคราะห์ตัวแบบเชิงเส้นทั่วไป เมื่อตัวแปรตามไม่ถูกแปลงค่าและถูกแปลงค่าให้อยู่ในรูป $\frac{1}{y}$, $\frac{1}{y^2}$, $\log(y)$ และแปลงให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ เพื่อสร้างรูปแบบการพยากรณ์ปริมาณการใช้น้ำมันสำเร็จรูปของประเทศไทย ใช้ข้อมูลปริมาณการจำหน่ายน้ำมันสำเร็จรูป จากกรมธุรกิจพลังงาน

1.3.3 หาปัจจัยที่ส่งผลต่อปริมาณการใช้น้ำมันสำเร็จรูปของประเทศไทยซึ่งได้แก่ ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91 ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95 ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20 ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E85 ปริมาณการใช้น้ำมันเบนซิน 91 ปริมาณการใช้น้ำมันเบนซิน 95 ปริมาณการใช้แก๊ส LPG และปริมาณการใช้แก๊ส NGV ใช้ข้อมูลราคาน้ำมันสำเร็จรูปแต่ละชนิดจากสำนักงานนโยบายและแผนพลังงาน โดยมีตัวแปรในการศึกษา ดังนี้

ตัวแปรตาม คือ ปริมาณการใช้น้ำมันสำเร็จรูปแต่ละประเภท ได้แก่

ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91	ปริมาณการใช้น้ำมันเบนซิน 91
ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95	ปริมาณการใช้น้ำมันเบนซิน 95
ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20	ปริมาณการใช้แก๊ส LPG
ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E85	ปริมาณการใช้แก๊ส NGV

ตัวแปรอิสระ ได้แก่

ราคาน้ำมันแก๊สโซฮอล์ 91	ราคาแก๊ส LPG
ราคาน้ำมันแก๊สโซฮอล์ 95	ราคาแก๊ส NGV
ราคาน้ำมันแก๊สโซฮอล์ E20	ราคาเอทานอล
ราคาน้ำมันแก๊สโซฮอล์ E85	ราคาน้ำมันดิบโลก

ราคาน้ำมันเบนซิน 91 อัตราแลกเปลี่ยนเงินตรา

ราคาน้ำมันเบนซิน 95

หมายเหตุ : การพยากรณ์ปริมาณการใช้น้ำมันสำเร็จรูปแต่ละชนิด ปริมาณการใช้ชนิดอื่นจะนำไปเป็นตัวแปรอิสระ

1.4 ประโยชน์ของการวิจัย

1.4.1 ได้ตัวแบบที่เหมาะสมในการแสดงความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระเมื่อข้อมูลมีการแจกแจงแบบเบ้ขวา

1.4.2 ได้สมการที่เหมาะสมในการพยากรณ์ปริมาณการใช้น้ำมันสำเร็จรูปของประเทศไทยเพื่อเป็นแนวทางในการวางแผนเกี่ยวกับการแก้ไขปัญหาการนำเข้าและส่งออกของน้ำมันสำเร็จรูปในภาพรวมของประเทศไทยโดยสามารถนำไปใช้ในการวางแผนด้านภาคอุตสาหกรรม ภาคครัวเรือนและภาคขนส่ง

1.5 นิยามศัพท์

ประสิทธิภาพ หมายถึง ค่าวัดระดับคุณภาพเชิงเปรียบเทียบภายใต้เกณฑ์ที่พิจารณา คือ ค่าเปอร์เซ็นต์ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percent Error : MAPE) และค่ารากของความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error : RMSE)

พหุน ปณฺ ฑิโต ชีเว

บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงเอกสารและงานวิจัยที่เกี่ยวข้อง โดยจะนำเสนอตามลำดับหัวข้อดังนี้

2.1 ตัวแบบเชิงเส้นทั่วไป (General Linear Model : GLM)

2.1.1 ตัวประมาณค่าพารามิเตอร์ของตัวแบบเชิงเส้นทั่วไป

2.1.2 เกณฑ์ที่ใช้พิจารณาความเหมาะสมของตัวแบบเชิงเส้นทั่วไป

2.1.3 สถิติที่ใช้ทดสอบความมีนัยสำคัญของพารามิเตอร์ในตัวแบบเชิงเส้นทั่วไป

2.2 ตัวแบบเชิงเส้นนัยทั่วไป (Generalized Linear Model : GLMs)

2.2.1 ตัวประมาณค่าพารามิเตอร์ของตัวแบบเชิงเส้นนัยทั่วไป

2.2.2 เกณฑ์ที่ใช้พิจารณาความเหมาะสมของตัวแบบเชิงเส้นนัยทั่วไป

2.2.3 สถิติที่ใช้ทดสอบความมีนัยสำคัญของพารามิเตอร์ในตัวแบบเชิงเส้นนัยทั่วไป

2.3 ลักษณะการแจกแจงที่ศึกษา

2.4 ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient)

2.5 การคัดเลือกตัวแปรอิสระด้วยวิธีการถดถอยแบบขั้นบันได

2.6 เกณฑ์ที่ใช้ในการคัดเลือกตัวแบบ

2.7 ข้อมูลน้ำมันสำเร็จรูปในประเทศไทย

2.8 งานวิจัยที่เกี่ยวข้อง

2.1 ตัวแบบเชิงเส้นทั่วไป (General Linear Model : GLM)

ตัวแบบเชิงเส้นทั่วไปหรือบางครั้งเรียกว่าตัวแบบเชิงเส้นทางสถิติทั่วไป (General Linear Statistical Model) เป็นตัวแบบที่พิจารณาค่าของ y จากการทราบค่าของ x ซึ่งมีรูปแบบดังนี้ (สำรวม จงเจริญ, 2551)

$$y = \mu_{y|x} + \varepsilon$$

โดยมีเงื่อนไขว่า Y และ ε ต้องเป็นตัวแปรสุ่ม และ $\mu_{y|x}$ เป็นค่าเฉลี่ยของ y เมื่อกำหนดค่า x ซึ่งเป็นฟังก์ชันของ x กับพารามิเตอร์ในตัวแบบ

X ไม่เป็นตัวแปรสุ่ม ตัวแปรสุ่ม ε มีค่าเฉลี่ยเป็น 0 และมีความแปรปรวนเป็นค่าคงที่ σ^2 กำหนดให้ Y เป็นตัวแปรตามหรือตัวแปรตอบสนอง และ X เป็นตัวแปรอิสระ โดยฟังก์ชันของ x กับพารามิเตอร์ในตัวแบบ $(\underline{\beta})$ นี้จะต้องเป็นฟังก์ชันเชิงเส้นของพารามิเตอร์ที่ไม่ทราบค่า ตัวอย่างของตัวแบบเชิงเส้นทั่วไป ได้แก่

$y = \beta_0 + \beta x + \varepsilon$ เมื่อ $E(\varepsilon) = 0, \quad Var(\varepsilon) = \sigma^2$

$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \varepsilon$ เมื่อ $E(\varepsilon) = 0, \quad Var(\varepsilon) = \sigma^2$

$y = \beta_0 + \beta_1 x + \beta_2 e^x + \varepsilon$ เมื่อ ε มีการแจกแจงเป็น $N(0, \sigma^2)$

เมื่อ β_i แทน พารามิเตอร์ซึ่งไม่ทราบค่า เมื่อ $i = 0, 1, \ldots, p$

ตัวแบบเชิงเส้นทั่วไปสามารถเขียนในรูปของเมทริกซ์ ได้ดังนี้

$$\underline{y} = \underset{\sim}{x}\underline{\beta} + \underline{\varepsilon}$$

เมื่อ $\underline{y}$ คือ เวกเตอร์ของตัวแปรตอบสนอง

$\underset{\sim}{x}$ เป็น เมทริกซ์ของตัวแปรอธิบายผลซึ่งเป็นค่าคงที่และทราบค่า

$\underline{\beta}$ เป็น เวกเตอร์ของพารามิเตอร์ซึ่งไม่ทราบค่า

$\underline{\varepsilon}$ เป็น เวกเตอร์ของความคลาดเคลื่อนสุ่มเป็นค่าที่ไม่สามารถสังเกตได้ที่มีค่าคาดหมายเป็น $\underline{0}$

2.1.1 ตัวประมาณค่าพารามิเตอร์ของตัวแบบเชิงเส้นทั่วไป

การประมาณค่าพารามิเตอร์โดยวิธีกำลังสองน้อยที่สุด (Parameter estimation by Least-square method)

จากตัวแบบ

$$\underline{y} = \underset{\sim}{x}\underline{\beta} + \underline{\varepsilon}$$

เมื่อ $\underline{y}$ คือ เป็นเวกเตอร์ที่มีขนาด $n \times 1$

$\underset{\sim}{x}$ คือ เป็นเมทริกซ์ของค่าคงที่ขนาด $n \times (k+1)$

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

$\underline{\varepsilon}$ คือ เป็นเวกเตอร์ของความคลาดเคลื่อนขนาด $n \times 1$

ในการประมาณค่าพารามิเตอร์ $\underline{\beta}$ โดยวิธีกำลังสองน้อยสุด ต้องการข้อสมมติแต่เพียงว่าเวกเตอร์ของความคลาดเคลื่อน $\underline{\varepsilon}$ มีค่าเฉลี่ยเป็นศูนย์และความแปรปรวนเท่ากับ $\sigma^2 \mathrm{I}$ เท่านั้น ซึ่งส่งผลให้ $\underline{y}$ มีค่าเฉลี่ย $\underset{\sim}{x}\underline{\beta}$ และความแปรปรวนเท่ากับ $\sigma^2 \mathrm{I}$ เช่นกัน และจากแนวคิดของวิธีกำลังสองน้อยที่สุดจะต้องหาตัวประมาณของพารามิเตอร์ $\underline{\beta}$ คือ $\underline{b}$ โดย

$$\underline{\hat{y}} = \underset{\sim}{x}\underline{b} = b_0 + b_1 x_1 + \ldots + b_k x_k$$

ซึ่งจะทำให้กำลังสองของความคลาดเคลื่อนของการประมาณมีค่าน้อยที่สุดนั่นคือ ถ้าให้ $\underline{y} - \underset{\sim}{x}\underline{b}$ เป็นเวกเตอร์ของความคลาดเคลื่อนของการประมาณแล้วกำลังสองของความคลาดเคลื่อนของการประมาณคือ

$$\left(\underline{y}-\underset{\sim}{x}\underline{b}\right)'\left(\underline{y}-\underset{\sim}{x}\underline{b}\right)\equiv Q(\underline{b})$$

การหา $\underline{b}$ ที่ทำให้ $Q(\underline{b})$ มีค่าน้อยที่สุด จะใช้หลักการของอนุพันธ์ โดยการหาค่าอนุพันธ์ของ $Q(\underline{b})$ เทียบกับ $\underline{b}$ แล้วกำหนดให้เท่ากับศูนย์ เพื่อแก้สมการหาค่า $\underline{b}$ นั่นคือ

$$\begin{aligned}Q(b)&=\underline{y}'\underline{y}-\underline{y}'(\underset{\sim}{x}\underline{b})-(\underset{\sim}{x}\underline{b})'\underline{y}+(\underset{\sim}{x}\underline{b})'(\underset{\sim}{x}\underline{b})\\&=\underline{y}'\underline{y}-2\underline{b}'\left(\underset{\sim}{x}'\underline{y}\right)+\underline{b}'\left(\underset{\sim}{x}'\underset{\sim}{x}\right)\underline{b}\end{aligned}$$

เนื่องจากทั้ง $\underline{y}'\left(\underset{\sim}{x}\underline{b}\right)$ และ $(\underset{\sim}{x}\underline{b})'\underline{y}$ เป็นสเกลาร์ เมื่อหาอนุพันธ์เทียบกับ $\underline{b}$ จะได้

$$\frac{\partial Q(\underline{b})}{\partial \underline{b}}=-2\left(\underset{\sim}{x}'\underline{y}\right)+2\left(\underset{\sim}{x}'\underset{\sim}{x}\right)\underline{b}$$

และเมื่อกำหนดให้ $\frac{\partial Q(\underline{b})}{\partial \underline{b}}=0$ แก้สมการหา $\underline{b}$ ได้เป็น

$$\underline{b}=\left(\underset{\sim}{x}'\underset{\sim}{x}\right)^{-1}\underset{\sim}{x}'\underline{y}$$

2.1.2 เกณฑ์ที่ใช้พิจารณาความเหมาะสมของตัวแบบเชิงเส้นทั่วไป

สัมประสิทธิ์การตัดสินใจเชิงพหุ (Coefficient of Multiple Determination : r^2) ความหมายของสัมประสิทธิ์การตัดสินใจเชิงพหุนั้นจะเหมือนความหมายของสัมประสิทธิ์การตัดสินใจในกรณีที่มีตัวแปรอิสระเพียง 1 ตัว นั่นคือ ตัวแปรอิสระ $X_1,\ldots,X_k$ สามารถอธิบายการเปลี่ยนแปลงของตัวแปร Y ได้มากน้อยเพียงใด โดยใช้สัญลักษณ์ R^2 หรือ r^2 แทนสัมประสิทธิ์การตัดสินใจเชิงพหุ (บังอร กุมพล, 2544)

$$r^2=\frac{SSR}{SST}=\frac{\sum\left(\hat{Y}-\bar{Y}\right)^2}{\sum\left(Y-\bar{Y}\right)^2}=1-\frac{\sum\left(\hat{Y}-\bar{Y}\right)^2}{\sum\left(Y-\bar{Y}\right)^2}=1-\frac{SSE}{SST}$$

โดยที่ $0\le r^2\le 1$ นั่นคือ

ถ้า r^2 มีค่าเข้าใกล้ 1 จะหมายถึง $X_1,\ldots,X_k$ สามารถอธิบายการเปลี่ยนแปลงของ Y ได้มาก แต่ถ้า r^2 มีค่าเข้าใกล้ 0 จะหมายถึง $X_1,\ldots,X_k$ สามารถอธิบายการเปลี่ยนแปลงของ Y ได้ค่อนข้างน้อยมาก

2.1.3 สถิติที่ใช้ทดสอบความมีนัยสำคัญของพารามิเตอร์ในตัวแบบเชิงเส้นทั่วไป

2.1.3.1 ทดสอบโดยใช้ F – test

สมมติฐานการทดสอบ

$H_0:\beta_1=\beta_2=\ldots=\beta_k=0$

H_1 :มี $\beta_i\neq 0$ อย่างน้อย 1 ค่า ; $i=1,2,\ldots,k$

สถิติในการทดสอบ

$$F_0 = \frac{MSR}{MSE}$$

สรุปผลเมื่อจะปฏิเสธ H_0 ถ้า $F_0 \geq F_{\alpha,(1,n-2)}$ นั่นคือ มีตัวแปรอิสระอย่างน้อย 1 ตัว มีอิทธิพลต่อตัวแปรตาม

2.4.2 ทดสอบโดยใช้ T – Test

สมมติฐานการทดสอบ

$H_0 : \beta_i = 0; i = 1,2,...,k$

$H_1 : \beta_i \neq 0$

สถิติในการทดสอบ

$$T = \frac{\hat{\beta}_i}{S.E(\hat{\beta}_i)}$$

สรุปผล เขตปฏิเสธ H_0 : จะปฏิเสธ H_0 ถ้า $T \geq t_{\frac{\alpha}{2}}$ หรือ $T \leq -t_{\frac{\alpha}{2}}$ นั่นคือมีตัวแปรอิสระตัวที่ i มีอิทธิพลต่อตัวแปรตาม

2.2 ตัวแบบเชิงเส้นนัยทั่วไป (Generalized Linear Model : GLMs)

ตัวแบบเชิงเส้นนัยทั่วไป (Generalized Linear Model) พัฒนามาจากตัวแบบเชิงเส้นหรือตัวแบบเชิงเส้นทั่วไปเป็นการอธิบายความสัมพันธ์ระหว่างตัวแปรตาม (Response Variable : Y) และตัวแปรอิสระ (Explanatory Variables : X) หรือตัวแปรพยากรณ์ (Predictor Variables) โดยที่ตัวแบบสามารถเขียนให้อยู่ในรูป

$$y = g(\mu_{y|x}) + \varepsilon$$

โดยที่ $g(\mu_{y|x})$ เป็นฟังก์ชันของค่าเฉลี่ยของ y เมื่อกำหนด x และเป็นฟังก์ชันทางเดียวที่สามารถหาอนุพันธ์ได้ (สุจิตตา สุระภี, 2560)

การวิเคราะห์ด้วย GLM พบว่ามีข้อจำกัดในการวิเคราะห์และการนำไปประยุกต์ใช้ดังนี้

1. ตัวแปรสุ่ม y ต้องเป็นตัวแปรสุ่มต่อเนื่อง (Continuous Random Variable)
2. ค่าความคลาดเคลื่อนต้องมีการแจกแจงปรกติ (Normal Distribution)
3. ความแปรปรวนของค่าความคลาดเคลื่อนต้องมีค่าคงที่ (Constant Variance)

ซึ่งถ้าความคลาดเคลื่อนมีการแจกแจงปรกติและมีความแปรปรวนคงที่จะสามารถพิสูจน์ได้ว่าตัวแปรตาม (Y) ก็จะมีการแจกแจงปรกติและมีค่าความแปรปรวนคงที่เช่นเดียวกัน

ลักษณะของข้อมูลส่วนใหญ่มักจะไม่เป็นไปตามข้อจำกัดในการวิเคราะห์ของตัวแบบเชิงเส้นทั่วไป รวมทั้งข้อมูลปริมาณการใช้น้ำมันสำเร็จรูปของประเทศไทยไม่ได้มีคุณสมบัติตามข้อกำหนดที่กล่าวไว้ข้างต้นเสมอไป ด้วยสาเหตุดังกล่าวจึงได้ใช้ตัวแบบเชิงเส้นนัยทั่วไปในการพิจารณาความสัมพันธ์ของตัวแปรเพื่อให้สามารถรองรับกับข้อมูลที่มีค่าต่อเนื่องแต่ไม่ได้มีการแจกแจงปรกติตลอดจนข้อมูลที่มีค่าความแปรปรวนไม่คงที่ และที่สำคัญคือในวิธีการวิเคราะห์ตัวแบบเชิงเส้นนัยทั่วไปจะมีทางเลือกให้เลือกหลากหลายในส่วนประกอบเชิงสุ่มคือส่วนประกอบเชิงสุ่มไม่จำเป็นต้องมีการแจกแจงปรกติและฟังก์ชันเชื่อมโยงมีให้เลือกหลายรูปแบบและเป็นอิสระกับส่วนประกอบเชิงสุ่ม

ด้วยสาเหตุที่ตัวแบบเชิงเส้นนัยทั่วไปมีรากฐานมาจากตัวแบบเชิงเส้น โดยตัวแบบเชิงเส้นนัยทั่วไป y ต้องมีฟังก์ชันการแจกแจงความน่าจะเป็น (Probability Function : $f(y)$) อยู่ในวงศ์เลขชี้กำลัง (Exponential Family Form)

ตัวแบบเชิงเส้นนัยทั่วไปเป็นตัวแบบที่ถูกแนะนำครั้งแรกโดย Nelder และ Nedderburn ในปี ค.ศ.1972 โดยแยกส่วนประกอบต่างๆ ของตัวแบบเชิงเส้นนัยทั่วไปออกเพื่อให้ง่ายต่อการเข้าใจ

โดยส่วนประกอบของตัวแบบเชิงเส้นนัยทั่วไปจะแบ่งออกเป็น 3 ส่วนดังต่อไปนี้ (Agresti A., 2009)

1. ส่วนประกอบเชิงสุ่ม (Random Component)

หมายถึงตัวแปรตาม (Response Variable : Y) และการแจกแจงความน่าจะเป็นของตัวแปรตามโดยกำหนดให้ $Y_1, Y_2, \ldots, Y_N$ ต้องเป็นอิสระกัน โดยเงื่อนไขที่สำคัญของการแจกแจงของส่วนประกอบเชิงสุ่ม จะต้องมีฟังก์ชันการแจกแจงความน่าจะเป็น (Probability Function) อยู่ในวงศ์เลขชี้กำลัง (Exponential Family) เช่น การแจกแจงเกาส์ (Gaussian Distribution) การแจกแจงทวินาม (Binomial Distribution) การแจกแจงปัวซง (Poisson Distribution) การแจกแจงแกมมา (Gamma Distribution) และการแจกอินเวอร์สเกาส์เซียน (Inverse – Gaussian Distribution) เป็นต้น โดยฟังก์ชันความน่าจะเป็นที่อยู่ในวงศ์เลขชี้กำลังคือ ฟังก์ชันความน่าจะเป็นที่สามารถเขียนอยู่ในรูป

$$f(y \mid \theta, \phi) = \exp\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right]$$

เมื่อ θ คือ คานอนิคัลพารามิเตอร์ (canonical parameter)

ϕ คือ พารามิเตอร์การกระจาย (dispersion parameter)

$b(\theta)$ คือ เป็นฟังก์ชันของพารามิเตอร์ที่ไม่ขึ้นกับตัวแปรสุ่ม y

$c(y, \phi)$ คือ เป็นฟังก์ชันของตัวแปรสุ่มกับ dispersion parameter

ตัวอย่างเช่น

การแจกแจงปรกติ (Normal Distribution) หรือการแจกแจงเกาส์ (Gaussian Distribution)

$$f(y\,|\,\theta,\phi)=\frac{1}{\sqrt{2\pi\sigma}}\exp\left[-\frac{(y-\mu)^2}{2\sigma}\right]$$

$$=\exp\left[\frac{y\mu-\mu^2/2}{\sigma^2}-\frac{1}{2}\left(\frac{y^2}{\sigma^2}+\log(2\pi\sigma^2)\right)\right]$$

ดังนั้น $\theta=\mu$, $\phi=\sigma^2$, $a(\phi)=\phi$, $b(\theta)=\frac{\theta^2}{2}$ และ $c(y,\phi)=-\frac{\left(\frac{y^2}{\phi}+\log(2\pi\phi)\right)}{2}$

การแจกแจงแกมมา (Gamma Distribution)

$$f(y;\alpha,\beta)=\frac{y^{\alpha-1}}{\beta^{\alpha}\Gamma(\alpha)}\exp\left(\frac{-y}{\beta}\right)$$

$$=\exp\left(\frac{-y\left(\frac{1}{\beta}\right)-\log(\beta)}{\frac{1}{\alpha}}+(\alpha-1)\log(y)+\alpha\log(\alpha)-\log(\Gamma(\alpha)\right)$$

ดังนั้น $\theta=-\frac{1}{\beta}$, $\phi=\frac{1}{\alpha}$, $a(\phi)=\phi=1$, $b(\theta)=\log(\beta)=-\log(-\theta)$ และ

$$c(y,\phi)=\left(\frac{1}{\phi}-1\right)\log(y)-\frac{\log(\phi)}{\phi}-\log(\Gamma(\frac{1}{\phi}))$$

การแจกอินเวอร์สเกาส์เซียน (Inverse – Gaussian Distribution)

$$f(y;\mu,\lambda)=\sqrt{\frac{\lambda}{2\pi y^3}}\exp\left(\frac{-\lambda(y-\mu)^2}{2\mu^2 y}\right)$$

$$=\exp\left(\frac{y(-\frac{1}{\mu^2})+\frac{2}{\mu}}{\frac{2}{\lambda}}-\frac{\lambda}{2y}+\frac{\log(\lambda)-\log(2\pi y^3)}{2}\right)$$

ดังนั้น $\theta=-\frac{1}{\mu^2}$, $\phi=\frac{2}{\lambda}$, $a(\phi)=\phi=\frac{2}{\lambda}$, $b(\theta)=-\frac{2}{\mu}=-2\sqrt{-\theta}$ และ

$$c(y,\phi)=-\frac{\log(\frac{\phi}{2})+\log(2\pi y^3)}{2}-\frac{1}{\phi y}$$

พหุน ปณฑิโต ชีเว

2. ส่วนประกอบเชิงระบบ (Systematic Component)

หมายถึงตัวแปรอิสระ (Explanatory Variables : X) ทุกตัวซึ่งเป็นตัวแปรที่ถูกใช้ในการพยากรณ์และถูกกำหนดให้อยู่ในรูปการรวมตัวกันเชิงเส้นของพารามิเตอร์ (Linear Combination of Parameter) นั่นคือจะจัดอยู่ในรูป

$$\alpha + \beta_1 x_1 + \cdots + \beta_k x_k$$

ถ้าส่วนประกอบเชิงระบบอยู่ในรูปการรวมตัวกันเชิงเส้นในลักษณะนี้เราจะเรียกอีกอย่างว่าตัวพยากรณ์เชิงเส้น (Linear Predictor)

3. ฟังก์ชันเชื่อมโยง (Link Function)

เป็นฟังก์ชันเชื่อมโยงระหว่างส่วนประกอบเชิงสุ่มและส่วนประกอบเชิงระบบ นั่นหมายถึงเรากำลังพิจารณาว่า $\mu = E(y \mid x)$ จะมีความเชื่อมโยงอย่างไรกับตัวแปรอิสระที่อยู่ในรูปตัวพยากรณ์เชิงเส้น โดยที่

$$g(\mu) = \alpha + \beta_1 x_1 + \cdots + \beta_k x_k$$

ฟังก์ชัน $g(.)$ จะถูกเรียกว่าฟังก์ชันเชื่อมโยง (Link Function) ซึ่งต้องเป็นฟังก์ชันทางเดียวและสามารถหาอนุพันธ์ได้ (Monotonic and differentiable function) โดยฟังก์ชันเชื่อมโยงที่ง่ายที่สุดก็คือ $g(\mu) = \mu$ เราจะเรียกฟังก์ชันเชื่อมโยงนี้ว่า ฟังก์ชันเชื่อมโยงเอกลักษณ์ (Identity Link) ดังนั้นตัวแบบเชิงเส้นสำหรับค่าเฉลี่ยของตัวแปรตาม ที่มีฟังก์ชันเชื่อมโยงเป็นฟังก์ชันเชื่อมโยงเอกลักษณ์จะเขียนอยู่ในรูป

$$\mu = \alpha + \beta_1 x_1 + \cdots + \beta_k x_k$$

ซึ่งถ้า Y มีการแจกแจงปรกติตัวแบบนี้จะเป็นรูปแบบของตัวแบบเชิงเส้นทั่วไป สำหรับฟังก์ชันเชื่อมโยงอื่น คือฟังก์ชันที่เปลี่ยนตัวแบบที่อยู่ในรูปค่าเฉลี่ยให้อยู่ในรูปที่ไม่เป็นเชิงเส้นของค่าเฉลี่ย เพื่อใช้เชื่อมโยงกับตัวพยากรณ์เชิงเส้น ยกตัวอย่างเช่น กำหนดให้ฟังก์ชันเชื่อมโยงเป็น $g(\mu) = \log(\mu)$ ถ้า μ มีค่าเป็นบวกเราจะเรียกฟังก์ชันเชื่อมโยงนี้ว่า “log link” ดังนั้นตัวแบบเชิงเส้นนัยทั่วไปใช้ “log link” เป็นฟังก์ชันเชื่อมโยง จะถูกเรียกว่าตัวแบบล็อกเชิงเส้น (loglinear model) จะอยู่ในรูป

$$\log(\mu) = \alpha + \beta_1 x_1 + \cdots + \beta_k x_k$$

แต่ถ้าฟังก์ชันเชื่อมโยงเป็น $g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$ ซึ่งก็คือ log ของ odds เราจะเรียกว่า “logit link” ดังนั้นตัวแบบที่มีฟังก์ชันเชื่อมโยงเป็น logit link จะถูกเรียกว่าตัวแบบโลจิท (logit model) จะอยู่ในรูป

$$\log\left(\frac{\mu}{1-\mu}\right) = \alpha + \beta_1 x_1 + \cdots + \beta_k x_k$$

สำหรับงานวิจัยในครั้งนี้ได้ศึกษากรณีที่ Y มีการแจกแจงแกมมา (Gamma Distribution) และมีการแจกแจงอินเวอร์สเกาส์เซียน (Inverse - Gaussian Distribution)

ในกรณีที่ Y มีการแจกแจงแบบแกมมา (Gamma Distribution) จะมีฟังก์ชันเชื่อมโยงคานอนิคัล (canonical link function) เป็น Inverse ซึ่งมีรูปแบบของตัวแบบดังนี้

$$\mu^{-1} = \alpha + \beta_1 x_1 + \cdots + \beta_k x_k$$

ในกรณีที่ Y มีการแจกแจงอินเวอร์สเกาส์เซียน (Inverse - Gaussian Distribution) จะมีฟังก์ชันเชื่อมโยงคานอนิคัล (canonical link function) เป็น Inverse - square ซึ่งมีรูปแบบของตัวแบบดังนี้

$$\mu^{-2} = \alpha + \beta_1 x_1 + \cdots + \beta_k x_k$$

ตาราง 2.1 ชนิดของตัวแบบสำหรับการวิเคราะห์ทางสถิติ

องค์ประกอบตัวแปรสุ่ม (random component)	ฟังก์ชันเชื่อมโยงคานอนิคัล (canonical link function)	องค์ประกอบเชิงระบบ (systematic component)	ตัวแบบ (model)
Normal	Identity	ตัวแปรต่อเนื่อง	Regression
Normal	Identity	ตัวแปรกลุ่ม	Analysis of Variance
Normal	Identity	ผสม	Analysis of Covariance
Poisson	Log	ผสม	Log Linear
Bernoulli	Logit	ผสม	Logistic Regression
Multinomial	Generalized logit	ผสม	Multinomial Response
Gamma	Inverse	ผสม	Gamma GLMs
Inverse Gaussian	Inverse-square	ผสม	Inverse - Gaussian GLMs

ที่มา : (Faraway J., 2016)

ในการวิเคราะห์ตัวแบบเชิงเส้นนัยทั่วไปไม่ได้มีข้อจำกัดว่าฟังก์ชันเชื่อมโยงต้องเป็นฟังก์ชันเชื่อมโยงคานอนิคัลเท่านั้นสามารถเป็นฟังก์ชันเชื่อมโยงใดก็ได้

ตัวแบบเชิงเส้นนัยทั่วไปสำหรับข้อมูลที่มีการแจกแจงแบบปรกติ (Normal GLMs)

ตัวแบบการถดถอยสามัญ (Ordinary Regression Model) และตัวแบบการวิเคราะห์ความแปรปรวน (ANOVA Model) ซึ่งตัวแปรตามเป็นข้อมูลเชิงปริมาณแบบต่อเนื่องเราจะถือว่าเป็นกรณีหนึ่งของตัวแบบการถดถอยเชิงเส้นนัยทั่วไป (GLMs) ที่องค์ประกอบเชิงสุ่ม (Random

Component) มีการแจกแจงปรกติและมีฟังก์ชันเชื่อมโยง (Link Function) เป็นฟังก์ชันเชื่อมโยงแบบเอกลักษณ์ (Identity Link) โดย $g(\mu)=\mu$ โดยตัวแบบการถดถอยสามัญ คือ

$$\mu_{y|x}=\alpha+\beta_1 x_1+\ldots+\beta_k x_k$$

โดยปรกติในการวิเคราะห์การถดถอยเราจะมีข้อตกลงเบื้องต้นที่สำคัญ 2 ข้อ คือตัวแปรตาม Y ต้องมีการแจกแจงปรกติที่มีค่าความแปรปรวนคงที่ ถ้าข้อมูลไม่เป็นไปตามข้อตกลงนี้จะต้องทำการแปลงข้อมูล (Transform) ให้อยู่ในรูปแบบต่างๆ และถึงแม้ว่าหลังจากแปลงข้อมูลแล้วรูปแบบที่แปลงค่าของ Y จะมีการแจกแจงปรกติ แต่ก็เป็นไปได้ยากที่ค่าความแปรปรวนของค่าที่เราแปลงค่าไปจะมีค่าคงที่

ดังนั้นการใช้ทฤษฎีหรือวิธีการของตัวแบบเชิงเส้นนัยทั่วไป (GLMs) จะช่วยแก้ปัญหานี้ได้โดยไม่จำเป็นต้องแปลงข้อมูล ในการประมาณค่าพารามิเตอร์ในตัวแบบ จะใช้วิธีภาวะน่าจะเป็นสูงสุด ส่วนการวิเคราะห์การถดถอยสามัญ (Ordinary Regression) จะใช้วิธีกำลังสองน้อยสุด (Least Squared Method) และที่สำคัญคือในวิธีการวิเคราะห์ตัวแบบเชิงเส้นนัยทั่วไปจะมีทางเลือกให้เลือกหลากหลายในส่วนประกอบเชิงสุ่ม คือส่วนประกอบเชิงสุ่มไม่จำเป็นต้องมีการแจกแจงปรกติและฟังก์ชันเชื่อมโยงมีให้เลือกหลายรูปแบบและเป็นอิสระกับส่วนประกอบเชิงสุ่ม

2.2.1 ตัวประมาณค่าพารามิเตอร์ของตัวแบบเชิงเส้นนัยทั่วไป

การประมาณค่าพารามิเตอร์ในตัวแบบเชิงเส้นนัยทั่วไปจะใช้วิธีภาวะน่าจะเป็นสูงสุด (Maximum Likrlihood Estimates : MLE) โดยที่ฟังก์ชันล็อกภาวะน่าจะเป็นของแต่ละค่าสังเกตจะเขียนอยู่ในรูป (Faraway J., 2016)

$$\ell(\beta;y_i)=w_i\left[\frac{y_i\theta_i-b(\theta_i)}{\phi}\right]+c(y_i,\phi)$$

เมื่อ w_i คือ ค่าถ่วงน้ำหนัก

β คือ พารามิเตอร์ในตัวแบบ

$b(\theta_i)$ คือ ฟังก์ชันของพารามิเตอร์

$y_i\theta_i$ คือ ฟังก์ชันของตัวแปรเชิงสุ่มกับพารามิเตอร์

ϕ คือ dispersion parameter

$c(y_i,\phi)$ คือ ฟังก์ชันของตัวแปรเชิงสุ่มกับ dispersion parameter

เราต้องการหาค่า $\underset{\sim}{\beta}$ ที่ทำให้ $\ell(\beta; y_i)$ มีค่าสูงที่สุด โดยการหาอนุพันธ์ของ ℓ เทียบกับ β_j ดังนี้

$$\frac{\partial \ell}{\partial \beta_j} = \frac{1}{\phi}\sum_i w_i \left(y_i \frac{\partial \theta_i}{\partial \beta_j} - b'(\theta_i)\frac{\partial \theta_i}{\partial \beta_j} \right)$$

โดยกฎลูกโซ่ จะได้

$$\frac{\partial \theta_i}{\partial \beta_j} = \frac{\partial \theta_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial \beta_j}$$

เมื่อ $\frac{\partial \mu_i}{\partial \theta_i} = b''(\theta_i)$ จะได้

$$\frac{\partial \ell}{\partial \beta_j} = \frac{1}{\phi}\sum_i \frac{\left(y_i - b'(\theta_i)\right)}{b''(\theta_i)/w_i}\frac{\partial \mu_i}{\partial \beta_j}$$

และสามารถเขียนให้อยู่ในรูปของค่าเฉลี่ยและค่าความแปรปรวน และกำหนดให้อนุพันธ์ย่อยมีค่าเป็น 0 เพื่อให้ได้ค่าประมาณด้วยวิธีภาวะน่าจะเป็นสูงสุด ได้ดังนี้

$$\sum_i \frac{(y_i - \mu_i)}{V(\mu_i)}\frac{\partial \mu_i}{\partial \beta_j} = 0 \quad ; \quad \forall_j$$

ถ้าเราทราบฟังก์ชันความแปรปรวน $V(\mu)$ แล้วเราจะได้ค่าประมาณพารามิเตอร์ที่ทำให้ $\sum_i \frac{(y_i - \mu_i)^2}{V(\mu_i)}$ มีค่าน้อยที่สุด

การหาค่าประมาณ $\sum_i \frac{(y_i - \mu_i)}{V(\mu_i)}\frac{\partial \mu_i}{\partial \beta_j} = 0 \quad ; \quad \forall_j$ จะใช้ algorithm เดียวกันกับการใช้ iteratively reweighted least squares (IRWLS) โดยมีขั้นตอนดังนี้

1. ใส่ค่าเริ่มต้น $\hat{\mu}^0$ และ $\hat{\eta}^0$
2. สร้าง adjusted dependent variable $z^i = \hat{\eta}^i + \left(y - \hat{\mu}^i\right)\frac{d\eta}{d\mu}|\hat{\eta}^i$
3. สร้าง weights $1/w^0 = \left(\frac{d\eta}{d\mu}\right)^2 |\hat{\eta}^i V(\hat{\mu}^0)$
4. ประมาณค่า $\hat{\beta}^{i+1}$ และ $\hat{\eta}^{i+1}$

ทำซ้ำตามขั้นตอนที่ 2 – 4 จนกระทั่งลู่เข้า

สังเกตว่าขั้นตอนการหาตัวแบบที่เหมาะสมจะพิจารณา $\eta = g(\mu)$ และ $V(\mu)$ แต่ไม่ได้สนใจเกี่ยวกับการแจกแจงของ y ประเด็นนี้จะมีความสำคัญเมื่อพิจารณาวิธี quasi – likelihood

ค่าประมาณของความแปรปรวนจะได้จากทฤษฎีภาวะน่าจะเป็นโดย

$$\hat{\text{var}}(\hat{\beta}) = (X^T W X)^{-1}\hat{\phi}$$

เมื่อ W คือเมทริกซ์แนวทแยงที่ได้จากค่าถ่วงน้ำหนักของ w นี้เป็นรูปแบบเดียวกับการใช้วิธีถ่วงน้ำหนักแบบกำลังน้อยที่สุด ยกเว้นแต่ว่าค่าถ่วงน้ำหนักจะเป็นฟังก์ชันของ y ใน GLMs

2.2.2 เกณฑ์ที่ใช้พิจารณาความเหมาะสมของตัวแบบเชิงเส้นนัยทั่วไป

ค่าดิเวียนส์ส่วนเหลือ (Residual Deviance)

ตัวสถิติดิเวียนส์

$$D = -2\ln(\frac{l_0}{l_1})$$
$$= \sum_{j=1}^{J} d(y_j, \hat{\mu}_j)^2$$

เมื่อ J คือ จำนวน covariate pattern

$d(y_j, \hat{\mu}_j)$ คือ ค่าดิเวียนส์ส่วนเหลือ

สำหรับ y ที่มีการแจกแจงแกมมาจะได้ $D = 2\sum_i \left[-\log(\frac{y_i}{\hat{\mu}_i}) + \frac{(y_i - \hat{\mu}_i)}{\hat{\mu}_i} \right]$ และ y ที่มีการแจกแจงอินเวอร์สเกาส์เซียน $D = \sum_i \frac{(y_i - \hat{\mu}_i)^2}{(\hat{\mu}_i^2 y_i)}$

สรุป เมื่อตัวแบบที่มีค่าดิเวียนส์ส่วนเหลือต่ำจะเป็นตัวแบบที่มีประสิทธิภาพในการพยากรณ์ที่มีความแม่นยำสูง

2.2.3 สถิติที่ใช้ทดสอบความมีนัยสำคัญของพารามิเตอร์ในตัวแบบเชิงเส้นนัยทั่วไป

2.2.3.1 Likelihood Ratio Test

สมมติฐาน คือ

$H_0 : \beta_1 = \beta_2 = ... = \beta_k = 0$

H_1 :มี $\beta_i \neq 0$ อย่างน้อย 1 ค่า ; $i = 1,2,...,k$

สถิติในการทดสอบ

คำนวณสถิติทดสอบจากสูตร $D = -2\ln(\frac{l_0}{l_1})$

l_0 คือ ค่าสูงสุดของฟังก์ชันภาวะน่าจะเป็นตามตัวแบบของ H_0

l_1 คือ ค่าสูงสุดของฟังก์ชันภาวะน่าจะเป็นตามตัวแบบของ H_1

ค่าสถิติ D ที่คำนวณได้จะมีการแจกแจงแบบ χ^2 และมี $df = df_2 - df_1$ โดย df_1 และ df_2 คือ จำนวนพารามิเตอร์ใน null model และ alternative model ตามลำดับ การทดสอบสมมติฐานจึงต้องมีการเปรียบเทียบค่า D กับค่า χ^2 วิกฤต ที่ระดับความเชื่อมั่นต่างๆ หากค่า

D มากกว่า χ^2 วิกฤต เราจะปฏิเสธสมมติฐานหลัก เมื่อมีตัวแปรอิสระอย่างน้อย 1 ตัว มีอิทธิพลต่อตัวแปรตาม ซึ่งตัวแปรตามคือ ปริมาณการใช้น้ำมันสำเร็จรูปแต่ละชนิด ที่ระดับนัยสำคัญ 0.05

2.2.3.2 สถิติทดสอบของวาลด์ (Wald Statistic)

สมมติฐาน คือ

$H_0 : \beta_i = 0; i = 1,2,...,k$

$H_1 : \beta_i \neq 0$

สถิติในการทดสอบ

$$\text{Wald Statistic} = \left(\frac{\hat{\beta}_i}{S.E(\hat{\beta}_i)} \right)^2$$

ในการทดสอบถ้าผลการทดสอบมีหลักฐานพอที่จะไม่ปฏิเสธ H_0 (ยอมรับ) แสดงว่าตัวแปรทำนายไม่มีผลต่อการเปลี่ยนแปลงของตัวแปรตามผ่านฟังก์ชันเชื่อมโยง ถ้าทดสอบพบว่า ปฏิเสธ H_0 หรือข้อสรุปจะอยู่ใน H_1 และค่าสัมประสิทธิ์เป็นบวก แสดงว่าตัวแปรทำนายนั้นมีผลต่อตัวแปรตามในทางเดียวกันผ่านฟังก์ชันเชื่อมโยงและถ้าหากค่าสัมประสิทธิ์เป็นลบ แสดงว่าตัวแปรทำนายนั้นส่งผลต่อตัวแปรตามในทางตรงกันข้ามผ่านฟังก์ชันเชื่อมโยง สถิติทดสอบของวาลด์ (Wald Statistic) จะมีการแจกแจงแบบ χ^2 และมี $df = 1$ ค่าวิกฤตคือ $\chi^2_{\alpha,1}$ จะปฏิเสธสมมติฐานหลัก เมื่อ Wald Statistic มากว่าค่าวิกฤต $T \leq -t_{\frac{\alpha}{2}}$ นั่นคือ มีตัวแปรอิสระตัวที่ i มีอิทธิพลต่อตัวแปรตาม

2.3 ลักษณะการแจกแจงที่ศึกษา

2.3.1 การแจกแจงแกมมา (Gamma Distribution)

การแจกแจงแกมมามักจะใช้ในการบรรยายตัวแปรสุ่มซึ่งเป็นผลจากการวัดใดๆ ที่มีค่าเป็นลบไม่ได้ เช่นปริมาณน้ำฝนที่ตกใน 1 วัน (ถ้ามีฝนตก) อายุการใช้งานของผลิตภัณฑ์ เป็นต้น

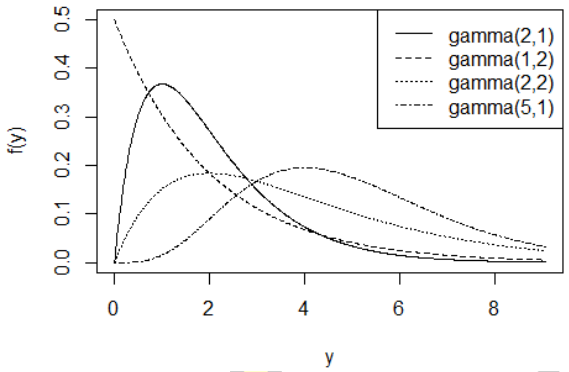
ถ้า Y เป็นตัวแปรสุ่มต่อเนื่อง มีการแจกแจงแกมมา โดยมีพารามิเตอร์รูปร่าง $\alpha > 0$ และพารามิเตอร์ขนาด $\beta > 0$ มีฟังก์ชันความหนาแน่นความน่าจะเป็น (Probability Density Function : pdf) ดังนี้ (ธีระศักดิ์ อุรัจนานนท์, 2546)

$$f(y;\alpha,\beta) = \frac{y^{\alpha-1}}{\beta^{\alpha}\Gamma(\alpha)} \exp\left(\frac{-y}{\beta}\right) \quad ;0 \leq y < \infty$$

ค่าคาดหมายและความแปรปรวน คือ

ค่าคาดหมาย เท่ากับ $\alpha\beta$

ความแปรปรวน เท่ากับ $\alpha\beta^2$



รูปภาพที่ 2.1 กราฟแสดงรูปร่างของฟังก์ชันความน่าจะเป็นของการแจกแจงแกมมา ที่มีพารามิเตอร์ (α, β) เปลี่ยนแปลงไป

2.3.2 การแจกแจงอินเวอร์สเกาส์เซียน (Inverse Gaussian Distribution)

เสนอขึ้นครั้งแรกโดย Schrodinger ในปี ค.ศ. 1915 ซึ่งเป็นการแจกแจงหนึ่งที่นำมาประยุกต์ใช้กับปัญหาต่างๆ เช่น การวิเคราะห์ราคาหุ้น ปัญหาด้านชีววิทยา อุทกภูมิศาสตร์ อุตุนิยมวิทยา เป็นต้น ที่มีพารามิเตอร์แสดงตำแหน่ง $\mu > 0$ และพารามิเตอร์รูปร่าง $\lambda > 0$ (วรางคณา กีรติวิบูลย์, 2556)

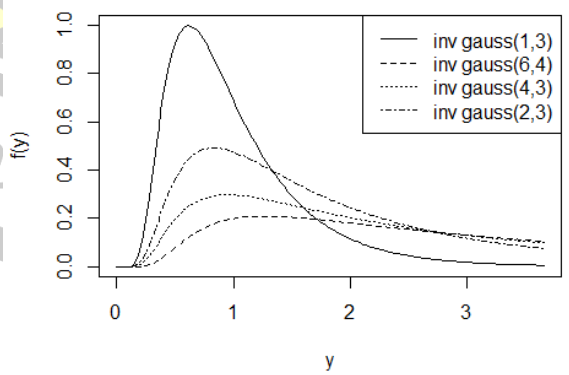
ฟังก์ชันความหนาแน่นความน่าจะเป็น (Probability Density Function : pdf) ของการแจกแจงอินเวอร์สเกาส์เซียน เขียนได้ดังนี้

$$f(y;\mu,\lambda) = \sqrt{\frac{\lambda}{2\pi y^3}} \exp\left(\frac{-\lambda(y-\mu)^2}{2\mu^2 y}\right) \qquad \text{เมื่อ} \quad y > 0, \mu, \lambda > 0$$

ค่าคาดหมายและความแปรปรวน คือ

ค่าคาดหมาย เท่ากับ μ

ความแปรปรวน เท่ากับ $\frac{\mu^3}{\lambda}$



รูปภาพที่ 2.2 กราฟแสดงรูปร่างของฟังก์ชันความน่าจะเป็นของการแจกแจงอินเวอร์สเกาส์เซียน ที่มีพารามิเตอร์ (μ, λ) เปลี่ยนแปลงไป

2.4 ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation coefficient)

โดยทั่วไปนิยมใช้สัญญาลักษณ์ r แทนสัมประสิทธิ์สหสัมพันธ์ของกลุ่มตัวอย่าง และ ρ แทนสัมประสิทธิ์สหสัมพันธ์ของประชากร ค่าสัมประสิทธิ์สหสัมพันธ์ที่ใช้วัดขนาดของความสัมพันธ์กันระหว่างตัวแปร มี 2 ลักษณะ คือ $-1 \le r \le 1$ และ $0 \le r \le 1$

การบอกระดับหรือขนาดของความสัมพันธ์จะใช้ตัวเลขของค่าสัมประสิทธ์สหสัมพันธ์ หากค่าสัมประสิทธ์สหสัมพันธ์มีค่าเข้าใกล้ -1 หรือ 1 แสดงถึงการมีความสัมพันธ์กันในระดับสูง แต่หากมีค่าเข้าใกล้ 0 แสดงถึงการมีความสัมพันธ์กันในระดับน้อยหรือไม่มีเลย สำหรับการพิจารณาค่าสัมประสิทธ์สหสัมพันธ์โดยทั่วไปอาจใช้เกณฑ์ดังนี้ (Hinkle, D.E, William, 1998)

ค่า r	ระดับของความสัมพันธ์
0.90 - 1.00	มีความสัมพันธ์กันสูงมาก
0.70 - 0.90	มีความสัมพันธ์กันในระดับสูง
0.50 - 0.70	มีความสัมพันธ์กันในระดับปานกลาง
0.30 - 0.50	มีความสัมพันธ์กันในระดับต่ำ
0.00 - 0.30	มีความสัมพันธ์กันในระดับต่ำมาก

เครื่องหมาย +,- หน้าตัวเลขสัมประสิทธิ์สหสัมพันธ์จะบอกถึงทิศทางของความสัมพันธ์

r มีเครื่องหมาย + หมายถึง การมีความสัมพันธ์กันไปในทิศทางเดียวกัน (ตัวแปรหนึ่งมีค่าสูง อีกตัวหนึ่งจะมีค่าสูงไปด้วย)

r มีเครื่องหมาย - หมายถึง การมีความสัมพันธ์กันไปในทิศทางตรงกันข้าม (ตัวแปรหนึ่งมีค่าสูง ตัวแปรอีกตัวหนึ่งจะมีค่าต่ำ)

ยกเว้นค่าสัมประสิทธิ์สหสัมพันธ์บางชนิดที่มีลักษณะ $0 \le r \le 1$ ซึ่งจะบอกได้เพียงขนาดหรือระดับของความสัมพันธ์เท่านั้น ไม่สามารถบอกทิศทางของความสัมพันธ์ได้

2.5 การคัดเลือกตัวแปรอิสระด้วยวิธีการถดถอยแบบขั้นบันได

การคัดเลือกตัวแปรอิสระด้วยวิธีการถดถอยแบบขั้นบันได (Stepwise regressions) ในตัวแบบเชิงเส้นทั่วไปเป็นวิธีการคัดเลือกตัวแปรอิสระที่นิยมใช้อย่างมาก วิธี Stepwise regressions จะแก้ปัญหาโดยตัวแปรอิสระที่เข้ามาก่อนจะถูกคัดออกจากตัวแบบการถดถอยเมื่อมีตัวแปรอิสระอื่นที่มีความสัมพันธ์กับมันถูกคัดเข้ามา การกำหนดระดับนัยสำคัญของการคัดตัวแปรอิสระเข้าและออกจากตัวแบบการถดถอยอาจกำหนดให้เท่ากันหรือต่างกัน ในกรณีกำหนดให้ต่างกันจะกำหนดระดับนัยสำคัญของการคัดตัวแปรอิสระเข้าให้ต่ำกว่าระดับนัยสำคัญของการคัด

ตัวแปรอิสระออก เพื่อให้ตัวแปรอิสระที่ถูกคัดเข้ามาก่อนมีโอกาสน้อยที่จะถูกคัดออกเมื่อมีตัวแปรอิสระใหม่ถูกคัดเข้ามาในตัวแบบการถดถอยภายหลัง (ธิดาเดียว มยุรีสวรรค์, 2559)

ส่วนการคัดเลือกตัวแปรอิสระตามวิธีการถดถอยแบบขั้นบันได ในตัวแบบเชิงเส้นนัยทั่วไปจะใช้ Akaike's Information Criterion (AIC) เป็นเกณฑ์ในการพิจารณาการคัดตัวแปรเข้าหรือคัดตัวแปรออกจากสมการ

2.6 เกณฑ์ที่ใช้ในการคัดเลือกตัวแบบ

2.6.1 ค่าเปอร์เซ็นต์ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percent Error : MAPE)

การพยากรณ์ที่ดีนั้นควรมีค่าความคลาดเคลื่อนในการพยากรณ์ต่ำ จึงจะสามารถพยากรณ์ได้อย่างแม่นยำ ซึ่งการวัดความแม่นในการพยากรณ์ครั้งนี้ใช้ค่าเปอร์เซ็นต์ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAPE) เป็นค่าวัดความถูกต้องของการพยากรณ์ที่วัดจากขนาดของความคลาดเคลื่อนของการพยากรณ์กับข้อมูลจริง โดยถ้าค่า MAPE มีค่าน้อย ตัวแบบที่ได้ก็จะให้ค่าพยากรณ์ที่ความแม่นสูง

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\%$$

โดยที่

y_i คือ ข้อมูลจริง

$\hat{y}_i$ คือ ค่าพยากรณ์

n คือ จำนวนข้อมูล

ตาราง 2.2 เกณฑ์การประเมินผลการพยากรณ์โดยค่าเปอร์เซ็นต์ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย

MAPE (%)	การประเมินผล
MAPE ≤ 10%	การพยากรณ์มีความแม่นสูง
10% < MAPE ≤ 20%	ความแม่นในการพยากรณ์อยู่ในระดับดี
20% < MAPE ≤ 50%	ความแม่นในการพยากรณ์อยู่ในระดับพอใช้
MAPE > 50%	การพยากรณ์ไม่แม่นยำ

ที่มา : (Lewis, 1982)

2.6.2 ค่ารากของความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error : RMSE)

เป็นวิธีการวัดความคลาดเคลื่อนจากค่าที่พยากรณ์จากตัวแบบจำลองกับค่าจริงที่เกิดขึ้น หากค่า RMSE มีค่าน้อยแสดงว่าแบบจำลองสามารถประมาณค่าได้ใกล้เคียงกับค่าจริง ดังนั้นหากค่านี้มีค่าเท่ากับศูนย์แล้วจะหมายความว่าไม่เกิดความคลาดเคลื่อนในแบบจำลองนี้ ค่า RMSE สามารถคำนวณได้จากสมการดังนี้ (ณัฐภัทร แก้วรัตนภัทร์ และคณะ, 2555)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}$$

โดยที่

y_i = ข้อมูลจริง

$\hat{y}_i$ = ค่าการพยากรณ์

n = จำนวนข้อมูล

2.7 ข้อมูลน้ำมันสำเร็จรูปในประเทศไทย

2.7.1 น้ำมันเชื้อเพลิง (มารู้จักน้ำมันเชื้อเพลิง, 2560)

น้ำมันเชื้อเพลิง หมายถึง ของเหลวที่ได้จากการกลั่นน้ำมันดิบ หลังจากนั้นจึงนำมาปรับปรุงคุณภาพให้เหมาะสมต่อการใช้งาน เพื่อใช้เผาให้เป็นพลังงานในการขับเคลื่อนของเครื่องยนต์เครื่องเทอร์ไบน์ หรือใช้งานในโรงงานอุตสาหกรรม เช่น ต้มน้ำในหม้อไอน้ำ (Boiler) ใช้ในเตาอบเครื่องปั้นดินเผา หรือในโรงงานเซรามิก และใช้ในการทำความร้อน ให้แสงสว่าง เป็นต้น โดยทั่วไปน้ำมันเชื้อเพลิงที่จำหน่ายในสถานีบริการ จะมีอยู่ 2 ประเภท คือ น้ำมันเชื้อเพลิงที่ใช้กับเครื่องยนต์เบนซิน และน้ำมันเชื้อเพลิงที่ใช้กับเครื่องยนต์ดีเซล

น้ำมันเบนซิน (Gasoline) เป็นน้ำมันเชื้อเพลิงส่วนที่เบาที่สุด ที่ได้จากการกลั่นน้ำมันดิบส่วนนี้เรียกว่า แนฟธา (Naphtha) แล้วจึงนำมาปรับปรุงคุณภาพ ที่สำคัญคือการเพิ่มค่าออกเทน น้ำมันเบนซินปัจจุบันมี 2 ประเภท คือ น้ำมันเบนซินรถยนต์ (Motor Gasoline) และน้ำมันเบนซินอากาศยาน (Aviation Gasoline) ซึ่งทั้งสองประเภทนี้มีข้อแตกต่างกันตรงที่น้ำมันอากาศยานจะมีค่าออกเทนสูงกว่าน้ำมันเบนซินรถยนต์มาก

ประเภทของน้ำมันเชื้อเพลิง ได้แก่

น้ำมันแก๊สโซฮอล์ 91 มีคุณสมบัติตามมาตรฐานที่กำหนด และสามารถใช้ทดแทนน้ำมันเบนซิน 91 ธรรมดาได้ โดยมีส่วนผสมระหว่างเอทานอลหรือเอทิล แอลกอฮอล์ มีความบริสุทธิ์ 99.5% ผสมกับน้ำมันเบนซิน 91 ในอัตรา ส่วน น้ำมัน 9 ส่วน เอทานอล 1 ส่วนผลดีต่อเครื่องยนต์

ไม่มีผลกระทบต่อสมรรถนะเครื่องยนต์ และอัตราการเร่ง ไม่แตกต่างจากน้ำมันเบนซิน 91 สามารถเติมผสมกับน้ำมันเบนซินที่อยู่ในถังได้

น้ำมันแก๊สโซฮอล์ 95 มีคุณสมบัติตามมาตรฐานที่กำหนด และสามารถใช้ทดแทนน้ำมันเบนซิน 95 ธรรมดาได้มีส่วนผสมระหว่างเอทานอลหรือเอทิลแอลกอฮอล์มีความบริสุทธิ์ 99.5% ผสมกับน้ำมันเบนซิน 95 ในอัตรา ส่วน น้ำมัน 9 ส่วน เอทานอล 1 ส่วนน้ำมันแก๊สโซฮอล์ 95 แตกต่างจากน้ำมันเบนซิน 95 โดยน้ำมันแก๊สโซฮอล์ 95 ผลิตจากน้ำมันเบนซินออกเทน 91 ผสมกับเอทานอลซึ่งเป็นตัวเพิ่มค่าออกเทน ทำให้ได้แก๊สโซฮอล์ที่มีออกเทนเท่ากับน้ำมันเบนซิน 95 ที่ใช้สาร MTBE (Methyl Tertiary Butyl Ether) เป็นสารเพิ่มค่าออกเทนแต่สาร MTBE มีข้อเสียคือ ทำให้เกิดการปนเปื้อนกับน้ำใต้ดินและน้ำดื่ม

น้ำมันแก๊สโซฮอล์ อี 85 คือ น้ำมันที่มีส่วนผสมน้ำมันเชื้อเพลิงที่ได้จากการนำน้ำมันเบนซินไร้สารตะกั่วผสมกับเอทานอล หรือเอทิลแอลกอฮอล์ ซึ่งเป็นแอลกอ ฮอล์บริสุทธิ์ 99.5% ในอัตราส่วนเบนซิน 15% ต่อเอทา นอล 85% ได้เป็นน้ำมัน

น้ำมันแก๊สโซฮอล์ อี 20 น้ำมันที่มีส่วนผสมน้ำมันเชื้อเพลิงที่ได้จากการนำน้ำมันเบนซินไร้สารตะกั่วผสมกับเอทานอล หรือเอทิลแอลกอฮอล์ ซึ่งเป็นแอลกอฮอล์บริสุทธิ์ 99.5% ในอัตราส่วนเบนซิน 80% ต่อเอทานอล 20%

น้ำมันเบนซิน ออกเทน 95 น้ำมันชนิดนี้ รถทุกคันที่ใช้น้ำมันเชื้อเพลิงชนิดเบนซิน สามารถใช้ได้หมด เนื่องจากเป็นน้ำมันที่ไม่มีส่วนผสมของ เอทิลแอลกอฮอล์ และมีออกเทนที่ให้ค่าสูง มีการเผาไหม้ที่ดีที่สุดของน้ำมันในขณะนี้ และมีการป้องกันการน็อคของเครื่องยนต์สูง การเผาไหม้ของเครื่องยนต์จึงสมบูรณ์ และให้กำลังของการจุดระเบิดสูงตามมา สมรรถนะการขับขี่ตอบสนองได้เร็ว แต่ติดตรงที่มีราคาที่สูงกว่าน้ำมันชนิดอื่นๆ เป็นกฎธรรมชาติที่ว่า ของดีต้องแพงไว้ก่อน

น้ำมันเบนซิน ออกเทน 91 น้ำมันชนิดนี้ รถทุกคันอาจจะสามารถใช้ได้ ยกเว้นรถที่มีระบุไว้ว่า เติมน้ำมันชนิด เบนซินออกเทน 95 เท่านั้น ซึ่งหากเราฝืนเติมออกเทน 91 เข้าไปในรถที่มีระบุคำว่า เบนซินออกเทน 95 อาการที่รถจะแสดงออกมาให้เราทราบว่าใช้น้ำมันผิดประเภท ก็อาจจะแค่เครื่องยนต์สะดุด เดินเบาไม่เรียบ แต่รถสามารถวิ่งได้ เป็นน้ำมันที่ไม่มีส่วนผสมของ เอทิลแอลกอฮอล์ และมีค่าออกเทนที่ให้ค่าต่ำกว่าออกเทน 95 ลงมา สมรรถนะการขับขี่ตอบสนองได้เร็ว เป็นรอง ออกเทน 95 เล็กน้อย แทบจะไม่เห็นผล ต้องพิสูจน์ด้วยการนำรถเข้า Test ที่ห้องแล็บ จึงจะรู้ได้อย่างชัดเจน น้ำมันชนิดนี้จะหาเติมได้ยาก เพราะบางปั๊มไม่มีให้บริการ เนื่องจากมีชนิดของน้ำมันเชื้อเพลิงอื่นเข้ามาแทนที่ได้ ส่วนราคาสูงรองมาจาก ออกเทน 95

ก๊าซ LPG หรือ ก๊าซหุงต้ม เป็นก๊าซที่ได้จากกระบวนการแยกก๊าซธรรมชาติและขบวนการกลั่นน้ำมันเป็น สารประกอบพวกไฮโดรคาร์บอน ประกอบด้วยก๊าซโปรเปน (Propane) และบิวเทน (Butane) เป็นส่วนประกอบหลัก มีคุณสมบัติหนักกว่าอากาศประมาณ 1.5 – 2 เท่า ไม่มีสี ไม่มีกลิ่น

ไม่เป็นพิษ จึงต้องเติมกลิ่นเหม็น (Ethyl Mercaptan) ลงไปเพื่อให้รู้ว่าก๊าซรั่ว ซึ่งอาจทำให้ติดไฟได้ และขยายตัวเมื่ออุณหภูมิสูง (จากสถานะของเหลวกลายเป็นไอ ขยายตัวประมาณ 250 เท่า) จะมีสถานะเป็นไอที่อุณหภูมิปกติและความดันบรรยากาศ มีออกเทน 105 – 110 ก๊าซ LPG 1 ลิตร หนักประมาณ 0.5 กิโลกรัม LPG เมื่อเผาไหม้จะมีมลภาวะต่ำกว่าน้ำมัน

ก๊าซ NGV (Natural Gas for Vehicle : NGV) มีภาษาเชิงวิชาการว่า ก๊าซซีเอ็นจี (Compressed Natural Gas : CNG) คือ ก๊าซธรรมชาติที่มี “มีเทน” เป็นส่วนประกอบหลักและถูกอัดจนมีความดันสูง ซึ่งในบางประเทศเรียกว่า “ก๊าซธรรมชาติอัด” (ซีเอ็นจี) ซึ่งถูกอัดที่แรงดัน 200 bar หรือ 3,000 psi และถูกกักเก็บไว้ในถังบรรจุก๊าซธรรมชาติอัดที่ถูกผลิตขึ้นมาเป็นพิเศษให้สามารถรองรับแรงดันได้ โดยมีสภาพเป็นก๊าซหรือไอที่อุณหภูมิและความดันบรรยากาศ โดยมีค่าความถ่วงจำเพาะต่ำกว่าอากาศ จึงเบากว่าอากาศ เมื่อเกิดการรั่วไหลจะฟุ้งกระจายไปตามบรรยากาศอย่างรวดเร็ว จึงไม่มีการสะสมลุกไหม้บนพื้นราบ

2.7.2 การเปลี่ยนแปลงราคาน้ำมันสำเร็จรูปของประเทศไทย

การเปลี่ยนแปลงราคาปลีกน้ำมันของไทย ตั้งแต่อดีตจนถึงปัจจุบันจะเปลี่ยนแปลงขึ้นลงตามต้นทุนที่เปลี่ยนไป หรือการประกาศราคาของโรงกลั่น โดยช่วงยกเลิกควบคุมราคาขายปลีก แม้รัฐบาลจะควบคุมราคาขายปลีก ให้อยู่ในระดับคงที่เป็นระยะเวลาหนึ่ง แต่ในความเป็นจริงการกำหนดราคาของโรงกลั่นมีการเปลี่ยนแปลงทุกสัปดาห์ ตามราคาตลาดโลกที่เปลี่ยนแปลงไปโดยรัฐได้ใช้ ระบบกองทุนน้ำมันเชื้อเพลิง เพื่อรักษาระดับราคาขายส่งที่ออกจากโรงกลั่น และราคานำเข้าให้อยู่ในระดับคงที่ซึ่งส่งผลให้ราคาขายปลีกไม่เปลี่ยนแปลงหลังจากมีการยกเลิกการควบคุมแล้วราคาขายส่งจะมีการเปลี่ยนแปลงตาม ณ โรงกลั่น ซึ่งโรงกลั่นเป็นผู้กำหนดราคาและส่งผลให้ราคาขายปลีกเปลี่ยนแปลงตามในที่สุด จากการที่ประเทศไทยเป็นประเทศผู้นำเข้าน้ำมันโดยร้อยละ 90 ของการใช้น้ำมันเชื้อเพลิงต้องนำเข้ามาจากต่างประเทศในรูปของน้ำมันดิบและน้ำมันสำเร็จรูปบางส่วนประกอบกับการค้าน้ำมันเป็นไปอย่างเสรี ดังนั้นการกำหนดราคาน้ำมันของโรงกลั่นจึงขึ้นอยู่กับการเปลี่ยนแปลงของราคาน้ำมันในตลาดโลก และการเปลี่ยนแปลงของอัตราแลกเปลี่ยนเงินตรา (หรือค่าเงินบาท) ซึ่งเป็นต้นทุนในการนำเข้าน้ำมันเชื้อเพลิง (พรพิมล ศรีประเสริฐรัตน์, 2549)

2.7.3 การเปลี่ยนแปลงของราคาน้ำมันสำเร็จรูปในตลาดโลก

ตลาดซื้อขายน้ำมันระหว่างประเทศที่สำคัญ คือ ตลาดในภูมิภาคนั้นๆ ได้แก่ ตลาดในสหรัฐอเมริกา ยุโรป ตะวันออกกลางสิงคโปร์ ราคาตลาดเหล่านี้เรียกว่า ราคาตลาดอ้างอิง เพราะตลาดในประเทศแถบภูมิภาคนั้นๆ มักจะกำหนดราคาโดยคำนึงถึงตลาดเหล่านี้ ราคาตลาดอ้างอิงจะเปลี่ยนแปลงสอดคล้องกับต้นทุน คือราคาน้ำมันดิบ ประกอบกับความต้องการปริมาณการผลิตใน

ภูมิภาคนั้นๆ จากข้อมูลในอดีตพบว่า ราคาน้ำมันสำเร็จรูปในทุกตลาดปรับตัวเคลื่อนไหวไปในทิศทางและระดับเดียวกัน บางครั้งราคาน้ำมันสำเร็จรูปในบางตลาดอาจจะปรับตัวแตกต่างจากตลาดอื่นเนื่องจากมีอุปสงค์และอุปทานในภูมิภาคที่แตกต่างกันไปแต่ก็จะเป็นเพียงช่วงสั้นๆ เท่านั้น เพราะระดับราคาที่แตกต่างกันจะทำให้เกิดภาวะของการไหลเข้า/ออกของน้ำมันจากตลาดอื่นจนระดับราคาของตลาดนั้นปรับตัว สู่ภาวะสมดุลกับตลาดอื่น (พรพิมล ศรีประเสริฐรัตน์, 2549)

2.7.4 การกำหนดราคาน้ำมันสำเร็จรูปของไทย

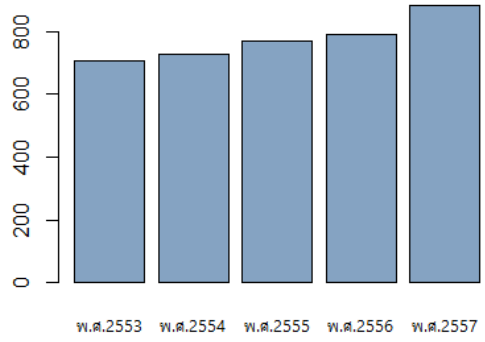
การกำหนดราคาน้ำมันสำเร็จรูปของโรงกลั่นไทย ต้องแข่งขันกับการนำเข้าจากต่างประเทศ ดังนั้นการกำหนดราคาน้ำมันสำเร็จรูปประดับค้าส่งจึงต้องกำหนดราคาในระดับที่แข่งขันกับราคานำเข้าที่ถูกที่สุด ซึ่งหมายถึงต้นทุนการส่งออกจากต่างประเทศมายังประเทศไทยในระดับต่ำสุด การกำหนดราคาน้ำมันสำเร็จรูปของโรงกลั่นน้ำมันจึงใช้หลักการเสมอภาคกับการนำเข้า (Import Parity Basis) และได้ใช้ตลาดสิงคโปร์เป็นตลาดอ้างอิงการกำหนดราคาน้ำมันสำเร็จรูปดังกล่าว ในการกำหนดราคาหากโครงการกำหนดราคาสูงกว่าการนำเข้าจากสิงคโปร์ ผู้ค้าน้ำมันจะนำเข้าแทนการซื้อจากโรงกลั่นในประเทศ แต่หากกำหนดราคาต่ำกว่าราคานำเข้าจะทำให้โรงกลั่นได้รับผลตอบแทนต่ำกว่าที่ควรย่อมไม่เกิดแรงจูงใจให้เกิดการลงทุนธุรกิจการกลั่นในประเทศไทย (พรพิมล ศรีประเสริฐรัตน์, 2549)

2.7.5 สาเหตุที่ใช้ราคาในตลาดจรสิงคโปร์เป็นฐานการกำหนดราคาน้ำมันสำเร็จรูปของไทย

ประเทศสิงคโปร์ถือเป็นประเทศส่งออกน้ำมันสำเร็จรูปรายใหญ่ที่สุดในภูมิภาคเอเชีย คำว่า “ราคาสิงคโปร์” ไม่ใช่ราคาน้ำมันสำเร็จรูปที่ประกาศโดยรัฐบาล หรือโรงกลั่นของประเทศสิงคโปร์ หรือราคาขายปลีกในประเทศสิงคโปร์ แต่เป็นราคาซื้อขายน้ำมันระหว่างผู้ค้าน้ำมันมากกว่า 300 ราย ในภูมิภาคเอเชียที่ตกลงกันผ่านตลาดกลางที่ประเทศสิงคโปร์เป็นปริมาณมหาศาล การที่ประเทศไทยตั้งอยู่ในภูมิภาคเอเชียและตั้งอยู่ใกล้กับประเทศสิงคโปร์ จึงเลือกที่จะอ้างอิงราคาน้ำมันสำเร็จรูปจากประเทศสิงคโปร์ ซึ่งสะท้อนระดับราคาที่สมดุลกับกลไกระบบการค้าเสรีของตลาดในภูมิภาคนี้ ซึ่งยากต่อการปั่นราคา อีกทั้งการเปลี่ยนแปลงของราคาก็สอดคล้องอย่างเป็นสากลกับตลาดซื้อขายน้ำมันอื่นๆ ทั่วโลก (ศูนย์ข้อมูล & ข่าวสืบสวนเพื่อสิทธิพลเมือง, 2557)

2.7.6 ปริมาณการใช้น้ำมันสำเร็จรูปในประเทศไทย

ปริมาณการใช้น้ำมันสำเร็จรูปในประเทศไทยตั้งแต่ปี พ.ศ. 2553 ถึง พ.ศ. 2557 โดยในปีพ.ศ. 2553 ถึง พ.ศ. 2557 มีการใช้น้ำมันสำเร็จรูป อยู่ที่ระดับ 704, 728, 768, 790 และ 881 พันบาร์เรลต่อวัน ตามลำดับ จะเห็นว่ามีปริมาณการใช้น้ำมันสำเร็จรูปเพิ่มขึ้นทุกปี สาเหตุที่มีการใช้น้ำมันสำเร็จรูปเพิ่มสูงขึ้น เนื่องจากมีการฟื้นตัวของเศรษฐกิจโลก และนักลงทุนก็มีความมั่นใจมากขึ้น ทำให้ปริมาณการใช้น้ำมันเพิ่มขึ้น (ศูนย์พยากรณ์และสารสนเทศพลังงาน, 2556)



รูปภาพที่ 2.3 ปริมาณการใช้น้ำมันสำเร็จรูปในประเทศไทย (พันบาร์เรล/วัน)

2.8 งานวิจัยที่เกี่ยวข้อง

(ฐิติรัตน์ แพทย์มงคล, 2552) ได้ศึกษาการกำหนดราคาขายปลีกน้ำมันดีเซล และสถานการณ์ความเคลื่อนไหวของราคาขายปลีกน้ำมันดีเซลในประเทศไทย และปัจจัยที่มีผลต่อการปรับตัวของราคาขายปลีกน้ำมันดีเซลในประเทศไทย ระหว่างปี พ.ศ. 2546 – 2552 (ไตรมาสที่1) โดยมีการเก็บรวบรวมข้อมูลเป็นรายเดือน ตั้งแต่เดือนมกราคม ปี พ.ศ.2546 ถึงเดือนมีนาคม ปี พ.ศ.2552 รวมทั้งหมด 75 เดือน เกี่ยวกับราคาน้ำมันดีเซลภายในประเทศและปัจจัยต่างๆ ที่มีผลต่อการปรับตัวของราคาน้ำมันดีเซล โดยทำการทดสอบข้อมูลว่ามีเสถียรภาพหรือไม่ โดยใช้วิธี Unit Root ทดสอบความสัมพันธ์ในระยะยาว โดยใช้เทคนิค Cointegration แล้วจึงนำตัวแปรอิสระที่ผ่านการทดสอบมาประมาณค่าด้วยสมการถดถอยเชิงซ้อน (Multiple Regression) จากผลการศึกษาพบว่า น้ำมันดีเซลที่ใช้ในการบริโภคเป็นการนำเข้าในรูปของน้ำมันดิบส่วนใหญ่มาจากกลุ่มโอเปค และอีกส่วนหนึ่งมาจากการนำเข้าน้ำมันดีเซลสำเร็จรูป และมาจากการผลิตภายในประเทศไทย ราคาขายปลีกน้ำมันดีเซลในประเทศไทยปรับตัวสูงขึ้นมาตลอด เนื่องจากการเข้ามาเก็งกำไรของกองทุน (Head Fund) ทำให้ราคาน้ำมันดิบในตลาดโลกสูงขึ้นราคาน้ำมันดีเซลในประเทศไทยจึงสูงขึ้นตามราคาน้ำมันดิบ ส่วนการศึกษาปัจจัยที่มีผลต่อการปรับตัวของราคาน้ำมันดีเซลในประเทศ

ไทย พบว่า ค่าธรรมเนียมกองทุนน้ำมัน ราคาน้ำมันดีเซลในตลาดสิงคโปร์ ราคาน้ำมันดิบในตลาดโลก ราคาก๊าซเอ็นจีวี และอัตราแลกเปลี่ยน เป็นปัจจัยที่มีผลต่อการปรับตัวของราคาน้ำมันดีเซลในประเทศไทย

(เรณู นิยมเดชา, 2554) ได้นำตัวแบบเชิงเส้นนัยทั่วไปมาประยุกต์ใช้กับการศึกษาต้นทุนความเสียหายของการประกันอัคคีภัย ซึ่งในงานวิจัยครั้งนี้ได้แบ่งออกเป็น 2 ส่วน ได้แก่ ส่วนที่หนึ่ง การคำนวณต้นทุนความเสียหายโดยใช้ตัวแบบเชิงเส้นนัยทั่วไปในการพยากรการณ์ และส่วนที่สองได้นำกระบวนการทางคณิตศาสตร์ประกันภัยมาใช้ในการวิเคราะห์ พบว่ามีปัจจัยเสี่ยง 2 ปัจจัย ปัจจัยแรกคือ บ้านอยู่อาศัยที่ตั้งอยู่บริเวณที่เป็นภัยโดดเดี่ยวธรรมดาทั่วไปมีความเสี่ยงสูงกว่าบ้านอยู่อาศัยที่ตั้งอยู่บริเวณภัยไม่โดดเดี่ยว ปัจจัยที่สองคือ บ้านอยู่อาศัยในเขตกรุงเทพมหานครมีโอกาสที่ความถี่และความรุนแรงจะสูงกว่าบ้านอยู่อาศัยในต่างจังหวัด

(กมลวรรณ สารพานิช, 2555) ใช้วิธีอารีมาและอารีแมกซ์ ในการพยากรณ์ราคาน้ำมันดิบล่วงหน้าในตลาดฟิวเจอร์ไนเม็กซ์ ประเทศสหรัฐอเมริกา โดยใช้ข้อมูลอนุกรมเวลาแบบรายสัปดาห์ตั้งแต่วันที่ 7 มกราคม พ.ศ. 2543 ถึง 27 มกราคม พ.ศ. 2555 รวมจำนวน 630 ตัวอย่าง พบว่าแบบจำลองอารีมา AR(1) AR(8) MA(12) เป็นแบบจำลองที่มีความเหมาะสมที่สุด เพราะมีค่า Root Mean Squared Error (RMSE) และค่า Theil's Inequality Coefficient ต่ำที่สุด จึงสรุปได้ว่าแบบจำลองนี้มีความเหมาะสมที่จะเป็นตัวแทนการพยากรณ์ราคาน้ำมันดิบล่วงหน้า จึงได้นำแบบจำลองดังกล่าวไปพยากรณ์ราคาน้ำมันดิบล่วงหน้า 2 สัปดาห์ถัดไป ได้ราคาคือ 104.717 และ 105.053 ดอลลาร์ต่อบาร์เรล์ ต่อมาได้ใช้แบบจำลองอารีแมกซ์ในการพยากรณ์ราคาน้ำมันดิบล่วงหน้า AR(1) MA(12) SPOT(-2) GOLD(-2) USDEUR(-2) เป็นแบบจำลองที่มีความเหมาะสมที่สุด จึงสรุปได้ว่าแบบจำลองนี้มีความเหมาะสมที่จะเป็นตัวแทนการพยากรณ์ราคาน้ำมันดิบล่วงหน้า จึงได้นำแบบจำลองดังกล่าวไปพยากรณ์ราคาน้ำมันดิบล่วงหน้า 2 สัปดาห์ถัดไป ได้ราคาคือ 108.723 และ 110.593 ดอลลาร์ต่อบาร์เรล์ เมื่อได้ผลการพยากรณ์จากแบบจำลองอารีมาและอารีแมกซ์ จึงนำค่า RMSE ทั้ง 2 แบบจำลองมาเปรียบเทียบ เพื่อหาค่า Relative Root Mean Squared Error (RRMSE) เปรียบเทียบ ซึ่งค่าที่คำนวณได้เท่ากับ 1.004 ซึ่งหมายถึงแบบจำลองทั้ง 2 แบบให้ผลการพยากรณ์ที่แม่นยำไม่แตกต่างกัน

(วรางคณา กีรติวิบูลย์, 2556) ศึกษาการสร้างตัวแบบพยากรณ์ที่เหมาะสมสำหรับราคาขายปลีกรายวันของน้ำมันแก๊สโซฮอล์ 95 ในเขตกรุงเทพมหานครและปริมณฑล ตั้งแต่วันที่ 1 มกราคม 2555 ถึงวันที่ 25 กุมภาพันธ์ 2556 จำนวน 422 ค่า โดยแบ่งข้อมูลเป็น 2 ชุด ชุดที่ 1 จำนวน 412 ค่า ตั้งแต่วันที่ 1 มกราคม 2555 ถึงวันที่ 15 กุมภาพันธ์ 2556 ใช้สำหรับการสร้างตัวแบบพยากรณ์ โดยวิธีบอกซ์-เจนกินส์ วิธีการทำให้เรียบแบบเอกซ์โพเนนเชียลด้วยวิธีของโฮลต์ วิธีการทำให้เรียบแบบเอกซ์โพเนนเชียลด้วยวิธีของบราวน์ และวิธีการทำให้เรียบ

แบบเอกซ์โพเนนเชียลที่มีแนวโน้มแบบ damped ชุดที่ 2 จำนวน 10 ค่า ตั้งแต่วันที่ 16 ถึง 25 กุมภาพันธ์ 2556 สำหรับนำมาตรวจสอบความแม่นยำของตัวแบบพยากรณ์ด้วยเกณฑ์เปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ยที่ต่ำที่สุด พบว่า วิธีการทำให้เรียบแบบเอกซ์โพเนนเชียลด้วยวิธีของบราวน์เป็นวิธีที่มีความเหมาะสมกับอนุกรมเวลาชุดนี้มากที่สุด

(รจนาถ จันทร์มี, 2555) ศึกษาถึงประสิทธิภาพของช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยของประชากรที่มีการแจกแจงเลขชี้กำลัง และการแจกแจงแกมมา โดยใช้ค่ามัธยฐานของตัวอย่าง ด้วยวิธีตัวสถิติที่วิธีค่าเบี่ยงเบนมาตรฐานเทียม วิธีตัวประมาณดาวน์ตัน วิธีค่าเบี่ยงเบนสัมบูรณ์มัธยฐาน โดยเกณฑ์ที่ใช้ในการพิจารณาคือ ความน่าจะเป็นที่ครอบคลุมและความกว้างเฉลี่ยของช่วงความเชื่อมั่น กำหนดให้ขนาดตัวอย่างเป็น 10 25 50 และ 100 และกำหนดให้สัมประสิทธิ์ความเชื่อมั่นเป็น 0.95 ทำการศึกษาโดยการจำลองด้วยเทคนิคมอนติคาร์โล โดยการทำซ้ำ 5,000 ครั้ง ผลการศึกษาคือ เมื่อพิจารณาจากความน่าจะเป็นที่ครอบคลุม ภายใต้การแจกแจงเลขชี้กำลังและการแจกแจงแกมมา วิธีตัวสถิติที่ให้ความน่าจะเป้นครอบคลุมใกล้เคียงค่าสัมประสิทธ์ความเชื่อมั่น 0.95 มากที่สุด ในทุกขนาดตัวอย่างและทุกค่าพารามิเตอร์ เมื่อพิจารณาจากความกว้างเฉลี่ยของช่วงความเชื่อมั่นเมื่อประชากรมีการแจกแจงเลขชี้กำลัง วิธีค่าเบี่ยงเบนสัมบูรณ์มัธยฐานให้ความกว้างของช่วงเฉลี่ยแคบที่สุด สำหรับพารามิเตอร์ เมื่อประชากรมีการแจกแจงแกมมาค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นวิธีตัวสถิติที่จะให้ค่าความกว้างเฉลี่ยของช่วงแคบที่สุดเมื่อพารามิเตอร์ มีค่าเท่ากับ 3, 7, 10 และวิธีค่าเบี่ยงเบนสัมบูรณ์มัธยฐานให้ความกว้างของช่วงเฉลี่ยแคบที่สุด เมื่อพารามิเตอร์ เท่ากับ 1

(สุนีย์ สัมมาทัต, 2557) ใช้ตัวแบบผสมเชิงเส้นวางนัยทั่วไป (Generalized Linear Mixed Model ; GLMM) ที่ตัวแปรตามมีการแจกแจงแบบปัวซง และมีอิทธิพลเชิงสุ่มแบบ Multivariate Conditional Autoregressive Model (MCAR) รวมอยู่ด้วย ในการประมาณอัตราป่วยโรคความดันโลหิตสูงและโรคหัวใจขาดเลือดในแต่ละจังหวัดของประเทศไทย หาปัจจัยที่มีอิทธิพลต่ออัตราป่วยเกิดโรคความดันโลหิตสูงและโรคหัวใจขาดเลือด และเพื่อสร้างแผนที่โรคความดันโลหิตสูงและโรคหัวใจขาดเลือดในแต่ละจังหวัดของประเทศไทย พบว่าปัจจัยที่มีอิทธิพลต่อความเสี่ยงของการป่วยโรคความดันโลหิตสูงและโรคหัวใจขาดเลือด คือ รายได้เฉลี่ยต่อหัว สัดส่วนนักเรียนนักศึกษา แผนที่โรคความดันโลหิตสูงและโรคหัวใจขาดเลือด ทำให้เห็นได้ง่ายกว่าพื้นที่ในมีอัตราป่วยสูง จึงใช้เป็นเครื่องมือที่สำคัญในการวางแผนป้องกันการเกิดโรคความดันโลหิตสูงและโรคหัวใจขาดเลือด

บทที่ 3
วิธีดำเนินการวิจัย

การดำเนินการวิจัยนี้จะมี 2 ส่วนที่นำเสนอคือ ส่วนแรกการวิจัยเพื่อเปรียบเทียบประสิทธิภาพของการพยากรณ์ โดยใช้ตัวแบบเชิงเส้นทั่วไป (General Linear Model) ใช้วิธีการประมาณค่าพารามิเตอร์ในตัวแบบด้วยวิธีตัวประมาณกำลังสองน้อยสุด (Least Square Estimation : LSE) และตัวแบบเชิงเส้นนัยทั่วไป (Generalized Linear Model) ใช้วิธีการประมาณค่าพารามิเตอร์ในตัวแบบด้วยวิธีตัวประมาณภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation : MLE) ส่วนที่ 2 จะศึกษาปริมาณการใช้น้ำมันสำเร็จรูปและหาปัจจัยที่ส่งผลต่อปริมาณการใช้น้ำมันสำเร็จรูปของประเทศไทย

3.1 แผนการดำเนินงานวิจัยในการสร้างแบบจำลอง

ในการวิจัยครั้งนี้กำหนดสถานการณ์ต่างๆ ดังนี้

3.1.1 กำหนดการแจกแจงของ Y

3.1.1.1 กำหนดให้ตัวแปรตาม Y มีการแจกแจงแกมมา ที่มีพารามิเตอร์รูปร่าง (Shape Parameter ; α) มีค่าเท่ากับ 2 และพารามิเตอร์ขนาด (Scale Parameter ; β) มีค่าเท่ากับ 1

3.1.1.2 กำหนดให้ตัวแปรตาม Y มีการแจกแจงอินเวอร์สเกาส์เซียน ที่มีพารามิเตอร์แสดงตำแหน่ง (Location Parameter ; μ) มีค่าเท่ากับ 1 และพารามิเตอร์รูปร่าง (Shape Parameter ; λ) มีค่าเท่ากับ 3

3.1.2 กำหนดให้ X มีการแจกแจงปรกติมาตรฐาน ที่มีพารามิเตอร์แสดงตำแหน่งหรือค่าเฉลี่ย (Location Parameter ; μ) มีค่าเท่ากับ 0 และพารามิเตอร์รูปร่างหรือความแปรปรวน (Shape Parameter ; σ^2) มีค่าเท่ากับ 1

3.1.3 กำหนดขนาดตัวอย่างในการศึกษา n = 30 , 50 และ 100 ตามลำดับ

3.1.4 กำหนดค่าความสัมพันธ์ระหว่าง X และ Y กำหนดเป็น 2 ระดับ คือ มีความสัมพันธ์ในระดับต่ำมาก โดยกำหนดค่าสัมประสิทธิ์สหสัมพันธ์น้อยกว่าหรือเท่ากับ 0.3 และมีความสัมพันธ์กันในระดับสูง โดยกำหนดค่าสัมประสิทธิ์สหสัมพันธ์มากกว่าหรือเท่ากับ 0.7 ส่วนค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร X_1 กับ X_2 กำหนดเป็นน้อยกว่าหรือเท่ากับ 0.3 และมากกว่าหรือเท่ากับ 0.7 เช่นเดียวกัน

3.1.5 วิเคราะห์ข้อมูลด้วยวิธีตัวแบบเชิงเส้นทั่วไปและวิธีตัวแบบเชิงเส้นนัยทั่วไป

3.1.5.1 สำหรับตัวแบบเชิงเส้นทั่วไป (General Linear Model) จะใช้ค่า Y ที่ไม่ถูกแปลงค่าและถูกแปลงค่าให้อยู่ในรูป $\frac{1}{y}$, $\frac{1}{y^2}$, $\log(y)$ และแปลงให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ ก่อนการประมาณค่าพารามิเตอร์ด้วยวิธีตัวประมาณกำลังสองน้อยสุด (Least Square Estimation : LSE) แล้วแปลงค่าข้อมูลกลับเพื่อหาค่าความคลาดเคลื่อนของการพยากรณ์

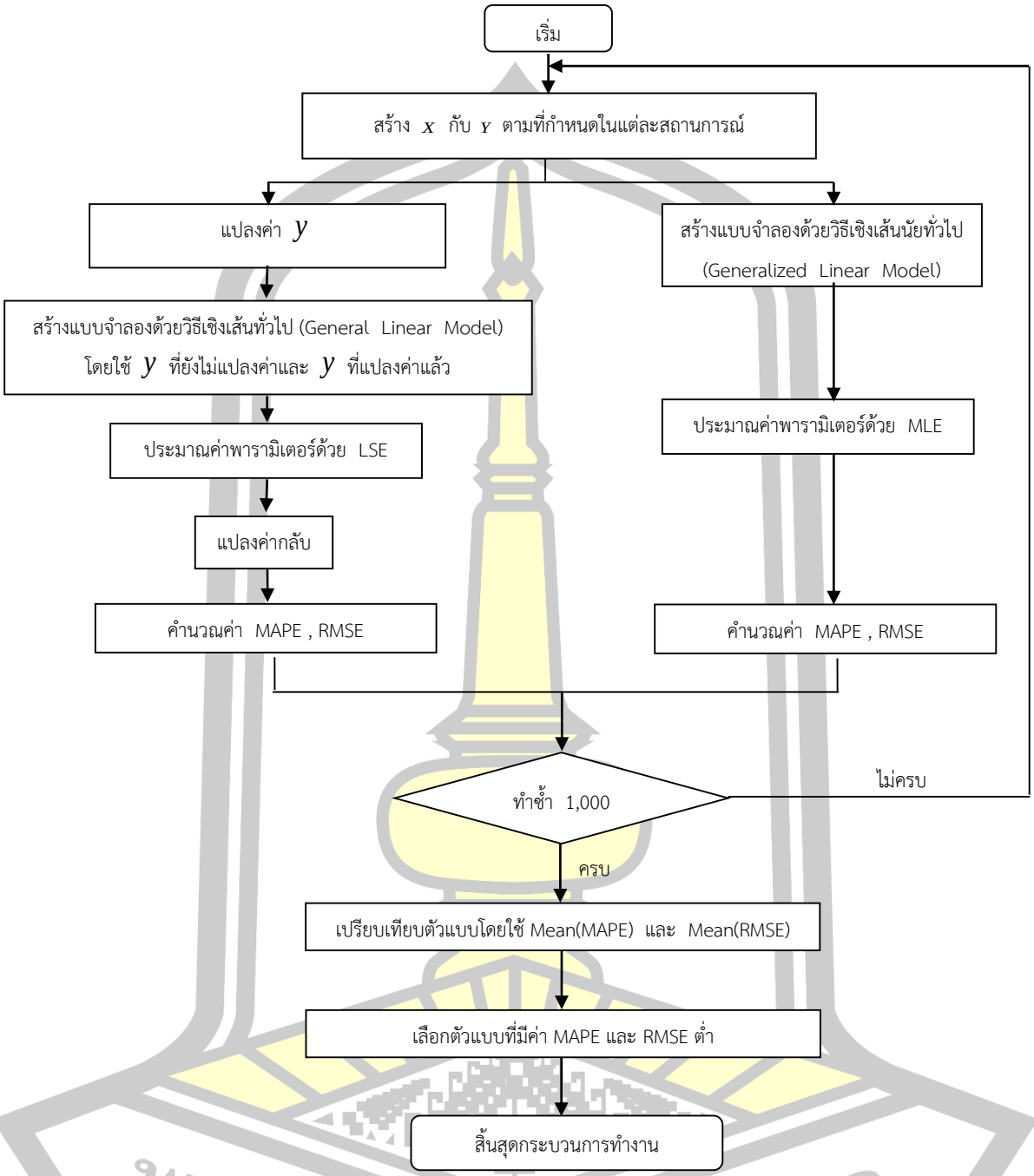
3.1.5.2 สำหรับตัวแบบเชิงเส้นนัยทั่วไป (Generalized Linear Model) ประมาณค่าพารามิเตอร์ด้วยวิธีตัวประมาณภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation : MLE) และองค์ประกอบเชิงสุ่มเป็นแกมมาที่มีฟังก์ชันเชื่อมโยงเป็น inverse และ $\log(\mu)$ และอินเวอร์สเกาส์เซียนที่มีฟังก์ชันเชื่อมโยงเป็น square root และ $\log(\mu)$

3.1.6 ใช้วิธีค่าเฉลี่ยของค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อน (MAPE) และรากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE) เป็นเกณฑ์ในการคัดเลือกตัวแบบ

3.1.7 กำหนดรอบประมวลผลในแต่ละสถานการณ์ 1,000 รอบ

3.1.8 เปรียบเทียบค่าเฉลี่ย MAPE และค่าเฉลี่ย RMSE

พหุน ปณฺฑิโต ชีเว



รูปภาพที่ 3.4 ขั้นตอนการจำลองข้อมูลเพื่อเปรียบเทียบประสิทธิภาพของตัวแบบเชิงเส้นทั่วไปและตัวแบบเชิงเส้นนัยทั่วไป

ตาราง 3.3 ตารางการกำหนดสถานการณ์ต่างๆ ในการจำลองข้อมูล

การแจกแจงของ Y	จำนวนตัวแปรอิสระ X	การแจกแจงของ X	ค่าสหสัมพันธ์ระหว่าง X กับ Y ; (r_{xy})	ค่าสหสัมพันธ์ระหว่าง X กับ X ; (r_{xx})	ขนาดตัวอย่าง (n)
$Gam(2,1)$	1	$X_1 \sim N(0,1)$	≤ 0.3	-	30
					50
					100
$Gam(2,1)$	1	$X_2 \sim N(0,1)$	≥ 0.7	-	30
					50
					100
$Gam(2,1)$	2	$X_1 \sim N(0,1)$ $X_2 \sim N(0,1)$	≤ 0.3 ≥ 0.7	≤ 0.3	30
					50
					100
$Gam(2,1)$	2	$X_1 \sim N(0,1)$ $X_2 \sim N(0,1)$	≤ 0.3 ≥ 0.7	≥ 0.7	30
					50
					100

ตาราง 3.3 ตารางการกำหนดสถานการณ์ต่างๆ ในการจำลองข้อมูล (ต่อ)

การแจกแจงของ Y	จำนวนตัวแปรอิสระ X	การแจกแจงของ X	ค่าสหสัมพันธ์ระหว่าง X กับ Y ; (r_{xy})	ค่าสหสัมพันธ์ระหว่าง X กับ X ; (r_{xx})	ขนาดตัวอย่าง (n)
$IG(1,3)$	1	$X_1 \sim N(0,1)$	≤ 0.3	-	30
					50
					100
$IG(1,3)$	1	$X_2 \sim N(0,1)$	≥ 0.7	-	30
					50
					100
$IG(1,3)$	2	$X_1 \sim N(0,1)$ $X_2 \sim N(0,1)$	≤ 0.3 ≥ 0.7	≤ 0.3	30
					50
					100
$IG(1,3)$	2	$X_1 \sim N(0,1)$ $X_2 \sim N(0,1)$	≤ 0.3 ≥ 0.7	≥ 0.7	30
					50
					100

3.2 ขั้นตอนในการประยุกต์ใช้ตัวแบบเชิงเส้นทั่วไปและตัวแบบเชิงเส้นนัยทั่วไป

3.2.1 รวบรวมข้อมูลเกี่ยวกับปริมาณการใช้น้ำมันสำเร็จรูปแต่ละชนิดซึ่งเป็นข้อมูลรายเดือน จำนวน 72 เดือน ตั้งแต่ปี พ.ศ. 2553 ถึง พ.ศ. 2558 ดังนี้

ตัวแปรตาม (Y)	**ตัวแปรอิสระ** (X)		
ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91 ใช้ข้อมูลตั้งแต่เดือน มกราคม พ.ศ. 2553 ถึง ธันวาคม พ.ศ.2558	P_G91 P_Ethanol G95	B95 E20	P_G95 P_E20 P_B95 WCO EXC
ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95 ใช้ข้อมูลตั้งแต่เดือน มกราคม พ.ศ. 2553 ถึง ธันวาคม พ.ศ.2558	P_E85 P_Ethanol B95	E20 E85	
ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20 ใช้ข้อมูลตั้งแต่เดือน มกราคม พ.ศ. 2553 ถึง กุมภาพันธ์ พ.ศ.2556	P_G91 P_Ethanol G95	G91 B95	
ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20 ใช้ข้อมูลตั้งแต่เดือน มีนาคม พ.ศ. 2556 ถึง ธันวาคม พ.ศ.2558	P_G91 P_Ethanol G95	G91 B95	
ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E85 ใช้ข้อมูลตั้งแต่เดือน มกราคม พ.ศ. 2553 ถึง ธันวาคม พ.ศ.2558	P_E85 P_G91 P_Ethanol G95	B95 G91 E20	
ปริมาณการใช้น้ำมันเบนซิน 91 ใช้ข้อมูลตั้งแต่เดือน มกราคม พ.ศ. 2553 ถึง ธันวาคม พ.ศ.2555	P_B91 P_G91 P_E85 P_NGV P_LPG G95	B95 G91 E20 E85 LPG NGV	
ปริมาณการใช้น้ำมันเบนซิน 95 ใช้ข้อมูลตั้งแต่เดือน กุมภาพันธ์ พ.ศ. 2553 ถึง มีนาคม พ.ศ.2557	P_G91 P_E85 P_NGV P_LPG G95	G91 E20 E85 LPG NGV	

ตัวแปรตาม (Y)	ตัวแปรอิสระ (X)		
	ตัวแปรอิสระ (ต่างกัน)		ตัวแปรอิสระ (เหมือนกัน)
ปริมาณการใช้แก๊ส LPG ใช้ข้อมูลตั้งแต่เดือน กุมภาพันธ์ พ.ศ. 2553 ถึง มีนาคม พ.ศ.2557	P_LPG	E20	P_G95
	P_G91	E85	P_E20
	P_E85	B91	P_B95
	P_NGV	B95	WCO
	P_B91	NGV	EXC
	G95	P_Ethanol	
	G91		
ปริมาณการใช้แก๊ส NGV ใช้ข้อมูลตั้งแต่เดือน กุมภาพันธ์ พ.ศ. 2553 ถึง มีนาคม พ.ศ.2557	P_NGV	E20	
	P_E85	E85	
	P_NGV	B91	
	P_B91	B95	
	G95	LPG	
	G91	P_Ethanol	

หมายเหตุ

G91 หมายถึง ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91 P_G91 หมายถึง ราคาน้ำมันแก๊สโซฮอล์ 91

G95 หมายถึง ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95 P_G95 หมายถึง ราคาน้ำมันแก๊สโซฮอล์ 95

E20 หมายถึง ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20 P_E20 หมายถึง ราคาน้ำมันแก๊สโซฮอล์ E20

E85 หมายถึง ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E85 P_E85 หมายถึง ราคาน้ำมันแก๊สโซฮอล์ E85

B91 หมายถึง ปริมาณการใช้น้ำมันเบนซิน 91 P_B91 หมายถึง ราคาน้ำมันเบนซิน 91

B95 หมายถึง ปริมาณการใช้น้ำมันเบนซิน 95 P_B95 หมายถึง ราคาน้ำมันเบนซิน 95

LPG หมายถึง ปริมาณการใช้แก๊ส LPG P_LPG หมายถึง ราคาแก๊ส LPG

NGV หมายถึง ปริมาณการใช้แก๊ส NGV P_NGV หมายถึง ราคาแก๊ส NGV

WCO หมายถึง ราคาน้ำมันดิบโลก P_Ethanol หมายถึง ราคาเอทานอล

EXC หมายถึง อัตราแลกเปลี่ยนเงินตรา

3.2.2 นำข้อมูลมาวิเคราะห์ด้วยวิธีตัวแบบเชิงเส้นทั่วไปและวิธีตัวแบบเชิงเส้นนัยทั่วไป

3.2.3 สำหรับวิธีการวิเคราะห์ตัวแบบเชิงเส้นทั่วไป ใช้ค่า Y ที่ไม่ถูกแปลงค่าและถูกแปลงค่าให้อยู่ในรูป $\frac{1}{y}$, $\frac{1}{y^2}$, $\log(y)$ และแปลงให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ แล้วประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุด (Least Square Estimation : LSE) และสถิติที่ใช้ทดสอบความมีนัยสำคัญของพารามิเตอร์ในตัวแบบ

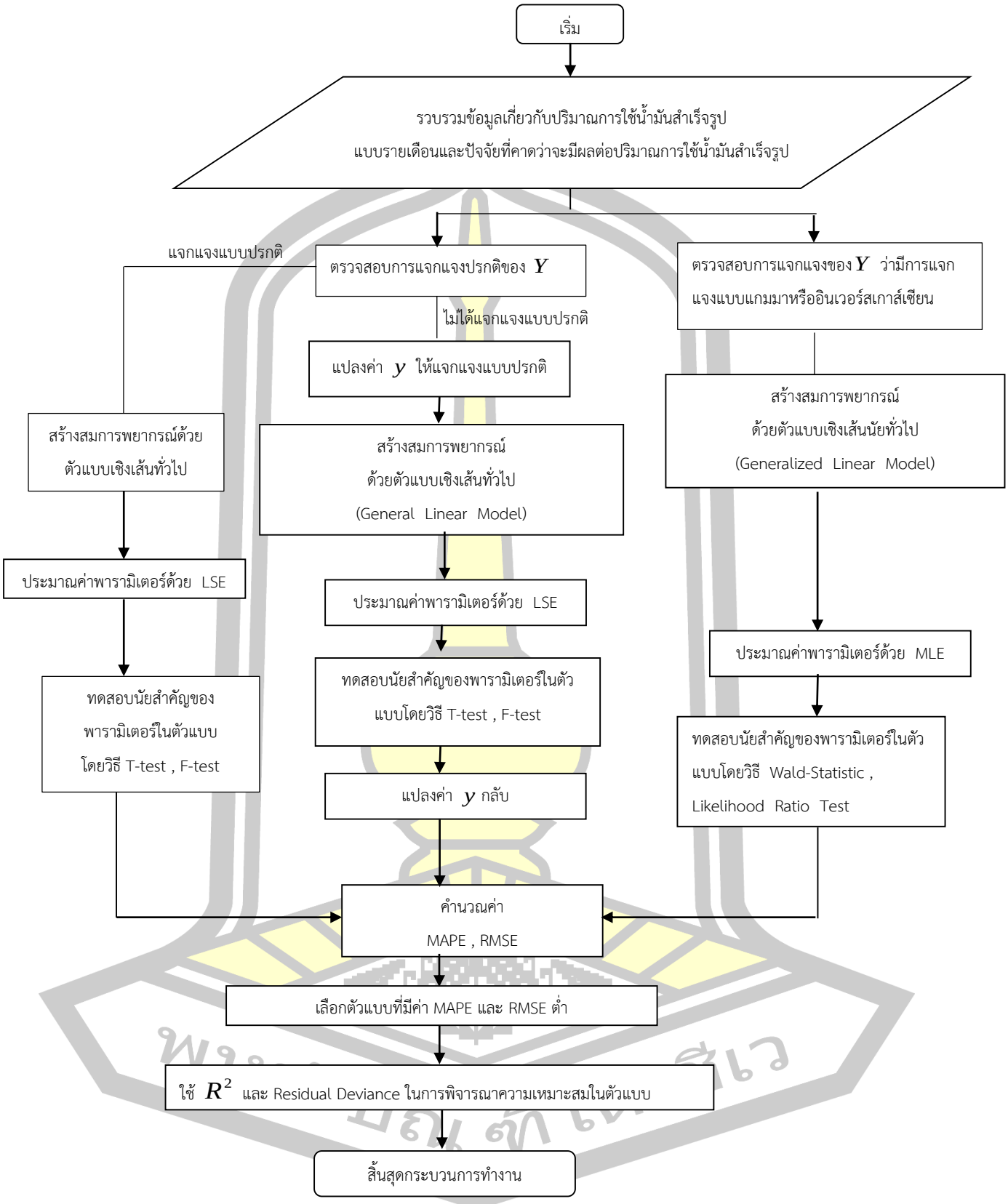
ตัดสินว่าตัวแปรอิสระตัวใดส่งผลต่อตัวแปรตามใช้ F-test และ T-test แล้วแปลงค่าข้อมูลกลับเพื่อหาค่าความคลาดเคลื่อนของการพยากรณ์

3.2.4 สำหรับวิธีการวิเคราะห์ตัวแบบเชิงเส้นนัยทั่วไป จะประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation : MLE) ใช้สถิติ Wald – Statistic และ Likelihood Ratio Test ทดสอบเพื่อพิจารณาตัวแปรอิสระที่ส่งผลต่อตัวแปรตาม

3.2.5 ใช้วิธีค่าเปอร์เซ็นต์ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAPE) และค่ารากของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ในการคัดเลือกตัวแบบ

3.2.6 ใช้ $\boldsymbol{R}^2$ และ Residual Deviance ในการพิจารณาความเหมาะสมของตัวแบบ

พหูน ปณ ฑิโต ชีเว



รูปภาพที่ 3.5 ขั้นตอนการหาสมการที่เหมาะสมในการพยากรณ์ปริมาณการใช้น้ำมันสำเร็จรูป

บทที่ 4
ผลการวิจัย

ในการนำเสนอผลการวิจัย เรื่อง ประสิทธิภาพของการพยากรณ์โดยตัวแบบเชิงเส้นนัยทั่วไปเมื่อข้อมูลมีการแจกแจงแบบเบ้ขวากับการประยุกต์สำหรับปริมาณการใช้น้ำมันสำเร็จรูปของประเทศไทย ผู้วิจัยจะนำเสนอเป็น 2 ส่วน ส่วนแรกนำเสนอผลการจำลองข้อมูลเพื่อเปรียบเทียบประสิทธิภาพในการพยากรณ์ เมื่อข้อมูลมีการแจกแจงแบบเบ้ขวาด้วยวิธีวิเคราะห์ตัวแบบเชิงเส้นนัยทั่วไป (Generalized Linear Model : GLMs) กับวิธีการวิเคราะห์ตัวแบบเชิงเส้นทั่วไป (General Linear Model : GLM) ส่วนที่ 2 นำเสนอผลการประยุกต์ใช้ตัวแบบเชิงเส้นนัยทั่วไป (Generalized Linear Model : GLMs) และวิธีการวิเคราะห์ตัวแบบเชิงเส้นทั่วไป (General Linear Model : GLM) กับข้อมูลปริมาณการใช้น้ำมันสำเร็จรูปของประเทศไทย

4.1 ผลการจำลองข้อมูลเพื่อหาตัวแบบที่มีประสิทธิภาพในการพยากรณ์ เมื่อข้อมูลมีการแจกแจงแบบเบ้ขวา

4.1.1 กรณีการจำลองข้อมูลให้ Y มีการแจกแจงแกมมา

ผลการจำลองข้อมูลเมื่อตัวแปรตาม Y มีการแจกแจงแกมมา ที่มีพารามิเตอร์รูปร่าง (α) มีค่าเท่ากับ 2 และพารามิเตอร์ขนาด (β) มีค่าเท่ากับ 1 และกำหนดให้ X มีการแจกแจงปรกติมาตรฐาน ที่มีพารามิเตอร์เป็น (0,1) ค่าความสัมพันธ์ระหว่าง X และ Y กำหนดเป็น 2 ระดับ คือ มีความสัมพันธ์ในระดับต่ำมากคือมีค่าสัมประสิทธิ์สหสัมพันธ์น้อยกว่าหรือเท่ากับ 0.3 และมีความสัมพันธ์กันในระดับสูง คือ มีค่าสัมประสิทธิ์สหสัมพันธ์มากกว่าหรือเท่ากับ 0.7 ในกรณีที่มีตัวแปรอิสระ (X) 2 ตัวจะกำหนดค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร X_1 กับ X_2 เป็นน้อยกว่าหรือเท่ากับ 0.3 และมากกว่าหรือเท่ากับ 0.7 เช่นเดียวกัน กำหนดขนาดตัวอย่าง (n) เป็น 30 50 และ 100 ตามลำดับ

พหุน ปณ ฑิโต ชีเว

ตาราง 4.4 ค่าเฉลี่ยของค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อน (MAPE) และรากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE) จำแนกตามขนาดตัวอย่าง ชนิดของตัวแบบและรูปแบบการแปลงค่าของ y เมื่อกำหนดให้ y มีการแจกแจงแกมมา (2,1) กรณีความสัมพันธ์ระหว่าง x และ y อยู่ในระดับต่ำ

ค่าสัมประสิทธิ์สหสัมพันธ์	n	ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	MAPE	RMSE
$r_{x_1,y} \le 0.3$	30	ตัวแบบเชิงเส้นทั่วไป	y	-	100.1349	**1.3371**
			$\frac{1}{y}$	-	**63.4600**	1.6009
			$\frac{1}{y^2}$	-	**59.6400**	1.8036
			$\log(y)$	-	76.1074	1.4049
			Normal	-	82.7375	1.3694
		ตัวแบบเชิงเส้นนัยทั่วไป	-	*Inverse*	100.4983	**1.3418**
			-	$\log(\mu)$	100.0033	**1.3429**
	50	ตัวแบบเชิงเส้นทั่วไป	y	-	103.8235	**1.3461**
			$\frac{1}{y}$	-	**65.2463**	1.6246
			$\frac{1}{y^2}$	-	**61.1844**	1.8378
			$\log(y)$	-	78.9964	1.4141
			Normal	-	85.9681	1.3784
		ตัวแบบเชิงเส้นนัยทั่วไป	-	*Inverse*	104.2087	**1.3510**
			-	$\log(\mu)$	103.7416	**1.3504**
	100	ตัวแบบเชิงเส้นทั่วไป	y	-	106.7163	**1.3582**
			$\frac{1}{y}$	-	**66.7734**	1.6479
			$\frac{1}{y^2}$	-	**63.9789**	1.9086
			$\log(y)$	-	81.4924	1.4253
			Normal	-	88.6474	1.3897
		ตัวแบบเชิงเส้นนัยทั่วไป	-	*Inverse*	107.1348	**1.3617**
			-	$\log(\mu)$	106.7073	**1.3602**

Normal : แปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ

จากตาราง 4.4 พบว่ากรณีที่ตัวแปรอิสระ X และตัวแปรตาม Y มีสัมประสิทธิ์สหสัมพันธ์ในระดับต่ำ เมื่อพิจารณาค่า MAPE พบว่าตัวแบบเชิงเส้นทั่วไป โดยการแปลงค่า y ให้อยู่ในรูป $\frac{1}{y}$ และแปลงค่าให้อยู่ในรูป $\frac{1}{y^2}$ จะให้ค่า MAPE ต่ำกว่ากรณีอื่นๆ แต่เมื่อพิจารณาที่ค่า RMSE กลับพบว่าการใช้ตัวแบบเชิงเส้นทั่วไปที่ไม่ได้แปลงค่า y รวมทั้งการใช้ตัวแบบเชิงเส้นนัยทั่วไปจะให้ค่า RMSE ต่ำกว่า และเมื่อพิจารณาขนาดของตัวอย่าง (n) พบว่าไม่มีผลต่อค่า MAPE และ RMSE

พหุน ปณ ฑิโต ชีเว

ตาราง 4.5 ค่าเฉลี่ยของค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อน (MAPE) และรากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE) จำแนกตามขนาดตัวอย่าง ชนิดของตัวแบบและรูปแบบการแปลงค่าของ y เมื่อกำหนดให้ y มีการแจกแจงแกมมา (2,1) กรณีความสัมพันธ์ระหว่าง x และ y อยู่ในระดับสูง

ค่าสัมประสิทธิ์สหสัมพันธ์	n	ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	MAPE	RMSE
$r_{x_2y} \geq 0.7$	30	ตัวแบบเชิงเส้นทั่วไป	y	-	70.6733	0.9704
			$\frac{1}{y}$	-	**51.2435**	1.3901
			$\frac{1}{y^2}$	-	55.3237	1.5942
			$\log(y)$	-	55.1983	0.9721
			Normal	-	58.7170	0.9431
		ตัวแบบเชิงเส้นนัยทั่วไป	-	*Inverse*	65.0818	0.9749
			-	$\log(\mu)$	63.7475	**0.9324**
	50	ตัวแบบเชิงเส้นทั่วไป	y	-	71.1627	0.9988
			$\frac{1}{y}$	-	**53.3040**	1.6550
			$\frac{1}{y^2}$	-	58.3065	1.7410
			$\log(y)$	-	56.1186	1.0048
			Normal	-	59.4238	**0.9702**
		ตัวแบบเชิงเส้นนัยทั่วไป	-	*Inverse*	70.8588	1.1346
			-	$\log(\mu)$	65.5852	**0.9700**
	100	ตัวแบบเชิงเส้นทั่วไป	y	-	70.1850	0.9910
			$\frac{1}{y}$	-	NA	NA
			$\frac{1}{y^2}$	-	NA	NA
			$\log(y)$	-	**57.9167**	0.9986
			Normal	-	60.5326	**0.9681**
		ตัวแบบเชิงเส้นนัยทั่วไป	-	*Inverse*	78.4225	1.2211
			-	$\log(\mu)$	68.5056	**0.9685**

Normal : แปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ

NA : ไม่สามารถคำนวณค่า MAPE และ RMSE ได้ เนื่องจากค่าพยากรณ์มีค่าเป็นลบ

จากตาราง 4.5 พบว่ากรณีที่ตัวแปรอิสระ X และตัวแปรตาม Y มีสัมประสิทธิ์สหสัมพันธ์ในระดับสูง เมื่อพิจารณาค่า MAPE พบว่าตัวแบบเชิงเส้นทั่วไป โดยการแปลงค่า y ให้อยู่ในรูป $\frac{1}{y}$ จะให้ค่า MAPE ต่ำกว่ากรณีอื่นๆ ยกเว้น $n=100$ ตัวแบบเชิงเส้นทั่วไปที่ให้ค่า MAPE ต่ำคือการแปลงค่า y ให้อยู่ในรูป $\log(y)$ แต่เมื่อพิจารณาที่ค่า RMSE กลับพบว่าการใช้ตัวแบบเชิงเส้นทั่วไปที่แปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ รวมทั้งการใช้ตัวแบบเชิงเส้นนัยทั่วไปที่มีฟังก์ชันเชื่อมโยงเป็น “log link” จะมีค่า RMSE ที่ต่ำกว่าและเมื่อพิจารณาขนาดของตัวอย่าง (n) พบว่าไม่มีผลต่อค่า MAPE และ RMSE

พหุน ปณฺฑิโต ชีเว

ตาราง 4.6 ค่าเฉลี่ยของค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อน (MAPE) และรากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE) จำแนกตามขนาดตัวอย่าง ชนิดของตัวแบบและรูปแบบการแปลงค่าของ y เมื่อกำหนดให้ y มีการแจกแจงแกมมา (2,1) กรณีมีตัวแปรอิสระ 2 ตัวที่มีระดับความสัมพันธ์อยู่ในระดับต่ำ

ค่าสัมประสิทธิ์สหสัมพันธ์	n	ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	MAPE	RMSE
$r_{x_1y} \le 0.3$ $r_{x_2y} \ge 0.7$ $r_{x_1x_2} \le 0.3$	30	ตัวแบบเชิงเส้นทั่วไป	y	-	68.3301	0.9719
			$\frac{1}{y}$	-	**50.6031**	1.4544
			$\frac{1}{y^2}$	-	**53.1243**	1.6076
			$\log(y)$	-	**52.9902**	0.9852
			Normal	-	56.4613	**0.9501**
		ตัวแบบเชิงเส้นนัยทั่วไป	-	*Inverse*	63.1518	0.9646
			-	$\log(\mu)$	61.8396	**0.9502**
	50	ตัวแบบเชิงเส้นทั่วไป	y	-	68.7375	1.0005
			$\frac{1}{y}$	-	**51.9836**	1.6452
			$\frac{1}{y^2}$	-	**50.2350**	1.5573
			$\log(y)$	-	**53.7943**	1.0115
			Normal	-	57.0604	**0.9749**
		ตัวแบบเชิงเส้นนัยทั่วไป	-	*Inverse*	66.9328	1.1086
			-	$\log(\mu)$	62.6220	**0.9803**
	100	ตัวแบบเชิงเส้นทั่วไป	y	-	72.2949	0.9421
			$\frac{1}{y}$	-	102.0471	8.5025
			$\frac{1}{y^2}$	-	NA	NA
			$\log(y)$	-	**59.5574**	0.9612
			Normal	-	**62.6174**	**0.9340**
		ตัวแบบเชิงเส้นนัยทั่วไป	-	*Inverse*	77.9443	1.0032
			-	$\log(\mu)$	71.1149	**0.9287**

Normal : แปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ

NA : ไม่สามารถคำนวณค่า MAPE และ RMSE ได้ เนื่องจากค่าพยากรณ์มีค่าเป็นลบ

จากตาราง 4.6 พบว่ากรณีที่ตัวแปรอิสระ X_1 และ X_2 มีสัมประสิทธิ์สหสัมพันธ์ในระดับต่ำ ส่วน X_1 กับ Y มีความสัมพันธ์ในระดับต่ำ และ X_2 กับ Y มีความสัมพันธ์ในระดับสูง เมื่อพิจารณาค่า MAPE พบว่าตัวแบบเชิงเส้นทั่วไป โดยการแปลงค่า y ให้อยู่ในรูป $\frac{1}{y}$ การแปลงค่า y ให้อยู่ในรูป $\frac{1}{y^2}$ และแปลงค่า y ให้อยู่ในรูป $\log(y)$ จะให้ค่า MAPE ต่ำกว่ากรณีอื่นๆ ยกเว้น $n=100$ ตัวแบบเชิงเส้นทั่วไปที่ให้ค่า MAPE ต่ำคือการแปลงค่า y ให้อยู่ในรูป $\log(y)$ และแปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ แต่เมื่อพิจารณาที่ค่า RMSE กลับพบว่าการใช้ตัวแบบเชิงเส้นทั่วไปที่แปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ รวมทั้งการใช้ตัวแบบเชิงเส้นนัยทั่วไปที่มีฟังก์ชันเชื่อมโยงเป็น “log link” จะมีค่า RMSE ที่ต่ำกว่า และเมื่อพิจารณาขนาดของตัวอย่าง (n) พบว่าไม่มีผลต่อค่า MAPE และ RMSE

พหุน ปณฺฑิโต ชีเว

ตาราง 4.7 ค่าเฉลี่ยของค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อน (MAPE) และรากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE) จำแนกตามขนาดตัวอย่าง ชนิดของตัวแบบและรูปแบบการแปลงค่าของ y เมื่อกำหนดให้ y มีการแจกแจงแกมมา (2,1) กรณีมีตัวแปรอิสระ (x) 2 ตัวที่มีระดับความสัมพันธ์อยู่ในระดับสูง

ค่าสัมประสิทธิ์สหสัมพันธ์	n	ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	MAPE	RMSE
$r_{x_1y} \le 0.3$ $r_{x_2y} \ge 0.7$ $r_{x_1x_2} \ge 0.7$	30	ตัวแบบเชิงเส้นทั่วไป	y	-	63.3690	**0.8833**
			$\frac{1}{y}$	-	**49.9198**	1.5637
			$\frac{1}{y^2}$	-	**49.6238**	1.5698
			$\log(y)$	-	**50.6281**	0.9285
			Normal	-	53.0616	**0.8591**
		ตัวแบบเชิงเส้นนัยทั่วไป	-	*Inverse*	63.0602	0.9788
			-	$\log(\mu)$	58.0106	**0.9023**
	50	ตัวแบบเชิงเส้นทั่วไป	y	-	63.7266	**0.9170**
			$\frac{1}{y}$	-	**49.2783**	1.7304
			$\frac{1}{y^2}$	-	52.5836	1.7960
			$\log(y)$	-	**50.1155**	**0.9343**
			Normal	-	52.7262	**0.8728**
		ตัวแบบเชิงเส้นนัยทั่วไป	-	*Inverse*	65.4546	1.1387
			-	$\log(\mu)$	57.6796	**0.9085**
	100	ตัวแบบเชิงเส้นทั่วไป	y	-	65.3649	**0.9334**
			$\frac{1}{y}$	-	**51.7010**	1.8637
			$\frac{1}{y^2}$	-	NA	NA
			$\log(y)$	-	**51.5313**	0.9908
			Normal	-	53.8900	**0.8971**
		ตัวแบบเชิงเส้นนัยทั่วไป	-	*Inverse*	70.1563	1.3605
			-	$\log(\mu)$	58.8861	0.9753

Normal : แปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ

NA : ไม่สามารถคำนวณค่า MAPE และ RMSE ได้ เนื่องจากค่าพยากรณ์มีค่าเป็นลบ

จากตาราง 4.7 พบว่ากรณีที่ตัวแปรอิสระ X_1 และ X_2 มีสัมประสิทธิ์สหสัมพันธ์ในระดับสูง ส่วน X_1 กับ Y มีความสัมพันธ์ในระดับต่ำ และ X_2 กับ Y มีความสัมพันธ์ในระดับสูง เมื่อพิจารณาค่า MAPE พบว่าตัวแบบเชิงเส้นทั่วไป โดยการแปลงค่า y ให้อยู่ในรูป $\frac{1}{y}$ และแปลงค่า y ให้อยู่ในรูป $\log(y)$ จะให้ค่า MAPE ต่ำกว่ากรณีอื่นๆ ยกเว้น $n=30$ ตัวแบบเชิงเส้นทั่วไปที่ให้ค่า MAPE ต่ำคือการแปลงค่า y ให้อยู่ในรูป $\frac{1}{y}$ การแปลงค่า y ให้อยู่ในรูป $\frac{1}{y^2}$ และแปลงค่า y ให้อยู่ในรูป $\log(y)$ แต่เมื่อพิจารณาที่ค่า RMSE กลับพบว่าการใช้ตัวแบบเชิงเส้นทั่วไปที่ยังไม่แปลงค่า y และแปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ รวมทั้งการใช้ตัวแบบเชิงเส้นนัยทั่วไปที่มีฟังก์ชันเชื่อมโยงเป็น “log link” จะมีค่า RMSE ที่ต่ำกว่า และเมื่อพิจารณาขนาดของตัวอย่าง (n) พบว่าไม่มีผลต่อค่า MAPE และ RMSE

4.1.2 กรณีการจำลองข้อมูลให้ Y มีการแจกแจงอินเวอร์สเกาส์เซียน

ผลการจำลองข้อมูลเมื่อตัวแปรตาม Y มีการแจกแจงอินเวอร์สเกาส์เซียน ที่มีพารามิเตอร์แสดงตำแหน่ง (μ) มีค่าเท่ากับ 1 และพารามิเตอร์รูปร่าง (λ) มีค่าเท่ากับ 3 และกำหนดให้ X มีการแจกแจงปรกติมาตรฐาน ที่มีพารามิเตอร์เป็น (0,1) ค่าความสัมพันธ์ระหว่าง X และ Y กำหนดเป็น 2 ระดับ คือ มีความสัมพันธ์ในระดับต่ำมาก มีค่าสัมประสิทธิ์สหสัมพันธ์เป็นน้อยกว่าหรือเท่ากับ 0.3 และมีความสัมพันธ์กันในระดับสูง คือ มีค่าสัมประสิทธิ์สหสัมพันธ์เป็นมากกว่าหรือเท่ากับ 0.7 ส่วนค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร X_1 กับ X_2 กำหนดเป็นน้อยกว่าหรือเท่ากับ 0.3 และมากกว่าหรือเท่ากับ 0.7 เช่นเดียวกัน กำหนดขนาดตัวอย่าง (n) เป็น 30 50 และ 100 ตามลำดับ

พหุน ปณฺฑิโต ชีเว

ตาราง 4.8 ค่าเฉลี่ยของค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อน (MAPE) และรากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE) จำแนกตามขนาดตัวอย่าง ชนิดของตัวแบบและรูปแบบการแปลงค่าของ y เมื่อกำหนดให้ y มีการแจกแจงอินเวอร์เกาส์เซียน (1,3) กรณีความสัมพันธ์ของ x และ y อยู่ในระดับต่ำ

ค่าสัมประสิทธิ์สหสัมพันธ์	n	ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	MAPE	RMSE
$r_{x_1 y} \le 0.3$	30	ตัวแบบเชิงเส้นทั่วไป	y	-	51.5350	**0.5350**
			$\frac{1}{y}$	-	**39.5511**	0.5836
			$\frac{1}{y^2}$	-	**39.2250**	0.6487
			$\log(y)$	-	43.5254	0.5497
			Normal	-	**40.6203**	0.5812
		ตัวแบบเชิงเส้นนัยทั่วไป	-	$sqrt(\mu)$	51.7701	**0.5397**
			-	$\log(\mu)$	51.7729	**0.5396**
	50	ตัวแบบเชิงเส้นทั่วไป	y	-	51.9298	**0.5439**
			$\frac{1}{y}$	-	**39.7958**	0.5916
			$\frac{1}{y^2}$	-	**39.4026**	0.6609
			$\log(y)$	-	43.9335	0.5584
			Normal	-	**40.9518**	0.5907
		ตัวแบบเชิงเส้นนัยทั่วไป	-	$sqrt(\mu)$	52.0627	**0.5465**
			-	$\log(\mu)$	52.0845	**0.5468**
	100	ตัวแบบเชิงเส้นทั่วไป	y	-	52.1961	**0.5498**
			$\frac{1}{y}$	-	**40.0494**	0.5965
			$\frac{1}{y^2}$	-	**39.1063**	0.6517
			$\log(y)$	-	44.2601	0.5639
			Normal	-	**41.1934**	0.5974
		ตัวแบบเชิงเส้นนัยทั่วไป	-	$sqrt(\mu)$	52.2582	**0.5509**
			-	$\log(\mu)$	52.2670	**0.5510**

Normal : แปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ

จากตาราง 4.8 พบว่ากรณีที่ตัวแปรอิสระ X และตัวแปรตาม Y มีสัมประสิทธิ์สหสัมพันธ์ในระดับต่ำ เมื่อพิจารณาค่า MAPE พบว่าตัวแบบเชิงเส้นทั่วไป โดยการแปลงค่า y ให้อยู่ในรูป $\frac{1}{y}$ การแปลงค่า y ให้อยู่ในรูป $\frac{1}{y^2}$ และแปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ จะให้ค่า MAPE ต่ำกว่ากรณีอื่นๆ แต่เมื่อพิจารณาที่ค่า RMSE กลับพบว่าการใช้ตัวแบบเชิงเส้นทั่วไปที่ยังไม่แปลงค่า y รวมทั้งการใช้ตัวแบบเชิงเส้นนัยทั่วไปจะมีค่า RMSE ที่ต่ำกว่า และเมื่อพิจารณาขนาดของตัวอย่าง (n) พบว่าไม่มีผลต่อค่า MAPE และ RMSE

พหุน ปณ ฑิโต ชีเว

ตาราง 4.9 ค่าเฉลี่ยของค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อน (MAPE) และรากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE) จำแนกตามขนาดตัวอย่าง ชนิดของตัวแบบและรูปแบบการแปลงค่าของ y เมื่อกำหนดให้ y มีการแจกแจงอินเวอร์เกาส์เซียน (1,3) กรณีความสัมพันธ์ของ x และ y อยู่ในระดับสูง

ค่าสัมประสิทธิ์สหสัมพันธ์	n	ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	MAPE	RMSE
$r_{x_2y} \geq 0.7$	30	ตัวแบบเชิงเส้นทั่วไป	y	-	38.1073	**0.3814**
			$\frac{1}{y}$	-	**32.2567**	0.4724
			$\frac{1}{y^2}$	-	**32.7793**	0.6875
			$\log(y)$	-	**32.5714**	**0.3706**
			Normal	-	34.4257	0.4930
		ตัวแบบเชิงเส้นนัยทั่วไป	-	$sqrt(\mu)$	36.0137	0.3865
			-	$\log(\mu)$	35.1753	**0.3702**
	50	ตัวแบบเชิงเส้นทั่วไป	y	-	38.1716	**0.3894**
			$\frac{1}{y}$	-	**32.8842**	0.5274
			$\frac{1}{y^2}$	-	34.2893	0.7527
			$\log(y)$	-	**32.8103**	**0.3793**
			Normal	-	34.5998	0.4989
		ตัวแบบเชิงเส้นนัยทั่วไป	-	$sqrt(\mu)$	36.2012	0.3924
			-	$\log(\mu)$	35.4438	**0.3772**
	100	ตัวแบบเชิงเส้นทั่วไป	y	-	38.3021	0.3973
			$\frac{1}{y}$	-	**33.3118**	0.6242
			$\frac{1}{y^2}$	-	NA	NA
			$\log(y)$	-	**32.7985**	**0.3868**
			Normal	-	34.5932	0.5052
		ตัวแบบเชิงเส้นนัยทั่วไป	-	$sqrt(\mu)$	36.2391	0.3989
			-	$\log(\mu)$	35.4986	**0.3831**

Normal : แปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ

NA : ไม่สามารถคำนวณค่า MAPE และ RMSE ได้ เนื่องจากค่าพยากรณ์มีค่าเป็นลบ

จากตาราง 4.9 พบว่ากรณีที่ตัวแปรอิสระ X และตัวแปรตาม Y มีสัมประสิทธิ์สหสัมพันธ์ในระดับสูง เมื่อพิจารณาค่า MAPE พบว่าตัวแบบเชิงเส้นทั่วไป โดยการแปลงค่า y ให้อยู่ในรูป $\frac{1}{y}$ และแปลงค่า y ให้อยู่ในรูป $\log(y)$ จะให้ค่า MAPE ต่ำกว่ากรณีอื่นๆ แต่เมื่อพิจารณาที่ค่า RMSE กลับพบว่าการใช้ตัวแบบเชิงเส้นทั่วไปที่ยังไม่แปลงค่า y และแปลงค่า y ให้อยู่ในรูป $\log(y)$ รวมทั้งการใช้ตัวแบบเชิงเส้นนัยทั่วไปที่มีฟังก์ชันเชื่อมโยงเป็น “log link” จะมีค่า RMSE ที่ต่ำกว่า และเมื่อพิจารณาขนาดของตัวอย่าง (n) พบว่าไม่มีผลต่อค่า MAPE และ RMSE

พหูน ปณ ฑิโต ชีเว

ตาราง 4.10 ค่าเฉลี่ยของค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อน (MAPE) และรากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE) จำแนกตามขนาดตัวอย่าง ชนิดของตัวแบบและรูปแบบการแปลงค่าของ y เมื่อกำหนดให้ y มีการแจกแจงอินเวอร์เกาส์เซียน (1,3) กรณีมีตัวแปรอิสระ (x) 2 ตัวที่มีระดับความสัมพันธ์อยู่ในระดับต่ำ

ค่าสัมประสิทธิ์สหสัมพันธ์	n	ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	MAPE	RMSE
$r_{x_1y} \le 0.3$ $r_{x_2y} \ge 0.7$ $r_{x_1x_2} \le 0.3$	30	ตัวแบบเชิงเส้นทั่วไป	y	-	37.9501	0.3804
			$\frac{1}{y}$	-	**32.1397**	0.4719
			$\frac{1}{y^2}$	-	33.7571	0.7000
			$\log(y)$	-	**32.3176**	**0.3721**
			Normal	-	**32.1800**	**0.3715**
		ตัวแบบเชิงเส้นนัยทั่วไป	-	$sqrt(\mu)$	35.7249	0.3882
			-	$\log(\mu)$	35.0208	**0.3751**
	50	ตัวแบบเชิงเส้นทั่วไป	y	-	37.9930	0.3883
			$\frac{1}{y}$	-	**32.6136**	0.5250
			$\frac{1}{y^2}$	-	**31.9058**	0.6147
			$\log(y)$	-	**32.3738**	**0.3792**
			Normal	-	**32.2323**	**0.3787**
		ตัวแบบเชิงเส้นนัยทั่วไป	-	$sqrt(\mu)$	35.8014	0.3935
			-	$\log(\mu)$	35.0902	**0.3796**
	100	ตัวแบบเชิงเส้นทั่วไป	y	-	37.9144	0.3957
			$\frac{1}{y}$	-	**33.1957**	0.6424
			$\frac{1}{y^2}$	-	NA	NA
			$\log(y)$	-	**32.4270**	**0.3854**
			Normal	-	**32.2832**	**0.3849**
		ตัวแบบเชิงเส้นนัยทั่วไป	-	$sqrt(\mu)$	35.8726	0.3983
			-	$\log(\mu)$	35.1317	**0.3830**

Normal : แปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ

NA : ไม่สามารถคำนวณค่า MAPE และ RMSE ได้ เนื่องจากค่าพยากรณ์มีค่าเป็นลบ

จากตาราง 4.10 พบว่ากรณีที่ตัวแปรอิสระ X_1 และ X_2 มีสัมประสิทธิ์สหสัมพันธ์ในระดับต่ำ ส่วน X_1 กับ Y มีความสัมพันธ์ในระดับต่ำ และ X_2 กับ Y มีความสัมพันธ์ในระดับสูง เมื่อพิจารณาค่า MAPE และ RMSE พบว่าตัวแบบเชิงเส้นทั่วไป โดยการแปลงค่า y ให้อยู่ในรูป $\frac{1}{y}$ การแปลงค่า y ให้อยู่ในรูป $\log(y)$ และแปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ จะให้ค่า MAPE และ RMSE ต่ำกว่ากรณีอื่นๆ แต่เมื่อพิจารณาที่ค่า RMSE การใช้ตัวแบบเชิงเส้นนัยทั่วไปที่มีฟังก์ชันเชื่อมโยงเป็น “log link” ก็ให้ค่า RMSE ที่ต่ำเช่นเดียวกัน และเมื่อพิจารณาขนาดของตัวอย่าง (n) พบว่าไม่มีผลต่อค่า MAPE และ RMSE

ตาราง 4.11 ค่าเฉลี่ยของค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อน (MAPE) และรากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE) จำแนกตามขนาดตัวอย่าง ชนิดของตัวแบบและรูปแบบการแปลงค่าของ y เมื่อกำหนดให้ y มีการแจกแจงอินเวอร์เกาส์เซียน (1,3) กรณีมีตัวแปรอิสระ (x) 2 ตัวที่มีระดับความสัมพันธ์อยู่ในระดับสูง

ค่าสัมประสิทธิ์สหสัมพันธ์	n	ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	MAPE	RMSE
$r_{x_1y} \le 0.3$ $r_{x_2y} \ge 0.7$ $r_{x_1x_2} \ge 0.7$	30	ตัวแบบเชิงเส้นทั่วไป	y	-	35.4120	0.3463
			$\frac{1}{y}$	-	**30.9953**	0.4963
			$\frac{1}{y^2}$	-	32.0941	0.6877
			$\log(y)$	-	**29.9735**	**0.3314**
			Normal	-	**29.8229**	**0.3308**
		ตัวแบบเชิงเส้นนัยทั่วไป	-	$sqrt(\mu)$	33.4023	0.3574
			-	$\log(\mu)$	32.2539	**0.3385**
	50	ตัวแบบเชิงเส้นทั่วไป	y	-	35.2652	0.3552
			$\frac{1}{y}$	-	31.6984	0.5860
			$\frac{1}{y^2}$	-	**30.3803**	0.5968
			$\log(y)$	-	**30.0489**	**0.3423**
			Normal	-	**29.8978**	**0.3416**
		ตัวแบบเชิงเส้นนัยทั่วไป	-	$sqrt(\mu)$	33.1881	0.3605
			-	$\log(\mu)$	32.2927	**0.3463**
	100	ตัวแบบเชิงเส้นทั่วไป	y	-	35.4012	0.3630
			$\frac{1}{y}$	-	32.3472	0.7235
			$\frac{1}{y^2}$	-	NA	NA
			$\log(y)$	-	**30.0981**	**0.3484**
			Normal	-	**29.9469**	**0.3474**
		ตัวแบบเชิงเส้นนัยทั่วไป	-	$sqrt(\mu)$	33.1696	0.3654
			-	$\log(\mu)$	32.2952	**0.3489**

Normal : แปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ

NA : ไม่สามารถคำนวณค่า MAPE และ RMSE ได้ เนื่องจากค่าพยากรณ์มีค่าเป็นลบ

จากตาราง 4.11 พบว่ากรณีที่ตัวแปรอิสระ X_1 และ X_2 มีสัมประสิทธิ์สหสัมพันธ์ในระดับสูง ส่วน X_1 กับ Y มีความสัมพันธ์ในระดับต่ำ และ X_2 กับ Y มีความสัมพันธ์ในระดับสูง เมื่อพิจารณาค่า MAPE และ RMSE พบว่าตัวแบบเชิงเส้นทั่วไป โดยการแปลงค่า y ให้อยู่ในรูป $\log(y)$ และแปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ จะให้ค่า MAPE และ RMSE ต่ำกว่ากรณีอื่นๆ แต่เมื่อพิจารณาที่ค่า RMSE การใช้ตัวแบบเชิงเส้นนัยทั่วไปที่มีฟังก์ชันเชื่อมโยงเป็น “log link” ก็ให้ค่า RMSE ที่ต่ำเช่นเดียวกัน และเมื่อพิจารณาขนาดของตัวอย่าง (n) พบว่าไม่มีผลต่อค่า MAPE และ RMSE

4.2 ผลการวิเคราะห์หาสมการที่เหมาะสมในการพยากรณ์ปริมาณการใช้น้ำมันสำเร็จรูปและหาปัจจัยที่ส่งผลต่อปริมาณการใช้น้ำมันสำเร็จรูปของประเทศไทย

โดยใช้วิธีการวิเคราะห์ตัวแบบเชิงเส้นทั่วไป (General Linear Model : GLM) กับตัวแบบเชิงเส้นนัยทั่วไป (Generalized Linear Model : GLMs) ด้วยวิธีการคัดเลือกตัวแปรแบบขั้นตอน (Stepwise Selection)

4.2.1 ผลการวิเคราะห์หาสมการที่เหมาะสมในการพยากรณ์**ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91** และหาปัจจัยที่ส่งผลต่อการใช้ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91 ซึ่งตัวแปรอิสระในการนำมาวิเคราะห์ประกอบด้วย

P_G91 คือ ราคาน้ำมันแก๊สโซฮอล์ 91 G95 คือ ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95
P_G95 คือ ราคาน้ำมันแก๊สโซฮอล์ 95 B95 คือ ปริมาณการใช้น้ำมันเบนซิน 95
P_E20 คือ ราคาน้ำมันแก๊สโซฮอล์ E20 E20 คือ ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20
P_B95 คือ ราคาน้ำมันเบนซิน 95 WCO คือ ราคาน้ำมันดิบโลก
P_Ethanol คือ ราคาเอทานอล EXC คือ อัตราแลกเปลี่ยนเงินตรา

พหุน ปณฺฑิโต ชีเว

ตาราง 4.12 สมการพยากรณ์*ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91* และค่าสถิติแสดงความคลาดเคลื่อนและความเหมาะสมของตัวแบบพยากรณ์

ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	ผลการทดสอบการแจกแจง	สมการพยากรณ์	จำนวนตัวแปรอิสระ	MAPE	RMSE	Adjusted R-squared	Residual deviance
ตัวแบบเชิงเส้นทั่วไป	y	-	ไม่ได้มีการแจกแจงปรกติ (p-value ≤ 0.000)	$\hat{G91} = 325.72772 - 1.84787(P_Ethanol) - 6.32264(EXC) + 3.14346(P_G95) - 2.68144(P_B95) + 0.20289(G95) + 1.74737(E20)$	6	**3.9054**	**9.2026**	0.9853	-
	$\frac{1}{y}$	-	ไม่ได้มีการแจกแจงปรกติ (p-value ≤ 0.000)	$\frac{1}{\hat{G91}} = -0.001135 + 0.00006939(P_Ethanol) + 0.00027723(EXC) - 0.00003789(P_G95) - 0.0004610(E20)$	4	7.7743	30.0454	0.9464	-
	$\log(y)$	-	ไม่ได้มีการแจกแจงปรกติ (p-value ≤ 0.000)	$\log(\hat{G91}) = 6.456456 - 0.010954(P_Ethanol) - 0.043623(EXC) + 0.008254(E20) + 0.001947(B95)$	4	4.9754	13.4919	0.9720	-

ตาราง 4.12 สมการพยากรณ์*ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91* และค่าสถิติแสดงความคลาดเคลื่อนและความเหมาะสมของตัวแบบพยากรณ์ (ต่อ)

ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	ผลการทดสอบการแจกแจง	สมการพยากรณ์	จำนวนตัวแปรอิสระ	MAPE	RMSE	Adjusted R-squared	Residual deviance
ตัวแบบเชิงเส้นนัยทั่วไปที่มีการแจกแจงแกมมา	-	*Inverse*	ไม่ได้มีการแจกแจงแกมมา (p-value = 0.011)	$\frac{1}{\hat{\mu}_{G91}} = 0.002301 + 0.0001581(EXC) - 0.00001774(B95) - 0.00003113(E20)$	3	7.0297	16.8486	-	0.4890
	-	$\log(\mu)$		$\log(\hat{\mu}_{G91}) = 6.635289 - 0.009487(P_Ethanol) - 0.050498(EXC) + 0.009104(E20)$	3	5.3342	15.4748	-	0.2762
ตัวแบบเชิงเส้นนัยทั่วไปที่มีการแจกแจงอินเวอร์สเกาส์เซียน	-	$\frac{1}{\mu^2}$	มีการแจกแจงอินเวอร์สเกาส์เซียน (p-value = 0.221)	$\frac{1}{\hat{\mu}^2_{G91}} = 0.000001827 - 0.000003156(P_G91) + 0.000001498(EXC) + 0.000003281(P_E20) - 0.0000001367(B95) - 0.0000002837(E20)$	5	8.4309	24.1364	-	0.0036
	-	$\log(\mu)$		$\log(\hat{\mu}_{G91}) = 6.223332 - 0.009204(P_Ethanol) - 0.044261(EXC) + 0.005703(P_G95) + 0.009202(E20)$	4	5.2355	17.1232	-	0.0012
	-	$sqrt(\mu)$		$sqrt(\hat{\mu}_{G91}) = 18.72161 - 0.04490(P_Ethanol) - 0.25964(EXC) + 0.03992(P_G95) + 0.06563(E20)$	4	4.4958	14.0266	-	0.0009

จากตาราง 4.12 สมการที่ใช้ในการพยากรณ์*ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91* เมื่อพิจารณาค่า MAPE พบว่า ในการใช้ตัวแบบเชิงเส้นทั่วไปและตัวแบบเชิงเส้นนัยทั่วไป เป็นตัวแบบที่มีการพยากรณ์มีความแม่นสูง (MAPE ≤ 10) แต่ตัวแบบที่มีค่า MAPE ต่ำที่สุดคือ ตัวแบบเชิงเส้นทั่วไปที่ไม่ได้แปลงค่า y เช่นเดียวกับการพิจารณาค่า RMSE การใช้ตัวแบบเชิงเส้นทั่วไปที่ไม่ได้แปลงค่า y ก็จะให้ค่า RMSE ต่ำที่สุด ดังนั้นตัวแปรที่ส่งผลต่อปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91 ในทิศทางเดียวกัน คือ ราคาน้ำมันแก๊สโซฮอล์ 95 ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95 และปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20 นั้นหมายความว่า ถ้าราคาน้ำมันแก๊สโซฮอล์ 95 ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95 และปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20 เพิ่มขึ้น จะส่งผลให้ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91 เพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติ ส่วนตัวแปรที่ส่งผลในทิศทางตรงกันข้าม คือ ราคาเอทานอล อัตราแลกเปลี่ยนเงินตรา และราคาน้ำมันเบนซิน 95 นั้นหมายความว่า ถ้าราคาเอทานอล อัตราแลกเปลี่ยนเงินตรา และราคาน้ำมันเบนซิน 95 ลดลง ส่งผลให้ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91 เพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติ

4.2.2 ผลการวิเคราะห์หาสมการที่เหมาะสมในการพยากรณ์**ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95** และหาปัจจัยที่ส่งผลต่อการใช้ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95 ซึ่งตัวแปรอิสระในการนำมาวิเคราะห์ประกอบด้วย

P_G95	คือ ราคาน้ำมันแก๊สโซฮอล์ 95	B95	คือ ปริมาณการใช้น้ำมันเบนซิน 95
P_B95	คือ ราคาน้ำมันเบนซิน 95	E20	คือ ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20
P_E20	คือ ราคาน้ำมันแก๊สโซฮอล์ E20	E85	คือ ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E85
P_E85	คือ ราคาน้ำมันแก๊สโซฮอล์ E85	WCO	คือ ราคาน้ำมันดิบโลก
P_Ethanol	คือ ราคาเอทานอล	EXC	คือ อัตราแลกเปลี่ยนเงินตรา

พหุน ปณ ฑิโต ชีเว

ตาราง 4.13 สมการพยากรณ์*ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95* และค่าสถิติแสดงความคลาดเคลื่อนและความเหมาะสมของตัวแบบพยากรณ์

ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	ผลการทดสอบการแจกแจง	สมการพยากรณ์	จำนวนตัวแปรอิสระ	MAPE	RMSE	Adjusted R-squared	Residual deviance
ตัวแบบเชิงเส้นทั่วไป	y	-	ไม่ได้มีการแจกแจงปรกติ (p-value ≤ 0.000)	$\hat{G95} = 192.5080 - 26.6521(P_G95) + 1.8926(P_Ethanol) + 10.6764(P_E20) + 12.3691(P_B95) - 0.9674(E85) + 2.3789(B95)$	6	**4.4268**	**11.8153**	0.9207	-
	$\frac{1}{y}$	-	ไม่ได้มีการแจกแจงปรกติ (p-value ≤ 0.000)	$\frac{1}{\hat{G95}} = 0.006676 + 0.0008873(P_G95) - 0.00006773(P_Ethanol) - 0.0004580(P_E20) + 0.0001537(P_E85) - 0.0004376(P_B95) - 0.00005940(B95)$	6	5.3989	14.5577	0.8715	-
	$\frac{1}{y^2}$	-	ไม่ได้มีการแจกแจงปรกติ (p-value ≤ 0.000)	$\frac{1}{\hat{G95}} = 0.00006146 + 0.00001138(P_G95) - 0.0000007668(P_Ethanol) - 0.000005966(P_E20) + 0.000001503(P_E85) - 0.000005717(P_B95) + 0.0000006638(B95)$	6	6.8646	21.3591	0.8215	-
	$\log(y)$	-	ไม่ได้มีการแจกแจงปรกติ (p-value ≤ 0.000)	$\log(\hat{G95}) = 5.14408 - 0.15456(P_G95) + 0.06597(P_E20) + 0.07672(P_B95) + 0.01087(B95)$	4	5.1005	13.9649	0.8817	-

ตาราง 4.13 สมการพยากรณ์*ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95* และค่าสถิติแสดงความคลาดเคลื่อนและความเหมาะสมของตัวแบบพยากรณ์ (ต่อ)

ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	ผลการทดสอบการแจกแจง	สมการพยากรณ์	จำนวนตัวแปรอิสระ	MAPE	RMSE	Adjusted R-squared	Residual deviance
ตัวแบบเชิงเส้นทั่วไป	$q_normal(G95)$	-	มีการแจกแจงปรกติ (p-value = 0.0493)	$\hat{G95} = 188.9389 - 29.5048(P_G95) + 12.2857(P_E20) + 14.5372(P_B95) + 2.1628(B95)$	4	5.1300	13.9900	0.8888	-
ตัวแบบเชิงเส้นทั่วไปที่มีการแจกแจงแกมมา	-	*Inverse*	มีการแจกแจงแกมมา (p-value = 0.0493)	$\frac{1}{\hat{\mu}_{G95}} = 0.007127 + 0.0008436(P_G95) - 0.00005046(P_Ethanol) - 0.0003440(P_E20) - 0.0004365(P_B95) + 0.000007141(E20) - 0.00006338(B95)$	6	4.9933	13.4410	-	0.3371
	-	$\log(\mu)$		$\log(\hat{\mu}_{G95}) = 5.15980 - 0.15241(P_G95) + 0.06385(P_E20) + 0.07630(P_B95) + 0.01072(B95)$	4	5.0783	13.9446	-	0.3800

$q_normal(G95)$: แปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ

ตาราง 4.13 สมการพยากรณ์*ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95* และค่าสถิติแสดงความคลาดเคลื่อนและความเหมาะสมของตัวแบบพยากรณ์ (ต่อ)

ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	ผลการทดสอบการแจกแจง	สมการพยากรณ์	จำนวนตัวแปรอิสระ	MAPE	RMSE	Adjusted R-squared	Residual deviance
ตัวแบบเชิงเส้นนัยทั่วไปที่มีการแจกแจงอินเวอร์เกาส์เซียน	-	$\frac{1}{\mu^2}$	ไม่ได้มีการแจกแจงอินเวอร์เกาส์เซียน (p-value = 0.0037)	$\frac{1}{\hat{\mu}^2_{G95}} = 0.00005908 + 0.000009716(P_G95) - 0.0000005554(P_Ethanol) - 0.000004257(P_E20) - 0.000005040(P_B95) + 0.00000006790(E20) - 0.0000006466(B95)$	6	5.7628	15.6619	-	0.0023
	-	$\log(\mu)$		$\log(\hat{\mu}_{G95}) = 5.18123 - 0.15262(P_G95) + 0.06385(P_E20) + 0.07594(P_B95) + 0.01079(B95)$	4	5.0998	14.0356	-	0.0021
	-	$sqrt(\mu)$		$sqrt(\hat{\mu}_{G95}) = 12.39905 - 1.00234(P_G95) + 0.09235(P_Ethanol) + 0.42464(P_E20) + 0.46962(P_B95) - 0.03582(E85) + 0.08445(B95)$	6	4.4946	12.1712	-	0.0017

จากตาราง 4.13 สมการที่ใช้ในพยากรณ์ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95 เมื่อพิจารณาค่า MAPE พบว่า ในการใช้ตัวแบบเชิงเส้นทั่วไปและตัวแบบเชิงเส้นนัยทั่วไป เป็นตัวแบบที่มีการพยากรณ์มีความแม่นสูง (MAPE ≤ 10) แต่ตัวแบบที่มีค่า MAPE ต่ำที่สุดคือ ตัวแบบเชิงเส้นทั่วไปที่ไม่ได้แปลงค่า y เช่นเดียวกับการพิจารณาค่า RMSE การใช้ตัวแบบเชิงเส้นทั่วไปที่ไม่ได้แปลงค่า y ก็จะให้ค่า RMSE ต่ำที่สุด โดยตัวแปรที่ส่งผลต่อปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95 ในทิศทางเดียวกัน คือ ราคาเอทานอล ราคาน้ำมันแก๊สโซฮอล์ E20 ราคาน้ำมันเบนซิน 95 และปริมาณการใช้น้ำมันเบนซิน 95 นั้นหมายความว่า ถ้าราคาเอทานอล ราคาน้ำมันแก๊สโซฮอล์ E20 ราคาน้ำมันเบนซิน 95 และปริมาณการใช้น้ำมันเบนซิน 95 เพิ่มขึ้นจะส่งผลให้ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95 เพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติ ส่วนตัวแปรที่ส่งผลในทิศทางตรงกันข้าม คือ ราคาน้ำมันแก๊สโซฮอล์ 95 และปริมาณการใช้น้ำมันแก๊สโซฮอล์ E85 นั้นหมายความว่า ถ้าราคาน้ำมันแก๊สโซฮอล์ 95 และปริมาณการใช้น้ำมันแก๊สโซฮอล์ E85 ลดลง ส่งผลให้ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95 เพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติ

4.2.3 ผลการวิเคราะห์หาสมการที่เหมาะสมในการพยากรณ์**ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20** และหาปัจจัยที่ส่งผลต่อการใช้ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20 ซึ่งตัวแปรอิสระในการนำมาวิเคราะห์ประกอบด้วย **ข้อมูลตั้งแต่มกราคม พ.ศ.2553 ถึง กุมภาพันธ์ พ.ศ.2556**

P_E20 คือ ราคาน้ำมันแก๊สโซฮอล์ E20
P_G95 คือ ราคาน้ำมันแก๊สโซฮอล์ 95
P_B95 คือ ราคาน้ำมันเบนซิน 95
P_G91 คือ ราคาน้ำมันแก๊สโซฮอล์ 91
P_Ethanol คือ ราคาเอทานอล
G95 คือ ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95
B95 คือ ปริมาณการใช้น้ำมันเบนซิน 95
G91 คือ ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91
WCO คือ ราคาน้ำมันดิบโลก
EXC คือ อัตราแลกเปลี่ยนเงินตรา

พหุน ปณ ฑิโต ชีเว

ตาราง 4.14 สมการพยากรณ์*ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20* ข้อมูลตั้งแต่มกราคม พ.ศ.2553 ถึง กุมภาพันธ์ พ.ศ.2556 และค่าสถิติแสดงความคลาดเคลื่อนและความเหมาะสมของตัวแบบพยากรณ์

ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	ผลการทดสอบการแจกแจง	สมการพยากรณ์	จำนวนตัวแปรอิสระ	MAPE	RMSE	Adjusted R-squared	Residual deviance
ตัวแบบเชิงเส้นทั่วไป	y	-	ไม่ได้มีการแจกแจงปรกติ (p-value ≤ 0.000)	$\hat{E20} = -28.1531 - 7.0077(P_E20) + 6.8494(P_G91) + 0.2953(G91)$	3	8.6979	2.414	0.9624	-
	$\frac{1}{y}$	-	ไม่ได้มีการแจกแจงปรกติ (p-value ≤ 0.000)	$\frac{1}{\hat{E20}} = 0.1851 - 0.001190(P_Ethanol) - 0.0008841(G91) + 0.0001834(G95) + 0.0007798(B95)$	4	7.553	9.002	0.9848	-
	$\log(y)$	-	ไม่ได้มีการแจกแจงปรกติ (p-value ≤ 0.000)	$\log(\hat{E20}) = 1.474049 - 0.056592(P_E20) + 0.041112(P_B95) + 0.014565(G91) - 0.004212(G95)$	4	**4.7446**	1.5704	0.9859	-
	$q_normal(E20)$	-	มีการแจกแจงปรกติ (p-value ≤ 0.3758)	$\hat{E20} = -20.06698 - 4.14194(P_E20) + 3.01382(P_G91) + 0.82210(P_B95) + 0.34178(G91) - 0.08202(G95)$	5	5.0800	1.6000	0.9823	-

$q_normal(E20)$: แปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ

ตาราง 4.14 สมการพยากรณ์ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20 ข้อมูลตั้งแต่มกราคม พ.ศ.2553 ถึง กุมภาพันธ์ พ.ศ.2556 และค่าสถิติแสดงความคลาดเคลื่อนและความเหมาะสมของตัวแบบพยากรณ์ (ต่อ)

ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	ผลการทดสอบการแจกแจง	สมการพยากรณ์	จำนวนตัวแปรอิสระ	MAPE	RMSE	Adjusted R-squared	Residual deviance
ตัวแบบเชิงเส้นนัยทั่วไปที่มีการแจกแจงแกมมา	-	*Inverse*	มีการแจกแจงแกมมา (p-value = 0.3758)	$\frac{1}{\hat{\mu}_{E20}} = 0.1526 - 0.001220(P_Ethanol) + 0.002620(P_G95) - 0.002763(P_B95) - 0.0006810(G91) + 0.0003625(G95)$	5	**4.788**	**1.4188**	-	0.1374
	-	$\log(\mu)$		$\log(\hat{\mu}_{E20}) = 1.4732632 - 0.0566789(P_E20) + 0.0413128(P_B95) + 0.0145635(G91) - 0.0042306(G95)$	4	**4.7615**	1.5582	-	0.1184
ตัวแบบเชิงเส้นนัยทั่วไปที่มีการแจกแจงอินเวอร์เกาส์เซียน	-	$\log(\mu)$	มีการแจกแจงอินเวอร์เกาส์เซียน (p-value = 0.2249)	$\log(\hat{\mu}_{G95}) = 1.3947389 - 0.0480811(P_E20) + 0.0340468(P_B95) + 0.0147815(G91) - 0.0037818(G95)$	4	4.8369	1.8981	-	0.0053
	-	$sqrt(\mu)$		$sqrt(\hat{\mu}_{E20}) = 0.597534 - 0.389349(P_E20) + 0.314587(P_G91) + 0.051630(P_B95) + 0.030085(G91) - 0.005748(G95)$	5	5.5655	2.1207	-	0.0079

จากตาราง 4.14 สมการที่ใช้ในพยากรณ์ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20 เมื่อพิจารณาค่า MAPE พบว่า ในการใช้ตัวแบบเชิงเส้นทั่วไปและตัวแบบเชิงเส้นนัยทั่วไปที่ y เป็นตัวแบบที่มีการพยากรณ์มีความแม่นสูง (MAPE ≤ 10) แต่ตัวแบบที่ให้ค่า MAPE ต่ำสุด คือ ตัวแบบเชิงเส้นนัยทั่วไปเมื่อ y มีการแจกแจงแกมมาที่มีฟังก์ชันเชื่อมโยงเป็น inverse เช่นเดียวกับการพิจารณาค่า RMSE การใช้ตัวแบบเชิงเส้นนัยทั่วไปเมื่อ y มีการแจกแจงแกมมาที่มีฟังก์ชันเชื่อมโยงเป็น inverse ก็จะให้ค่า RMSE ต่ำที่สุด โดยตัวแปรที่ส่งผลต่อปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20 ในทิศทางเดียวกัน คือ ราคาเอทานอล ราคาน้ำมันเบนซิน 95 และปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95 นั้นหมายความว่า ถ้าราคาเอทานอล ราคาน้ำมันเบนซิน 95 และปริมาณการใช้ น้ำมันแก๊สโซฮอล์ 95 ลดลง จะส่งผลให้ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20 ลดลงอย่างมีนัยสำคัญทางสถิติ ส่วนตัวแปรที่ส่งผลในทิศทางตรงกันข้าม คือ ราคาน้ำมันแก๊สโซฮอล์ 95 และปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95 หมายความว่า ถ้าราคาน้ำมันแก๊สโซฮอล์ 95 และปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95 เพิ่มขึ้นจะส่งผลให้ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20 ลดลงอย่างมีนัยสำคัญทางสถิติ

หมายเหตุ : ตัวแบบเชิงเส้นนัยทั่วไปที่มีฟังก์ชันเชื่อมโยงเป็น inverse link ถ้าสัมประสิทธิ์เป็นลบหมายถึงมีความสัมพันธ์ในทิศทางเดียวกันกับค่าเฉลี่ยของ y

4.2.4 ผลการวิเคราะห์หาสมการที่เหมาะสมในการพยากรณ์**ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20** และหาปัจจัยที่ส่งผลต่อการใช้ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20 ซึ่งตัวแปรอิสระในการนำมาวิเคราะห์ประกอบด้วย **ข้อมูลตั้งแต่มีนาคม พ.ศ.2556 ถึง ธันวาคม พ.ศ.2558**

P_E20 คือ ราคาน้ำมันแก๊สโซฮอล์ E20 G95 คือ ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95
P_G95 คือ ราคาน้ำมันแก๊สโซฮอล์ 95 B95 คือ ปริมาณการใช้น้ำมันเบนซิน 95
P_B95 คือ ราคาน้ำมันเบนซิน 95 G91 คือ ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91
P_G91 คือ ราคาน้ำมันแก๊สโซฮอล์ 91 WCO คือ ราคาน้ำมันดิบโลก
P_Ethanol คือ ราคาเอทานอล EXC คือ อัตราแลกเปลี่ยนเงินตรา

พหุน ปณฺฑิโต ชีเว

ตาราง 4.15 สมการพยากรณ์*ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20* ข้อมูลตั้งแต่มีนาคม พ.ศ.2556 ถึง ธันวาคม พ.ศ.2558 และค่าสถิติแสดงความคลาดเคลื่อนและความเหมาะสมของตัวแบบพยากรณ์

ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	ผลการทดสอบการแจกแจง	สมการพยากรณ์	จำนวนตัวแปรอิสระ	MAPE	RMSE	Adjusted R-squared	Residual deviance
ตัวแบบเชิงเส้นทั่วไป	y	-	ไม่ได้มีการแจกแจงปรกติ (p-value ≤ 0.000)	$\hat{E20} = -96.60753 - 5.97443(P_E20) + 1.16751(EXC) + 0.38758(WCO) + 4.89662(P_G95) + 0.73108(G91) - 1.42407(B95)$	6	2.0874	2.5678	0.9774	-
	$\frac{1}{y}$	-	ไม่ได้มีการแจกแจงปรกติ (p-value ≤ 0.000)	$\frac{1}{\hat{E20}} = 0.02502 - 0.00004108(WCO) - 0.00006732(G91) + 0.0001644(B95)$	3	2.6349	3.6967	0.9653	-
	$\frac{1}{y^2}$	-	ไม่ได้มีการแจกแจงปรกติ (p-value ≤ 0.000)	$\frac{1}{(\hat{E20})^2} = +0.0005609 - 0.000005265(EXC) - 0.0000007882(WCO) - 0.000001425(G91) + 0.0000003739(G95) + 0.000002032(B95)$	5	2.7188	4.083	0.962	-
	$\log(y)$	-	ไม่ได้มีการแจกแจงปรกติ (p-value ≤ 0.000)	$\log(\hat{E20}) = 3.1622759 - 0.0539528(P_E20) + 0.0045106(WCO) + 0.0424161(P_G95) + 0.0071960(G91) - 0.0167266(B95)$	5	2.0981	2.7868	0.9752	-

ตาราง 4.15 สมการพยากรณ์*ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20* ข้อมูลตั้งแต่มีนาคม พ.ศ.2556 ถึง ธันวาคม พ.ศ.2558 และค่าสถิติแสดงความคลาดเคลื่อนและความเหมาะสมของตัวแบบพยากรณ์ (ต่อ)

ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	ผลการทดสอบการแจกแจง	สมการพยากรณ์	จำนวนตัวแปรอิสระ	MAPE	RMSE	Adjusted R-squared	Residual deviance
ตัวแบบเชิงเส้นทั่วไป	$q_normal(E20)$	-	มีการแจกแจงปรกติ (p-value = 0.3266)	$\hat{E20} = -95.4239 - 5.3113(P_E20) + 1.1539(EXC) + 0.4228(WCO) + 4.3638(P_G95) + 0.7339(G91) - 1.5453(B95)$	6	**1.9300**	2.4800	0.9785	-
ตัวแบบเชิงเส้นนัยทั่วไปที่มีการแจกแจงแกมมา	-	*Inverse*	มีการแจกแจงแกมมา (p-value = 0.3266)	$\frac{1}{\hat{\mu}_{E20}} = 0.02725 + 0.0003815(P_E20) - 0.0001133(EXC) - 0.00004226(WCO) - 0.0003174(P_G95) - 0.00006481(G91) + 0.0001644(B95)$	6	2.0609	2.7903	-	0.0251
	-	$\log(\mu)$		$\log(\hat{\mu}_{E20}) = 2.7422814 - 0.0476477(P_E20) + 0.0109384(EXC) + 0.0042112(WCO) + 0.0391259(P_G95) + 0.0069715(G91) - 0.0152211(B95)$	6	**1.919**	**2.4989**	-	0.0207

ตาราง 4.15 สมการพยากรณ์*ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20* ข้อมูลตั้งแต่มีนาคม พ.ศ.2556 ถึง ธันวาคม พ.ศ.2558 และค่าสถิติแสดงความคลาดเคลื่อนและความเหมาะสมของตัวแบบพยากรณ์ (ต่อ)

ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	ผลการทดสอบการแจกแจง	สมการพยากรณ์	จำนวนตัวแปรอิสระ	MAPE	RMSE	Adjusted R-squared	Residual deviance
ตัวแบบเชิงเส้นนัยทั่วไปที่มีการแจกแจงอินเวอร์สเกาส์เซียน	-	$\frac{1}{\mu^2}$	มีการแจกแจงอินเวอร์เกาส์เซียน (p-value = 0.2455)	$\frac{1}{\hat{\mu}^2_{E20}} = 0.0003497 - 0.0000007409(WCO) - 0.000001212(G91) + 0.000003474(B95)$	3	2.9122	4.0984	-	0.0005
	-	$\log(\mu)$		$\log(\hat{\mu}_{E20}) = 3.1174413 - 0.0538930(P_E20) + 0.0045617(WCO) + 0.0425616(P_G95) + 0.0072978(G91) - 0.0166335(B95)$	5	2.0905	2.8019	-	0.0002
	-	$sqrt(\mu)$	มีการแจกแจงอินเวอร์เกาส์เซียน (p-value = 0.2455)	$sqrt(\hat{\mu}_{E20}) = 2.317418 - 0.288297(P_E20) + 0.021378(WCO) + 0.229135(P_G95) + 0.037231(G91) - 0.079726(B95)$	5	2.0792	2.7492	-	0.0002

จากตาราง 4.15 สมการที่ใช้ในพยากรณ์ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20 เมื่อพิจารณาค่า MAPE พบว่า ในการใช้ตัวแบบเชิงเส้นทั่วไปและตัวแบบเชิงเส้นนัยทั่วไป เป็นตัวแบบที่มีการพยากรณ์มีความแม่นสูง (MAPE ≤ 10) แต่ตัวแบบที่มีค่า MAPE ต่ำที่สุดคือ ตัวแบบเชิงเส้นนัยทั่วไปที่ y มีการแจกแจงแกมมาที่มีฟังก์ชันเชื่อมโยงเป็น $\log(\mu)$ เช่นเดียวกับการพิจารณาค่า RMSE การใช้ตัวแบบเชิงเส้นนัยทั่วไปที่ y มีการแจกแจงแกมมาที่มีฟังก์ชันเชื่อมโยงเป็น $\log(\mu)$ ก็จะให้ค่า RMSE ต่ำที่สุด โดยตัวแปรที่ส่งผลต่อปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20 ในทิศทางเดียวกัน คือ อัตราแลกเปลี่ยนเงินตรา ราคาน้ำมันดิบโลก ราคาน้ำมันแก๊สโซฮอล์ 95 นั้นหมายความว่า ถ้าราคาเอทานอล ราคาน้ำมันเบนซิน 95 และปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95 และปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91 เพิ่มขึ้น จะส่งผลให้ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20 เพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติ ส่วนตัวแปรที่ส่งผลในทิศทางตรงกันข้าม คือ ราคาน้ำมันแก๊สโซฮอล์ E20 และปริมาณการใช้น้ำมันเบนซิน 95 นั้นหมายความว่า ถ้าราคาน้ำมันแก๊สโซฮอล์ E20 และปริมาณการใช้น้ำมันเบนซิน 95 ลดลงจะส่งผลให้ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20 เพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติ

4.2.5 ผลการวิเคราะห์หาสมการที่เหมาะสมในการพยากรณ์**ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E85** และหาปัจจัยที่ส่งผลต่อการใช้ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E85 ซึ่งตัวแปรอิสระในการนำมาวิเคราะห์ประกอบด้วย

P_E85	คือ ราคาน้ำมันแก๊สโซฮอล์ E85	G95	คือ ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95
P_G95	คือ ราคาน้ำมันแก๊สโซฮอล์ 95	B95	คือ ปริมาณการใช้น้ำมันเบนซิน 95
P_B95	คือ ราคาน้ำมันเบนซิน 95	G91	คือ ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91
P_G91	คือ ราคาน้ำมันแก๊สโซฮอล์ 91	E20	คือ ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20
P_E20	คือ ราคาน้ำมันแก๊สโซฮอล์ E20	WCO	คือ ราคาน้ำมันดิบโลก
P_Ethanol	คือ ราคาเอทานอล	EXC	คือ อัตราแลกเปลี่ยนเงินตรา

พหุน ปณฺ ฑิโต ชีเว

ตาราง 4.16 สมการพยากรณ์*ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E85* และค่าสถิติแสดงความคลาดเคลื่อนและความเหมาะสมของตัวแบบพยากรณ์

ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	ผลการทดสอบการแจกแจง	สมการพยากรณ์	จำนวนตัวแปรอิสระ	MAPE	RMSE	Adjusted R-squared	Residual deviance
ตัวแบบเชิงเส้นทั่วไป	y	-	ไม่ได้มีการแจกแจงปรกติ (p-value ≤ 0.000)	$\hat{E85} = 25.7712 - 0.6016(EXC) - 3.0650(P_G91) + 3.4032(P_E20) - 0.1600(G91) + 0.5694(E20)$	5	111.1643	1.5930	0.9809	-
	$\frac{1}{y}$	-	ไม่ได้มีการแจกแจงปรกติ (p-value ≤ 0.000)	$\frac{1}{\hat{E85}} = 2.62228 + 0.51549(P_E85) - 0.14588(P_Ethanol) + 0.28374(EXC) - 0.07069(WCO) - 0.14305(G91) + 0.03886(G95) + 0.11123(B95) + 0.12173(E20)$	8	155.6555	21.4352	0.8990	-
	$\log(y)$	-	ไม่ได้มีการแจกแจงปรกติ (p-value ≤ 0.000)	$\log(\hat{E85}) = -1.76773 - 0.15322(P_E85) + 0.16851(P_G91) - 0.19290(P_G95) + 0.10069(P_B95) + 0.03368(G91) - 0.02025(G95)$	6	13.0163	3.2100	0.9926	-
	$q_normal(E85)$	-	มีการแจกแจงปรกติ (p-value = 0.09399)	$\hat{E85} = -11.834081 + 0.664537(P_G91) - 0.887061(P_G95) + 0.428169(P_B95) + 0.090261(G91) - 0.074425(G95) + 0.154602(E20)$	6	**11.2900**	**1.3700**	0.9949	-

$q_normal(E85)$: แปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ

ตาราง 4.16 สมการพยากรณ์*ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E85* และค่าสถิติแสดงความคลาดเคลื่อนและความเหมาะสมของตัวแบบพยากรณ์ (ต่อ)

ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	ผลการทดสอบการแจกแจง	สมการพยากรณ์	จำนวนตัวแปรอิสระ	MAPE	RMSE	Adjusted R-squared	Residual deviance
ตัวแบบเชิงเส้นนัยทั่วไปที่มีการแจกแจงแกมมา	-	$\log(\mu)$	มีการแจกแจงแกมมา (p-value = 0.09399)	$\log(\hat{\mu}_{E85}) = -1.77363 - 0.15239(P_E85) + 0.16974(P_G91) - 0.19373(P_G95) + 0.10020(P_B95) + 0.03361(G91) - 0.02013(G95)$	6	13.1116	3.2900	-	1.7828
ตัวแบบเชิงเส้นนัยทั่วไปที่มีการแจกแจงอินเวอร์เกาส์เซียน	-	$sqrt(\mu)$	มีการแจกแจงอินเวอร์เกาส์เซียน (p-value = 0.4206)	$sqrt(\hat{\mu}_{E85}) = 0.057916 - 0.087661(P_E85) + 0.009148(P_Ethanol) + 0.017063(EXC) - 0.002578(WCO) - 0.098109(P_G95) + 0.139022(P_E20) + 0.009271(G91) - 0.005192(G95) + 0.034672(E20)$	9	11.8364	3.3016	-	0.3904

จากตาราง 4.16 สมการที่ใช้ในพยากรณ์ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E85 เมื่อพิจารณาค่า MAPE พบว่า ในการใช้ตัวแบบเชิงเส้นทั่วไปที่แปลงค่า y ให้อยู่ในรูป $\log(y)$ และ แปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ รวมทั้งการใช้ตัวแบบเชิงเส้นนัยทั่วไป เป็นตัวแบบที่มีความแม่นในการพยากรณ์อยู่ในระดับดี (10 < MAPE ≤ 20) แต่ตัวแบบที่มีค่า MAPE ต่ำที่สุดคือ ตัวแบบเชิงเส้นทั่วไปที่แปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ เช่นเดียวกับการพิจารณาค่า RMSE การใช้ตัวแบบเชิงเส้นทั่วไปที่แปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ ก็จะให้ค่า RMSE ต่ำที่สุด โดยตัวแปรที่ส่งผลต่อปริมาณการใช้น้ำมันแก๊สโซฮอล์ E85 ดังนั้นตัวแปรที่ส่งผลต่อปริมาณการใช้น้ำมันแก๊สโซฮอล์ E85 ในทิศทางเดียวกัน คือ ราคาน้ำมันแก๊สโซฮอล์ 91 ราคาน้ำมันเบนซิน 95 ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91 และปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20 นั้นหมายความว่า ถ้าราคาน้ำมันแก๊สโซฮอล์ 91 ราคาน้ำมันเบนซิน 95 ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91 และปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20 เพิ่มขึ้น จะส่งผลให้ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E85 เพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติ ส่วนตัวแปรที่ส่งผลในทิศทางตรงกันข้าม คือ ราคาน้ำมันแก๊สโซฮอล์ 95 และปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95 นั้นหมายความว่า ถ้าราคาน้ำมันแก๊สโซฮอล์ 95 และปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95 ลดลง ส่งผลให้ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E85 เพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติ

4.2.6 ผลการวิเคราะห์หาสมการที่เหมาะสมในการพยากรณ์**ปริมาณการใช้น้ำมันเบนซิน 91** และหาปัจจัยที่ส่งผลต่อการใช้ปริมาณการใช้น้ำมันเบนซิน 91 ซึ่งตัวแปรอิสระในการนำมาวิเคราะห์ประกอบด้วย ข้อมูลตั้งแต่ กุมภาพันธ์ พ.ศ. 2553 ถึง ธันวาคม พ.ศ. 2555

P_B91 คือ ราคาน้ำมันเบนซิน 91
P_G95 คือ ราคาน้ำมันแก๊สโซฮอล์ 95
P_B95 คือ ราคาน้ำมันเบนซิน 95
P_G91 คือ ราคาน้ำมันแก๊สโซฮอล์ 91
P_E20 คือ ราคาน้ำมันแก๊สโซฮอล์ E20
P_E85 คือ ราคาน้ำมันแก๊สโซฮอล์ E85
P_NGV คือ ราคาแก๊ส NGV
P_LPG คือ ราคาแก๊ส LPG
EXC คือ อัตราแลกเปลี่ยนเงินตรา
G95 คือ ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95
B95 คือ ปริมาณการใช้น้ำมันเบนซิน 95
G91 คือ ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91
E20 คือ ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20
E85 คือ ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E85
LPG คือ ปริมาณการใช้แก๊ส LPG
NGV คือ ปริมาณการใช้แก๊ส NGV
WCO คือ ราคาน้ำมันดิบโลก

ตาราง 4.17 สมการพยากรณ์*ปริมาณการใช้น้ำมันเบนซิน 91* และค่าสถิติแสดงความคลาดเคลื่อนและความเหมาะสมของตัวแบบพยากรณ์

ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	ผลการทดสอบการแจกแจง	สมการพยากรณ์	จำนวนตัวแปรอิสระ	MAPE	RMSE	Adjusted R-squared	Residual deviance
ตัวแบบเชิงเส้นทั่วไป	y	-	ไม่ได้มีการแจกแจงปรกติ (p-value ≤ 0.000)	$\hat{B91} = 133.2913 - 14.6544(P_B91) + 0.9172(WCO) + 15.6165(P_G91) + 11.4914(P_G95) - 17.5601(P_E20) + 0.5483(G91) - 0.8934(E20) + 34.7601(P_LPG) - 15.9177(P_NGV)$	9	**1.1848**	**3.6258**	0.9578	-
	$\frac{1}{y}$	-	ไม่ได้มีการแจกแจงปรกติ (p-value ≤ 0.000)	$\frac{1}{\hat{B91}} = 0.004757 + 0.0002125(P_B91) - 0.00001498(WCO) - 0.0002335(P_G91) - 0.0001561(P_G95) + 0.0002703(P_E20) - 0.000009029(G91) + 0.00001374(E20) - 0.0002451(P_LPG)$	8	1.4085	4.3281	0.9427	-
	$\frac{1}{y^2}$	-	ไม่ได้มีการแจกแจงปรกติ (p-value ≤ 0.000)	$\frac{1}{(\hat{B91})^2} = 0.00002231 + 0.000001641(P_B91) - 0.0000001207(WCO) - 0.000001927(P_G91) - 0.000001132(P_G95) + 0.000002176(P_E20) - 0.00000007188(G91) + 0.0000001140(E20) - 0.000002012(P_LPG)$	8	1.4866	4.5696	0.9364	-
	$\log(y)$	-	ไม่ได้มีการแจกแจงปรกติ (p-value ≤ 0.000)	$\log(\hat{B91}) = 5.340598 - 0.055287(P_B91) + 0.003735(WCO) + 0.056328(P_G91) + 0.043224(P_G95) - 0.067299(P_E20) + 0.002265(G91) - 0.003297(E20) + 0.059923(P_LPG)$	8	1.3393	4.1574	0.9471	-

ตาราง 4.17 สมการพยากรณ์*ปริมาณการใช้น้ำมันเบนซิน 91* และค่าสถิติแสดงความคลาดเคลื่อนและความเหมาะสมของตัวแบบพยากรณ์

ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	ผลการทดสอบการแจกแจง	สมการพยากรณ์	จำนวนตัวแปรอิสระ	MAPE	RMSE	Adjusted R-squared	Residual deviance
ตัวแบบเชิงเส้นทั่วไป	$q_normal(B91)$	-	มีการแจกแจงปรกติ (p-value = 0.929)	$\hat{B91} = 136.9830 - 14.7757(P_B91) + 0.9552(WCO) + 16.3747(P_G91) + 11.1704(P_G95) - 18.1928(P_E20) + 0.5724(G91) - 0.9421(E20) + 33.7328(P_LPG) - 14.4845(P_NGV)$	9	1.2400	3.7500	0.9543	-
ตัวแบบเชิงเส้นนัยทั่วไปที่มีการแจกแจงแกมมา	-	*Inverse*	มีการแจกแจงแกมมา (p-value = 0.929)	$\frac{1}{\hat{\mu}_{B91}} = 0.004726 + 0.0002110(P_B91) - 0.00001544(WCO) - 0.0002289(P_G91) - 0.0001509(P_G95) + 0.0002624(P_E20) - 0.000008910(G91) + 0.00001295(E20) - 0.0002420(P_LPG)$	8	1.4017	4.3007	-	0.0093
	-	$\log(\mu)$		$\log(\hat{\mu}_{B91}) = 5.341935 - 0.055320(P_B91) + 0.003739(WCO) + 0.056331(P_G91) + 0.043215(P_G95) - 0.067290(P_E20) + 0.002266(G91) - 0.003294(E20) + 0.059873(P_LPG)$	8	1.3383	4.1576	-	0.0087
ตัวแบบเชิงเส้นนัยทั่วไปที่มีการแจกแจงอินเวอร์เกาส์เซียน	-	$\frac{1}{\mu^2}$	มีการแจกแจงแบบอินเวอร์เกาส์เซียน (p-value = 0.1006)	$\frac{1}{\hat{\mu}^2_{B91}} = 0.00002193 + 0.000001618(P_B91) - 0.0000001246(WCO) - 0.000001860(P_G91) - 0.000001069(P_G95) + 0.000002066(P_E20) - 0.00000007026(G91) + 0.0000001037(E20) - 0.000001960(P_LPG)$	8	1.4732	4.4919	-	<0.0000

$q_normal(E85)$: แปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ

ตาราง 4.17 สมการพยากรณ์*ปริมาณการใช้น้ำมันเบนซิน 91* และค่าสถิติแสดงความคลาดเคลื่อนและความเหมาะสมของตัวแบบพยากรณ์ (ต่อ)

ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	ผลการทดสอบการแจกแจง	สมการพยากรณ์	จำนวนตัวแปรอิสระ	MAPE	RMSE	Adjusted R-squared	Residual deviance
ตัวแบบเชิงเส้นนัยทั่วไปที่มีการแจกแจงอินเวอร์เกาส์เซียน	-	$\log(\mu)$	มีการแจกแจงอินเวอร์เกาส์เซียน (p-value = 0.1006)	$\log(\hat{\mu}_{B91}) = 5.338932 - 0.055499(P_B91) + 0.003675(WCO) + 0.056690(P_G91) + 0.043963(P_G95) - 0.068142(P_E20) + 0.002277(G91) - 0.003382(E20) + 0.060164(P_LPG)$	8	1.3415	4.1678	-	<0.0000
	-	$sqrt(\mu)$		$sqrt(\hat{\mu}_{B91}) = 14.37484 - 0.44935(P_B91) + 0.02847(WCO) + 0.44145(P_G91) + 0.37173(P_G95) - 0.54351(P_E20) + 0.01807(G91) - 0.02676(E20) + 0.47262(P_LPG)$	8	1.3215	4.1089	-	<0.0000

จากตาราง 4.17 สมการที่ใช้ในพยากรณ์ปริมาณการใช้น้ำมันเบนซิน 91 เมื่อพิจารณาค่า MAPE พบว่า ในการใช้ตัวแบบเชิงเส้นทั่วไปและตัวแบบเชิงเส้นนัยทั่วไป เป็นตัวแบบที่มีการพยากรณ์มีความแม่นสูง (MAPE ≤ 10) แต่ตัวแบบที่มีค่า MAPE ต่ำที่สุดคือ ตัวแบบเชิงเส้นทั่วไปที่ไม่ได้แปลงค่า y เช่นเดียวกับการพิจารณาค่า RMSE การใช้ตัวแบบเชิงเส้นทั่วไปที่ไม่ได้แปลงค่า y ก็จะให้ค่า RMSE ต่ำที่สุด โดยตัวแปรที่ส่งผลต่อปริมาณการใช้น้ำมันเบนซิน 91 ในทิศทางเดียวกัน คือ ราคาน้ำมันดิบโลก ราคาน้ำมันแก๊สโซฮอล์ 91 ราคาน้ำมันแก๊สโซฮอล์ 95 ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91 และราคาแก๊ส LPG นั้นหมายความว่า ถ้าราคาน้ำมันดิบโลก ราคาน้ำมันแก๊สโซฮอล์ 91 ราคาน้ำมันแก๊สโซฮอล์ 95 ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91 และราคาแก๊ส LPG เพิ่มขึ้น จะส่งผลให้ปริมาณการใช้น้ำมันเบนซิน 91 เพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติ ส่วนตัวแปรที่ส่งผลในทิศทางตรงกันข้าม คือ ราคาน้ำมันเบนซิน 91 ราคาน้ำมันแก๊สโซฮอล์ E20 ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20 และราคาแก๊ส NGV นั้นหมายความว่า ถ้าราคาน้ำมันเบนซิน 91 ราคาน้ำมันแก๊สโซฮอล์ E20 ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20 และราคาแก๊ส NGV ลดลง ส่งผลให้ปริมาณการใช้น้ำมันเบนซิน 91 เพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติ

4.2.7 ผลการวิเคราะห์หาสมการที่เหมาะสมในการพยากรณ์**ปริมาณการใช้น้ำมันเบนซิน 95** และหาปัจจัยที่ส่งผลต่อการใช้ปริมาณการใช้น้ำมันเบนซิน 95 ซึ่งตัวแปรอิสระในการนำมาวิเคราะห์ประกอบด้วย

P_B95 คือ ราคาน้ำมันเบนซิน 95
P_G95 คือ ราคาน้ำมันแก๊สโซฮอล์ 95
P_G91 คือ ราคาน้ำมันแก๊สโซฮอล์ 91
P_E20 คือ ราคาน้ำมันแก๊สโซฮอล์ E20
P_E85 คือ ราคาน้ำมันแก๊สโซฮอล์ E85
P_NGV คือ ราคาแก๊ส NGV
P_LPG คือ ราคาแก๊ส LPG
EXC คือ อัตราแลกเปลี่ยนเงินตรา
G95 คือ ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95
G91 คือ ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91
E20 คือ ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20
E85 คือ ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E85
LPG คือ ปริมาณการใช้แก๊ส LPG
NGV คือ ปริมาณการใช้แก๊ส NGV
WCO คือ ราคาน้ำมันดิบโลก

พหุน ปณฺ ฑิโต ชีเว

ตาราง 4.18 สมการพยากรณ์*ปริมาณการใช้น้ำมันเบนซิน 95* และค่าสถิติแสดงความคลาดเคลื่อนและความเหมาะสมของตัวแบบพยากรณ์

ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	ผลการทดสอบการแจกแจง	สมการพยากรณ์	จำนวนตัวแปรอิสระ	MAPE	RMSE	Adjusted R-squared	Residual deviance
ตัวแบบเชิงเส้นทั่วไป	y	-	ไม่ได้มีการแจกแจงปรกติ (p-value ≤ 0.000)	$\hat{B95} = -40.44375 - 3.03114(P_B95) + 5.54552(P_G95) - 2.84934(P_E85) + 0.14041(G95) - 0.43299(E85) + 0.35024(G91) - 0.06663(LPG) - 0.13616(NGV) + 6.92681(P_LPG) - 5.66885(P_NGV)$	10	30.6112	**2.9343**	0.9769	-
	$\frac{1}{y}$	-	ไม่ได้มีการแจกแจงปรกติ (p-value ≤ 0.000)	$\frac{1}{\hat{B95}} = 1.6475143 + 0.0439945(P_B95) - 0.0161568(EXC) - 0.1043751(P_G95) + 0.0537772(P_E20) - 0.0031065(G95) - 0.0075888(E85) + 0.0022151(G91) - 0.0589233(P_LPG)$	8	258.3843	543.7466	0.9433	-
	$\log(y)$	-	ไม่ได้มีการแจกแจงปรกติ (p-value ≤ 0.000)	$\log(\hat{B95}) = -0.6324247 - 0.2445370(P_B95) + 0.5792734(P_G95) - 0.2472373(P_E20) + 0.0171556(G95) + 0.0328272(E85) - 0.0064584(LPG)$	6	**15.7507**	5.2694	0.9691	-

ตาราง 4.18 สมการพยากรณ์*ปริมาณการใช้น้ำมันเบนซิน 95* และค่าสถิติแสดงความคลาดเคลื่อนและความเหมาะสมของตัวแบบพยากรณ์ (ต่อ)

ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	ผลการทดสอบการแจกแจง	สมการพยากรณ์	จำนวนตัวแปรอิสระ	MAPE	RMSE	Adjusted R-squared	Residual deviance
ตัวแบบเชิงเส้นนัยทั่วไปที่มีการแจกแจงแกมมา	-	*Inverse*	ไม่ได้มีการแจกแจงแกมมา (p-value ≤ 0.000)	$\frac{1}{\hat{\mu}_{B95}} = 0.6571 + 0.01486(P_B95) - 0.005035(EXC) - 0.05113(P_G95) + 0.03342(P_E20) - 0.0008126(G95) - 0.001623(E85) + 0.0003016(LPG) - 0.01989(P_LPG)$	8	33.2887	11.1472	-	9.9401
	-	$\log(\mu)$		$\log(\hat{\mu}_{B95}) = -0.5560781 - 0.2411372(P_B95) + 0.5725605(P_G95) - 0.2449469(P_E20) + 0.0171894(G95) + 0.0323290(E85) - 0.0065433(LPG)$	6	16.2996	5.4012	-	3.0822
ตัวแบบเชิงเส้นนัยทั่วไปที่มีการแจกแจงอินเวอร์เกาส์เซียน	-	$\log(\mu)$	มีการแจกแจงอินเวอร์เกาส์เซียน (p-value =0.1204)	$\log(\hat{\mu}_{B95}) = 1.494558 - 0.234357(P_B95) + 0.371446(P_G95) - 0.277742(P_E85) + 0.013146(G95) + 0.022379(E20)$	5	26.5532	20.6212	-	0.4960 9

ตาราง 4.18 สมการพยากรณ์*ปริมาณการใช้น้ำมันเบนซิน 95* และค่าสถิติแสดงความคลาดเคลื่อนและความเหมาะสมของตัวแบบพยากรณ์ (ต่อ)

ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	ผลการทดสอบการแจกแจง	สมการพยากรณ์	จำนวนตัวแปรอิสระ	MAPE	RMSE	Adjusted R-squared	Residual deviance
ตัวแบบเชิงเส้นนัยทั่วไปที่มีการแจกแจงอินเวอร์เกาส์เซียน	-	$sqrt(\mu)$	มีการแจกแจงอินเวอร์เกาส์เซียน (p-value =0.1204)	$sqrt(\hat{\mu}_{B95}) = 1.696397 - 0.262731(P_B95) + 0.493669(P_G95) - 0.329854(P_E85) + 0.013257(G95) + 0.055586(E20) - 0.011476(NGV)$	6	22.2734	8.8334	-	0.5337 8

จากตาราง 4.18 สมการที่ใช้ในพยากรณ์ปริมาณการใช้น้ำมันเบนซิน 95 เมื่อพิจารณาค่า MAPE พบว่า ในการใช้ตัวแบบเชิงเส้นทั่วไปที่แปลงค่า y ให้อยู่ในรูป $\log(y)$ รวมทั้งการใช้ตัวแบบเชิงเส้นนัยทั่วไปที่ y มีการแจกแจงแกมมาที่มีฟังก์ชันเชื่อมโยงเป็น $\log(\mu)$ เป็นตัวแบบที่มีความแม่นในการพยากรณ์อยู่ในระดับดี (10 < MAPE ≤ 20) แต่ตัวแบบที่มีค่า MAPE ต่ำที่สุดคือ ตัวแบบเชิงเส้นทั่วไปที่แปลงค่า y ให้อยู่ในรูป $\log(y)$ โดยตัวแปรที่ส่งผลต่อปริมาณการใช้น้ำมันเบนซิน 95 ในทิศทางเดียวกัน คือ ราคาน้ำมันแก๊สโซฮอล์ 95 ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95 และปริมาณการใช้น้ำมันแก๊สโซฮอล์ E85 นั้นหมายความว่า ถ้าราคาน้ำมันแก๊สโซฮอล์ 95 ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95 และปริมาณการใช้น้ำมันแก๊สโซฮอล์ E85 เพิ่มขึ้น จะส่งผลให้ปริมาณการใช้น้ำมันเบนซิน 95 เพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติ ส่วนตัวแปรที่ส่งผลในทิศทางตรงกันข้าม คือ ราคาน้ำมันเบนซิน 95 ราคาน้ำมันแก๊สโซฮอล์ E20 และปริมาณการใช้แก๊ส LPG นั้นหมายความว่า ถ้าราคาน้ำมันเบนซิน 95 ราคาน้ำมันแก๊สโซฮอล์ E20 และปริมาณการใช้แก๊ส LPG ลดลง ส่งผลให้ปริมาณการใช้น้ำมันเบนซิน 95 เพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติ แต่เมื่อพิจารณาค่า RMSE กลับพบว่าตัวแบบที่มีค่า RMSE ต่ำสุดคือ ทั้งตัวแบบเชิงเส้นทั่วไปที่ไม่ได้ค่า y ดังนั้นตัวแปรที่ส่งผลต่อปริมาณการใช้น้ำมันเบนซิน 95 ในทิศทางเดียวกัน คือ ราคาน้ำมันแก๊สโซฮอล์ 95 ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95 ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91 และราคาแก๊ส LPG นั้นหมายความว่า ถ้าราคาน้ำมันแก๊สโซฮอล์ 95 ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95 ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91 และราคาแก๊ส LPG เพิ่มขึ้น จะส่งผลให้ปริมาณการใช้น้ำมันเบนซิน 95 เพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติ ส่วนตัวแปรที่ส่งผลในทิศทางตรงกันข้าม คือ ราคาน้ำมันเบนซิน 95 ราคาน้ำมันแก๊สโซฮอล์ E85 ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E85 ปริมาณการใช้แก๊ส LPG ปริมาณการใช้แก๊ส NGV และราคาแก๊ส NGV นั้นหมายความว่า ถ้าราคาน้ำมันเบนซิน 95 ราคาน้ำมันแก๊สโซฮอล์ E85 ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E85 ปริมาณการใช้แก๊ส LPG ปริมาณการใช้แก๊ส NGV และราคาแก๊ส NGV ลดลง ส่งผลให้ปริมาณการใช้น้ำมันเบนซิน 95 เพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติ

พหุน ปณฺ ฑิโต ชีเว

4.2.9 ผลการวิเคราะห์หาสมการที่เหมาะสมในการพยากรณ์**ปริมาณการใช้แก๊ส LPG** และหาปัจจัยที่ส่งผลต่อการใช้ปริมาณการใช้แก๊ส LPG ซึ่งตัวแปรอิสระในการนำมาวิเคราะห์ประกอบด้วย

P_LPG คือ ราคาแก๊ส LPG
P_G95 คือ ราคาน้ำมันแก๊สโซฮอล์ 95
P_G91 คือ ราคาน้ำมันแก๊สโซฮอล์ 91
P_E20 คือ ราคาน้ำมันแก๊สโซฮอล์ E20
P_E85 คือ ราคาน้ำมันแก๊สโซฮอล์ E85
P_NGV คือ ราคาแก๊ส NGV
P_B91 คือ ราคาน้ำมันเบนซิน 91
P_B95 คือ ราคาน้ำมันเบนซิน 95
EXC คือ อัตราแลกเปลี่ยนเงินตรา
G95 คือ ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95
G91 คือ ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91
E20 คือ ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20
E85 คือ ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E85
B91 คือ ปริมาณการใช้น้ำมันเบนซิน 91
B95 คือ ปริมาณการใช้น้ำมันเบนซิน 95
NGV คือ ปริมาณการใช้แก๊ส NGV
WCO คือ ราคาน้ำมันดิบโลก
P_Ethanol คือ ราคาเอทานอล

พหุน ปณ ฑิโต ชีเว

ตาราง 4.19 สมการพยากรณ์*ปริมาณการใช้แก๊ส LPG* และค่าสถิติแสดงความคลาดเคลื่อนและความเหมาะสมของตัวแบบพยากรณ์

ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	ผลการทดสอบการแจกแจง	สมการพยากรณ์	จำนวนตัวแปรอิสระ	MAPE	RMSE	Adjusted R-squared	Residual deviance
ตัวแบบเชิงเส้นทั่วไป	y	-	ไม่ได้มีการแจกแจงปรกติ (p-value ≤ 0.000)	$\hat{LPG} = 371.530 - 3.345(P_B95) + 15.269(P_E20) - 11.643(P_E85) - 28.444(P_NGV) - 3.121(B95) + 1.769(G91) - 2.924(E85)$	7	3.0368	18.2751	0.7746	-
	$\frac{1}{y}$	-	ไม่ได้มีการแจกแจงปรกติ (p-value ≤ 0.000)	$\frac{1}{\hat{LPG}} = 0.002578 + 0.00001533(P_B95) - 0.00008091(P_E20) + 0.00007175(P_E85) + 0.0001362(P_NGV) + 0.00001494(B95) - 0.000008580(G91) + 0.00001300(E85)$	7	3.0230	18.327	0.7909	-
	$\frac{1}{y^2}$	-	ไม่ได้มีการแจกแจงปรกติ (p-value ≤ 0.000)	$\frac{1}{(\hat{LPG})^2} = 0.000007178 - 0.0000002582(P_E20) + 0.0000002626(P_E85) + 0.0000005602(P_NGV) + 0.00000005937(B95) - 0.00000003335(G91) + 0.00000004859(E85)$	6	3.2410	19.5235	0.785	-
	$\log(y)$	-	ไม่ได้มีการแจกแจงปรกติ (p-value ≤ 0.000)	$\log(\hat{LPG}) = 5.862063 + 0.022426(P_E20) - 0.019085(P_E85) - 0.057324(P_NGV) - 0.006083(B95) + 0.003358(G91) - 0.005441(E85)$	6	3.2092	18.9891	0.7727	-

ตาราง 4.19 สมการพยากรณ์*ปริมาณการใช้แก๊ส LPG* และค่าสถิติแสดงความคลาดเคลื่อนและความเหมาะสมของตัวแบบพยากรณ์ (ต่อ)

ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	ผลการทดสอบการแจกแจง	สมการพยากรณ์	จำนวนตัวแปรอิสระ	MAPE	RMSE	Adjusted R-squared	Residual deviance
ตัวแบบเชิงเส้นนัยทั่วไปที่มีการแจกแจงแกมมา	-	*Inverse*	ไม่ได้มีการแจกแจงแกมมา (p-value = 0.01084)	$\frac{1}{\hat{\mu}_{LPG}} = 0.002572 + 0.00001620(P_B95) - 0.00007965(P_E20) + 0.00006726(P_E85) + 0.0001358(P_NGV) + 0.00001468(B95) - 0.000008454(G91) + 0.00001330(E85)$	7	3.0309	18.2471	-	0.1139
	-	$\log(\mu)$		$\log(\hat{\mu}_{LPG}) = 5.866749 + 0.022222(P_E20) - 0.018762(P_E85) - 0.057715(P_NGV) - 0.006087(B95) + 0.003356(G91) - 0.005422(E85)$	6	3.2274	18.9826	-	0.1217
ตัวแบบเชิงเส้นนัยทั่วไปที่มีการแจกแจงอินเวอร์เกาส์เซียน	-	$\frac{1}{\mu^2}$	มีการแจกแจงอินเวอร์เกาส์เซียน (p-value = 0.3302)	$\frac{1}{\hat{\mu}^2_{LPG}} = 0.0000005927 + 0.0000001664(EXC) + 0.000000008683(P_B95) - 0.0000003694(P_E20) + 0.0000003446(P_E85) + 0.0000005197(P_NGV) + 0.00000006312(B95) - 0.00000003466(G91) + 0.00000003693(E85)$	8	**3.0100**	**17.4752**	-	0.0002

ตาราง 4.19 สมการพยากรณ์*ปริมาณการใช้แก๊ส LPG* และค่าสถิติแสดงความคลาดเคลื่อนและความเหมาะสมของตัวแบบพยากรณ์ (ต่อ)

ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	ผลการทดสอบการแจกแจง	สมการพยากรณ์	จำนวนตัวแปรอิสระ	MAPE	RMSE	Adjusted R-squared	Residual deviance
ตัวแบบเชิงเส้นนัยทั่วไปที่มีการแจกแจงอินเวอร์เกาส์เซียน	-	$\log(\mu)$	มีการแจกแจงอินเวอร์เกาส์เซียน (p-value = 0.3302)	$\log(\hat{\mu}_{LPG}) = 5.877203 + 0.022799(P_E20) - 0.020266(P_E85) - 0.058480(P_NGV) - 0.006198(B950) + 0.003417(G91) - 0.005346(E85)$	6	3.2284	19.0231	-	0.0003
	-	$sqrt(\mu)$		$sqrt(\hat{\mu}_{LPG}) = 28.783825 - 0.193568(EXC) - 0.822170(P_LPG) + 0.090294(P_E20) - 0.030816(B95) + 0.009326(G91) + 0.015814(NGV)$	6	2.9164	17.5138	-	0.0002

จากตาราง 4.19 สมการที่ใช้ในพยากรณ์ปริมาณการใช้แก๊ส LPG เมื่อพิจารณาค่า MAPE พบว่าในการใช้ตัวแบบเชิงเส้นทั่วไปและตัวแบบเชิงเส้นนัยทั่วไป เป็นตัวแบบที่มีการพยากรณ์มีความแม่นสูง (MAPE ≤ 10) แต่ตัวแบบที่มีค่า MAPE ต่ำที่สุดคือ ตัวแบบเชิงเส้นนัยทั่วไปที่ y มีการแจกแจงอินเวอร์เกาส์เซียนที่มีฟังก์ชันเชื่อมโยงเป็น $\frac{1}{\mu^2}$ เช่นเดียวกับการพิจารณาค่า RMSE การใช้ตัวแบบเชิงเส้นนัยทั่วไปที่ y มีการแจกแจงอินเวอร์เกาส์เซียนที่มีฟังก์ชันเชื่อมโยงเป็น $\frac{1}{\mu^2}$ ก็จะให้ค่า RMSE ต่ำที่สุด โดยตัวแปรที่ส่งผลต่อปริมาณการใช้แก๊ส LPG ในทิศทางเดียวกัน คือ ราคาน้ำมันแก๊สโซฮอล์ E20 และปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91 นั้นหมายความว่า ถ้าราคาน้ำมันแก๊สโซฮอล์ E20 และปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91 ลดลง จะส่งผลให้ปริมาณการใช้แก๊ส LPG ลดลงอย่างมีนัยสำคัญทางสถิติ ส่วนตัวแปรที่ส่งผลในทิศทางตรงกันข้าม คือ อัตราแลกเปลี่ยนเงินตรา ราคาน้ำมันเบนซิน 95 ราคาน้ำมันแก๊สโซฮอล์ E85 ราคาแก๊ส NGV ปริมาณการใช้น้ำมันเบนซิน 95 และปริมาณการใช้น้ำมันแก๊สโซฮอล์ E85 หมายความว่า ถ้าอัตราแลกเปลี่ยนเงินตรา ราคาน้ำมันเบนซิน 95 ราคาน้ำมันแก๊สโซฮอล์ E85 ราคาแก๊ส NGV ปริมาณการใช้น้ำมันเบนซิน 95 และปริมาณการใช้น้ำมันแก๊สโซฮอล์ E85 เพิ่มขึ้นจะส่งผลให้ปริมาณการใช้แก๊ส LPG ลดลงอย่างมีนัยสำคัญทางสถิติ

หมายเหตุ : ตัวแบบเชิงเส้นนัยทั่วไปที่มีฟังก์ชันเชื่อมโยงเป็น inverse link ถ้าสัมประสิทธิ์เป็นลบหมายถึงมีความสัมพันธ์ในทิศทางเดียวกันกับค่าเฉลี่ยของ y

4.2.10 ผลการวิเคราะห์หาสมการที่เหมาะสมในการพยากรณ์**ปริมาณการใช้แก๊ส NGV** และหาปัจจัยที่ส่งผลต่อการใช้ปริมาณการใช้แก๊ส NGV ซึ่งตัวแปรอิสระในการนำมาวิเคราะห์ประกอบด้วย

P_NGV คือ ราคาแก๊ส NGV
P_G95 คือ ราคาน้ำมันแก๊สโซฮอล์ 95
P_G91 คือ ราคาน้ำมันแก๊สโซฮอล์ 91
P_E20 คือ ราคาน้ำมันแก๊สโซฮอล์ E20
P_E85 คือ ราคาน้ำมันแก๊สโซฮอล์ E85
P_ LPG คือ ราคาแก๊ส LPG
P_B91 คือ ราคาน้ำมันเบนซิน 91
P_B95 คือ ราคาน้ำมันเบนซิน 95
EXC คือ อัตราแลกเปลี่ยนเงินตรา
G95 คือ ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95
G91 คือ ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91
E20 คือ ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20
E85 คือ ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E85
B91 คือ ปริมาณการใช้น้ำมันเบนซิน 91
B95 คือ ปริมาณการใช้น้ำมันเบนซิน 95
LPG คือ ปริมาณการใช้แก๊ส LPG
WCO คือ ราคาน้ำมันดิบโลก
P_Ethanol คือ ราคาเอทานอล

ตาราง 4.20 สมการพยากรณ์*ปริมาณการใช้แก๊ส NGV* และค่าสถิติแสดงความคลาดเคลื่อนและความเหมาะสมของตัวแบบพยากรณ์

ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	ผลการทดสอบการแจกแจง	สมการพยากรณ์	จำนวนตัวแปรอิสระ	MAPE	RMSE	Adjusted R-squared	Residual deviance
ตัวแบบเชิงเส้นทั่วไป	y	-	ไม่ได้มีการแจกแจงปรกติ (p-value ≤ 0.000)	$\hat{NGV} = -34.1751 + 4.7870(P_G91) - 5.4784(P_E85) + 4.7439(P_LPG) - 0.7915(B95) + 1.3914(G91) - 0.2932(G95) - 1.0227(E20)$	7	**2.5855**	**7.4596**	**0.**9643	**-**
	$\frac{1}{y}$	-	ไม่ได้มีการแจกแจงปรกติ (p-value ≤ 0.000)	$\frac{1}{\hat{NGV}} = 0.01087 - 0.00004718(P_Ethanol) - 0.0001325(P_G91) + 0.0002206(P_E85) + 0.00003387(B95) - 0.00005029(G91) + 0.00001089(G95) + 0.00004329(E20)$	7	3.3204	9.4516	0.9621	-
	$\frac{1}{y^2}$	-	ไม่ได้มีการแจกแจงปรกติ (p-value ≤ 0.000)	$\frac{1}{(\hat{NGV})^2} = 0.00009310 - 0.0000007707(P_Ethanol) - 0.000001408(P_G91) + 0.000002394(P_E85) + 0.0000004268(B95) - 0.0000005404(G91) + 0.0000001261(G95) + 0.0000003889(E20) + 0.0000002776(E85)$	8	4.2571	13.7044	0.9534	-
	$\log(y)$	-	ไม่ได้มีการแจกแจงปรกติ (p-value ≤ 0.000)	$\log(\hat{NGV}) = 3.977755 + 0.008625(P_Ethanol) + 0.026290(P_G91) - 0.039779(P_E85) + 0.025893(P_LPG) - 0.005373(B95) + 0.008303(G91) - 0.001835(G95) - 0.006918(E20)$	8	2.7536	7.5671	0.9679	-

ตาราง 4.20 สมการพยากรณ์*ปริมาณการใช้แก๊ส NGV* และค่าสถิติแสดงความคลาดเคลื่อนและความเหมาะสมของตัวแบบพยากรณ์ (ต่อ)

ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	ผลการทดสอบการแจกแจง	สมการพยากรณ์	จำนวนตัวแปรอิสระ	MAPE	RMSE	Adjusted R-squared	Residual deviance
ตัวแบบเชิงเส้นนัยทั่วไปที่มีการแจกแจงแกมมา	-	*Inverse*	ไม่ได้มีการแจกแจงแกมมา (p-value = 0.003596)	$\frac{1}{\hat{\mu}_{NGV}} = 0.01006 - 0.0001078(P_G91) + 0.0001352(P_E85) + 0.00002616(B95) - 0.00004302(G91) + 0.000009565(G95) + 0.00003555(E20)$	6	3.2302	8.5399	-	0.1136
	-	$\log(\mu)$		$\log(\hat{\mu}_{NGV}) = 3.981280 + 0.008520(P_Ethanol) + 0.026188(P_G91) - 0.039471(P_E85) + 0.025668(P_LPG) - 0.005363(B95) + 0.008298(G91) - 0.001837(G95) - 0.006915(E20)$	8	2.7440	7.5576	-	0.0817
ตัวแบบเชิงเส้นนัยทั่วไปที่มีการแจกแจงอินเวอร์เกาส์เซียน	-	$\frac{1}{\mu^2}$	ไม่ได้มีการแจกแจงอินเวอร์เกาส์เซียน (p-value = 0.0009)	$\frac{1}{\hat{\mu}^2_{NGV}} = 0.00007579 - 0.000001049(P_G91) + 0.000001351(P_E85) + 0.0000002736(B95) - 0.0000004452(G91) + 0.0000001026(G95) + 0.0000003797(E20)$	6	3.9059	10.4958	-	0.0008
	-	$\log(\mu)$		$\log(\hat{\mu}_{NGV}) = 3.953037 + 0.010118(P_Ethanol) + 0.027599(P_G91) - 0.044575(P_E85) + 0.028738(P_LPG) - 0.005626(B95) + 0.008519(G91) - 0.001881(G95) - 0.007145(E20)$	8	2.7804	7.7403	-	0.0004

ตาราง 4.20 สมการพยากรณ์*ปริมาณการใช้แก๊ส NGV* และค่าสถิติแสดงความคลาดเคลื่อนและความเหมาะสมของตัวแบบพยากรณ์ (ต่อ)

ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	ผลการทดสอบการแจกแจง	สมการพยากรณ์	จำนวนตัวแปรอิสระ	MAPE	RMSE	Adjusted R-squared	Residual deviance
ตัวแบบเชิงเส้นนัยทั่วไปที่มีการแจกแจงอินเวอร์สเกาส์เซียน	-	$sqrt(\mu)$	ไม่ได้มีการแจกแจงอินเวอร์สเกาส์เซียน (p-value = 0.0009)	$sqrt(\hat{\mu}_{NGV}) = 3.132950 + 0.058355(P_Ethanol) + 0.164153(P_G91) - 0.291330(P_E85) + 0.372187(P_LPG) - 0.027840(B95) + 0.048672(G91) - 0.011635(G95) - 0.038478(E20) + 0.003901(LPG)$	9	2.6208	7.4755	-	0.0003

จากตาราง 4.20 สมการที่ใช้ในพยากรณ์ปริมาณการใช้แก๊ส NGV เมื่อพิจารณาค่า MAPE พบว่าในการใช้ตัวแบบเชิงเส้นทั่วไปและตัวแบบเชิงเส้นนัยทั่วไป เป็นตัวแบบที่มีการพยากรณ์มีความแม่นสูง (MAPE ≤ 10) แต่ตัวแบบที่มีค่า MAPE ต่ำที่สุดคือ ตัวแบบเชิงเส้นทั่วไปที่ไม่ได้มีการแปลงค่า y เช่นเดียวกับการพิจารณาค่า RMSE การใช้ตัวแบบเชิงเส้นทั่วไปที่ไม่ได้มีการแปลงค่า y ก็จะให้ค่า RMSE ต่ำที่สุด โดยตัวแปรที่ส่งผลต่อปริมาณการใช้แก๊ส NGV ในทิศทางเดียวกัน คือ ราคาน้ำมันแก๊สโซฮอล์ 91 ราคาแก๊ส LPG และปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91 นั้นหมายความว่า ถ้าราคาน้ำมันแก๊สโซฮอล์ 91 ราคาแก๊ส LPG และปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91 เพิ่มขึ้น จะส่งผลให้ปริมาณการใช้แก๊ส NGV เพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติ ส่วนตัวแปรที่ส่งผลในทิศทางตรงกันข้าม คือ ราคาน้ำมันแก๊สโซฮอล์ E85 ปริมาณการใช้น้ำมันเบนซิน 95 ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95 และปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20 นั้นหมายความว่า ถ้าราคาน้ำมันแก๊สโซฮอล์ E85 ปริมาณการใช้น้ำมันเบนซิน 95 ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95 และปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20 ลดลง ส่งผลให้ปริมาณการใช้แก๊ส NGV เพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติ

พหุน ปณ ฑิโต ชีเว

บทที่ 5
สรุปผล อภิปรายผล และข้อเสนอแนะ

ในการวิจัยเรื่อง ประสิทธิภาพของการพยากรณ์โดยตัวแบบเชิงเส้นนัยทั่วไปเมื่อข้อมูลมีการแจกแจงแบบเบ้ขวากับการประยุกต์สำหรับปริมาณการใช้น้ำมันสำเร็จรูปของประเทศไทย มีจุดมุ่งหมายเพื่อการเปรียบเทียบประสิทธิภาพของการพยากรณ์เมื่อข้อมูลมีการแจกแจงแบบเบ้ขวาด้วยวิธีการวิเคราะห์ตัวแบบเชิงเส้นนัยทั่วไป (Generalized Linear Model) กับวิธีการวิเคราะห์ตัวแบบเชิงเส้นทั่วไป (General Linear Model) โดยการแปลงข้อมูลใช้วิธีการจำลองข้อมูลขึ้นมาด้วยการใช้โปรแกรม R แล้วเปรียบเทียบประสิทธิภาพของตัวแบบทั้งสองจากค่าเปอร์เซ็นต์ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percent Error : MAPE) กับรากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error : RMSE) และเพื่อหาสมการ ปัจจัยที่ส่งผลต่อปริมาณการใช้น้ำมันสำเร็จรูปของประเทศไทย จากการศึกษาสามารถสรุปผลการศึกษา อภิปรายผล และข้อเสนอแนะ ได้ดังนี้

5.1 สรุปผล

5.2 อภิปรายผล

5.3 ข้อเสนอแนะ

5.1 สรุปผล

จากผลการวิเคราะห์ข้อมูลจะพบผลที่สำคัญอย่างหนึ่งคือ การใช้เกณฑ์ในการคัดเลือกตัวแบบโดย MAPE และ RMSE จะให้ผลไม่เป็นไปในทางเดียวกัน ดังนั้นผู้วิจัยจึงจะสรุปผลตัวแบบที่มีประสิทธิภาพของการพยากรณ์สูง เมื่อพิจารณาที่ตัวแบบที่ให้ค่า MAPE หรือ RMSE ต่ำ โดยแยกตามเกณฑ์การคัดเลือกตัวแบบ ดังนี้

พหุน ปณฺ ฑิโต ชีเว

5.1.1 สรุปตัวแบบที่มีประสิทธิภาพของการพยากรณ์สูงเมื่อ y **มีการแจกแจงแกมมา**ที่มีพารามิเตอร์รูปร่าง (α) มีค่าเท่ากับ 2 และพารามิเตอร์ขนาด (β) มีค่าเท่ากับ 1 จำแนกตามสถานการณ์และเกณฑ์การคัดเลือกตัวแบบ

การแจกแจงของ Y	จำนวนตัวแปรอิสระ X	การแจกแจงของ X	ค่าสหสัมพันธ์ระหว่าง X กับ Y ; (r_{xy})	ค่าสหสัมพันธ์ระหว่าง X กับ X ; (r_{xx})	ขนาดตัวอย่าง (n)	เกณฑ์การคัดเลือกตัวแบบ					
						MAPE			RMSE		
						ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง
$Gam(2,1)$	1	$X_1 \sim N(0,1)$	≤ 0.3 (มีความสัมพันธ์กันในระดับต่ำมาก)	-	30	GLM	$\frac{1}{y}$ $\frac{1}{y^2}$	-	GLM GLMs	ไม่แปลง y -	- *Inverse* $\log(\mu)$
					50	GLM	$\frac{1}{y}$ $\frac{1}{y^2}$	-	GLM GLMs	ไม่แปลง y -	- *Inverse* $\log(\mu)$
					100	GLM	$\frac{1}{y}$ $\frac{1}{y^2}$	-	GLM GLMs	ไม่แปลง y -	- *Inverse* $\log(\mu)$

5.1.1 สรุปตัวแบบที่มีประสิทธิภาพของการพยากรณ์สูงเมื่อ y **มีการแจกแจงแกมมา**ที่มีพารามิเตอร์รูปร่าง (α) มีค่าเท่ากับ 2 และพารามิเตอร์ขนาด (β) มีค่าเท่ากับ 1 จำแนกตามสถานการณ์และเกณฑ์การคัดเลือกตัวแบบ (ต่อ)

การแจกแจงของ Y	จำนวนตัวแปรอิสระ X	การแจกแจงของ X	ค่าสหสัมพันธ์ระหว่าง X กับ Y ; (r_{xy})	ค่าสหสัมพันธ์ระหว่าง X กับ X ; (r_{xx})	ขนาดตัวอย่าง (n)	เกณฑ์การคัดเลือกตัวแบบ					
						MAPE			RMSE		
						ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง
$Gam(2,1)$	1	$X_2 \sim N(0,1)$	≥ 0.7 (มีความสัมพันธ์กันในระดับสูง)	-	30	GLM	$\frac{1}{y}$	-	GLMs	-	$\log(\mu)$
					50	GLM	$\frac{1}{y}$	-	GLM	*Normal* *	-
									GLMs	-	$\log(\mu)$
					100	GLM	$\log(y)$	-	GLM	*Normal* *	-
									GLMs	-	$\log(\mu)$

Normal * หมายถึง การแปลงค่า y ให้มีการแจกแจงแบบปรกติมาตรฐานโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ

5.1.1 สรุปตัวแบบที่มีประสิทธิภาพของการพยากรณ์สูงเมื่อ y **มีการแจกแจงแกมมา**ที่มีพารามิเตอร์รูปร่าง (α) มีค่าเท่ากับ 2 และพารามิเตอร์ขนาด (β) มีค่าเท่ากับ 1 จำแนกตามสถานการณ์และเกณฑ์การคัดเลือกตัวแบบ (ต่อ)

การแจกแจงของ Y	จำนวนตัวแปรอิสระ X	การแจกแจงของ X	ค่าสหสัมพันธ์ระหว่าง X กับ Y ; (r_{xy})	ค่าสหสัมพันธ์ระหว่าง X กับ X ; (r_{xx})	ขนาดตัวอย่าง (n)	เกณฑ์การคัดเลือกตัวแบบ					
						MAPE			RMSE		
						ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง
$Gam(2,1)$	2	$X_1 \sim N(0,1)$ $X_2 \sim N(0,1)$	≤ 0.3 ≥ 0.7	≤ 0.3 (มีความสัมพันธ์กันในระดับต่ำมาก)	30	GLM	$\frac{1}{y}$ $\frac{1}{y^2}$ $\log(y)$	-	GLM GLMs	*Normal* * -	- $\log(\mu)$
					50	GLM	$\frac{1}{y}$ $\frac{1}{y^2}$ $\log(y)$	-	GLM GLMs	*Normal* * -	- $\log(\mu)$
					100	GLM	$\log(y)$ *Normal* *	-	GLM GLMs	*Normal* * -	- $\log(\mu)$

Normal * หมายถึง การแปลงค่า y ให้มีการแจกแจงแบบปรกติมาตรฐานโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ

5.1.1 สรุปตัวแบบที่มีประสิทธิภาพของการพยากรณ์สูงเมื่อ y **มีการแจกแจงแกมมา**ที่มีพารามิเตอร์รูปร่าง (α) มีค่าเท่ากับ 2 และพารามิเตอร์ขนาด (β) มีค่าเท่ากับ 1 จำแนกตามสถานการณ์และเกณฑ์การคัดเลือกตัวแบบ (ต่อ)

การแจกแจงของ Y	จำนวนตัวแปรอิสระ X	การแจกแจงของ X	ค่าสหสัมพันธ์ระหว่าง X กับ Y ; (r_{xy})	ค่าสหสัมพันธ์ระหว่าง X กับ X ; (r_{xx})	ขนาดตัวอย่าง (n)	เกณฑ์การคัดเลือกตัวแบบ					
						MAPE			RMSE		
						ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง
$Gam(2,1)$	2	$X_1 \sim N(0,1)$ $X_2 \sim N(0,1)$	≤ 0.3 ≥ 0.7	≥ 0.7 (มีความสัมพันธ์กันในระดับสูง)	30	GLM	$\frac{1}{y}$ $\frac{1}{y^2}$ $\log(y)$	-	GLM GLMs	ไม่แปลง y $\log(y)$ *Normal* * -	- $\log(\mu)$
					50	GLM	$\frac{1}{y}$ $\log(y)$	-	GLM GLMs	ไม่แปลง y $\log(y)$ *Normal* * -	- $\log(\mu)$
					100	GLM	$\frac{1}{y}$ $\log(y)$ *Normal* *	-	GLM	ไม่แปลง y *Normal* *	-

Normal * หมายถึง การแปลงค่า y ให้มีการแจกแจงแบบปรกติมาตรฐานโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ

จากตารางสรุปตัวแบบที่มีประสิทธิภาพของการพยากรณ์ เมื่อ y มีการแจกแจงแกมมาสามารถสรุปผลได้ ดังนี้

กรณีที่ตัวแปรอิสระ 1 ตัวแปร มีความสัมพันธ์กับตัวแปรตาม**น้อยกว่าหรือเท่ากับ 0.3** ขนาดตัวอย่างเท่ากับ 30 50 100 ตามลำดับ พบว่าตัวแบบที่เหมาะสมในการพยากรณ์เมื่อพิจารณาที่ค่า MAPE คือ ตัวแบบเชิงเส้นทั่วไป โดยการแปลงค่า y ให้อยู่ในรูป $\frac{1}{y}$ และแปลงค่าให้อยู่ในรูป $\frac{1}{y^2}$ เมื่อพิจารณาที่ค่า RMSE กลับพบว่าตัวแบบที่เหมาะสมคือ ตัวแบบเชิงเส้นทั่วไปไม่ได้แปลงค่า y รวมทั้งตัวแบบเชิงเส้นนัยทั่วที่มีฟังก์ชันเชื่อมโยงเป็น “inverse” และ “log link” และเมื่อพิจารณาขนาดตัวอย่างไม่มีผลต่อค่า MAPE กับ RMSE

กรณีที่ตัวแปรอิสระ 1 ตัวแปร มีความสัมพันธ์กับตัวแปรตาม**มากกว่าหรือเท่ากับ 0.7** ในกรณีขนาดตัวอย่างเท่ากับ 30 และ 50 พบว่าตัวแบบที่เหมาะสมในการพยากรณ์เมื่อพิจารณาที่ค่า MAPE คือ ตัวแบบเชิงเส้นทั่วไป โดยการแปลงค่า y ให้อยู่ในรูป $\frac{1}{y}$ ในกรณีขนาดตัวอย่างเท่ากับ 100 ค่า MAPE ที่เหมาะสมในการพยากรณ์ คือ ตัวแบบเชิงเส้นทั่วไป โดยการแปลงค่า y ให้อยู่ในรูป $\log(y)$ เมื่อพิจารณาที่ค่า RMSE กลับพบว่าตัวแบบที่เหมาะสม คือ ตัวแบบเชิงเส้นทั่วไปที่แปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ รวมทั้งตัวแบบเชิงเส้นนัยทั่วไปที่มีฟังก์ชันเชื่อมโยงเป็น “log link” และเมื่อพิจารณาขนาดตัวอย่างไม่มีผลต่อค่า MAPE กับ RMSE

กรณีที่ตัวแปรอิสระ 2 ตัวแปร มีความสัมพันธ์กัน**น้อยกว่าหรือเท่ากับ 0.3** และตัวแปรอิสระสัมพันธ์กับตัวแปรตามน้อยกว่าหรือเท่ากับ 0.3 กับ มากกว่าหรือเท่ากับ 0.7 ตามลำดับ และขนาดตัวอย่างเท่ากับ 30 และ 50 พบว่าตัวแบบที่เหมาะสมในการพยากรณ์เมื่อพิจารณาที่ค่า MAPE คือ ตัวแบบเชิงเส้นทั่วไป โดยการแปลงค่า y ให้อยู่ในรูป $\frac{1}{y}$ การแปลงค่า y ให้อยู่ในรูป $\frac{1}{y^2}$ และแปลงค่า y ให้อยู่ในรูป $\log(y)$ ในกรณีขนาดตัวอย่างเท่ากับ 100 ค่า MAPE ที่เหมาะสมในการพยากรณ์ คือ ตัวแบบเชิงเส้นทั่วไป โดยการแปลงค่า y ให้อยู่ในรูป $\log(y)$ และแปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) เมื่อพิจารณาค่า RMSE กลับพบว่าตัวแบบที่เหมาะสม คือ ตัวแบบเชิงเส้นทั่ว โดยการแปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) รวมทั้งตัวแบบเชิงเส้นนัยทั่วไปที่มีฟังก์ชันเชื่อมโยงเป็น “log link” และเมื่อพิจารณาขนาดตัวอย่างไม่มีผลต่อค่า MAPE กับ RMSE

กรณีที่ตัวแปรอิสระ 2 ตัวแปร มีความสัมพันธ์กัน**มากกว่าหรือเท่ากับ 0.7** และตัวแปรอิสระสัมพันธ์กับตัวแปรตามน้อยกว่าหรือเท่ากับ 0.3 กับ มากกว่าหรือเท่ากับ 0.7 ตามลำดับ ในกรณีขนาดตัวอย่างเท่ากับ 30 ค่า MAPE ที่เหมาะสมในการพยากรณ์ คือ ตัวแบบเชิงเส้นทั่วไป

โดยการแปลงค่า y ให้อยู่ในรูป $\frac{1}{y}$ การแปลงค่า y ให้อยู่ ในรูป $\frac{1}{y^2}$ และแปลงค่า y ให้อยู่ในรูป $\log(y)$ ในกรณีขนาดตัวอย่างเท่ากับ 50 ค่า MAPE ที่เหมาะสมในการพยากรณ์ คือ ตัวแบบเชิงเส้นทั่วไป โดยการแปลงค่า y ให้อยู่ในรูป $\frac{1}{y}$ และแปลงค่า y ให้อยู่ในรูป $\log(y)$ ในกรณีขนาดตัวอย่างเท่ากับ 100 ค่า MAPE ที่เหมาะสมในการพยากรณ์ คือ ตัวแบบเชิงเส้นทั่วไป โดยการแปลงค่า y ให้อยู่ในรูป $\frac{1}{y}$ การแปลงค่า y ให้อยู่ในรูป $\log(y)$ และแปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ เมื่อพิจารณาค่า RMSE กลับพบว่าตัวแบบที่เหมาะสม คือ ตัวแบบเชิงเส้นทั่วไปที่ไม่ได้แปลงค่า y การแปลงค่า y ให้อยู่ในรูป $\log(y)$ และแปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ รวมทั้งตัวแบบเชิงเส้นนัยทั่วไปที่มีฟังก์ชันเชื่อมโยงเป็น “log link” และเมื่อพิจารณาขนาดตัวอย่างไม่มีผลต่อค่า MAPE กับ RMSE

พหุน ปณฺฑิโต ชีเว

5.1.2 สรุปตัวแบบที่มีประสิทธิภาพของการพยากรณ์สูงเมื่อ y **มีการแจกแจงอินเวอร์สเกาส์เซียน**ที่มีพารามิเตอร์แสดงตำแหน่ง (μ) มีค่าเท่ากับ 1 และพารามิเตอร์รูปร่าง (λ) มีค่าเท่ากับ 3 จำแนกตามสถานการณ์และเกณฑ์การคัดเลือกตัวแบบ

การแจกแจงของ Y	จำนวนตัวแปรอิสระ X	การแจกแจงของ X	ค่าสหสัมพันธ์ระหว่าง X กับ Y ; (r_{xy})	ค่าสหสัมพันธ์ระหว่าง X กับ X ; (r_{xx})	ขนาดตัวอย่าง (n)	เกณฑ์การคัดเลือกตัวแบบ					
						MAPE			RMSE		
						ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง
$IG(1,3)$	1	$X_1 \sim N(0,1)$	≤ 0.3 (มีความสัมพันธ์กันในระดับต่ำมาก)	-	30	GLM	$\frac{1}{y}$	-	GLM	ไม่แปลง y	-
							$\log(y)$		GLMs	-	$sqrt(\mu)$
							*Normal **				$\log(\mu)$
					50	GLM	$\frac{1}{y}$	-	GLM	ไม่แปลง y	-
							$\log(y)$		GLMs	-	$sqrt(\mu)$
							*Normal **				$\log(\mu)$
$IG(1,3)$	1	$X_1 \sim N(0,1)$	≤ 0.3 (มีความสัมพันธ์กันในระดับต่ำมาก)	-	100	GLM	$\frac{1}{y}$	-	GLM	ไม่แปลง y	-
							$\log(y)$		GLMs	-	$sqrt(\mu)$
							*Normal **				$\log(\mu)$

*Normal ** หมายถึง การแปลงค่า y ให้มีการแจกแจงแบบปรกติมาตรฐานโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ

5.1.2 สรุปตัวแบบที่มีประสิทธิภาพของการพยากรณ์สูงเมื่อ y **มีการแจกแจงอินเวอร์สเกาส์เซียน**ที่มีพารามิเตอร์แสดงตำแหน่ง (μ) มีค่าเท่ากับ 1 และพารามิเตอร์รูปร่าง (λ) มีค่าเท่ากับ 3 จำแนกตามสถานการณ์และเกณฑ์การคัดเลือกตัวแบบ (ต่อ)

การแจกแจงของ Y	จำนวนตัวแปรอิสระ X	การแจกแจงของ X	ค่าสหสัมพันธ์ระหว่าง X กับ Y ; (r_{xy})	ค่าสหสัมพันธ์ระหว่าง X กับ X ; (r_{xx})	ขนาดตัวอย่าง (n)	เกณฑ์การคัดเลือกตัวแบบ					
						MAPE			RMSE		
						ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง
$IG(1,3)$	1	$X_2 \sim N(0,1)$	≥ 0.7 (มีความสัมพันธ์กันในระดับสูง)	-	30	GLM	$\frac{1}{y}$ $\frac{1}{y^2}$ $\log(y)$	-	GLM GLMs	ไม่แปลง y $\log(y)$ -	- $\log(\mu)$
					50	GLM	$\frac{1}{y}$ $\log(y)$	-	GLM GLMs	ไม่แปลง y $\log(y)$ -	- $\log(\mu)$
					100	GLM	$\frac{1}{y}$ $\log(y)$	-	GLM GLMs	$\log(y)$ -	- $\log(\mu)$

Normal * หมายถึง การแปลงค่า y ให้มีการแจกแจงแบบปรกติมาตรฐานโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ

5.1.2 สรุปตัวแบบที่มีประสิทธิภาพของการพยากรณ์สูงเมื่อ y **มีการแจกแจงอินเวอร์สเกาส์เซียน**ที่มีพารามิเตอร์แสดงตำแหน่ง (μ) มีค่าเท่ากับ 1 และพารามิเตอร์รูปร่าง (λ) มีค่าเท่ากับ 3 จำแนกตามสถานการณ์และเกณฑ์การคัดเลือกตัวแบบ (ต่อ)

การแจกแจงของ Y	จำนวนตัวแปรอิสระ X	การแจกแจงของ X	ค่าสหสัมพันธ์ระหว่าง X กับ Y ; (r_{xy})	ค่าสหสัมพันธ์ระหว่าง X กับ X ; (r_{xx})	ขนาดตัวอย่าง (n)	เกณฑ์การคัดเลือกตัวแบบ					
						MAPE			RMSE		
						ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง
$IG(1,3)$	2	$X_1 \sim N(0,1)$ $X_2 \sim N(0,1)$	≤ 0.3 ≥ 0.7	≤ 0.3 (มีความสัมพันธ์กันในระดับต่ำ)	30	GLM	$\frac{1}{y}$ $\log(y)$ *Normal **	- -	GLM GLMs	$\log(y)$ *Normal ** -	- $\log(\mu)$
					50	GLM	$\frac{1}{y}$ $\log(y)$ *Normal **	-	GLM GLMs	$\log(y)$ *Normal ** -	- $\log(\mu)$
					100	GLM	$\frac{1}{y}$ $\log(y)$ *Normal **	-	GLM GLMs	$\log(y)$ *Normal ** -	- $\log(\mu)$

*Normal ** หมายถึง การแปลงค่า y ให้มีการแจกแจงแบบปรกติมาตรฐานโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ

5.1.2 สรุปตัวแบบที่มีประสิทธิภาพของการพยากรณ์สูงเมื่อ y **มีการแจกแจงอินเวอร์สเกาส์เซียน**ที่มีพารามิเตอร์แสดงตำแหน่ง (μ) มีค่าเท่ากับ 1 และพารามิเตอร์รูปร่าง (λ) มีค่าเท่ากับ 3 จำแนกตามสถานการณ์และเกณฑ์การคัดเลือกตัวแบบ (ต่อ)

การแจกแจงของ Y	จำนวนตัวแปรอิสระ X	การแจกแจงของ X	ค่าสหสัมพันธ์ระหว่าง X กับ Y ; (r_{xy})	ค่าสหสัมพันธ์ระหว่าง X กับ X ; (r_{xx})	ขนาดตัวอย่าง (n)	เกณฑ์การคัดเลือกตัวแบบ					
						MAPE			RMSE		
						ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง	ตัวแบบ	รูปแบบการแปลงค่า y	ฟังก์ชันเชื่อมโยง
$IG(1,3)$	2	$X_1 \sim N(0,1)$ $X_2 \sim N(0,1)$	≤ 0.3 ≥ 0.7	≥ 0.7 (มีความสัมพันธ์กันในระดับสูง)	30	GLM	$\frac{1}{y}$ $\log(y)$ *Normal **	- -	GLM GLMs	$\log(y)$ *Normal ** -	- $\log(\mu)$
					50	GLM	$\frac{1}{y^2}$ $\log(y)$ *Normal **	-	GLM GLMs	$\log(y)$ *Normal ** -	- $\log(\mu)$
					100	GLM	$\log(y)$ *Normal **	-	GLM GLMs	$\log(y)$ *Normal ** -	- $\log(\mu)$

*Normal ** หมายถึง การแปลงค่า y ให้มีการแจกแจงแบบปรกติมาตรฐานโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ

จากตารางสรุปตัวแบบที่มีประสิทธิภาพของการพยากรณ์ เมื่อ y มีการแจกแจงอินเวอร์สเกาส์เซียนสามารถสรุปผลได้ ดังนี้

กรณีที่**ตัวแปรอิสระ 1 ตัวแปร** มีความสัมพันธ์กับตัวแปรตาม**น้อยกว่าหรือเท่ากับ 0.3** และขนาดตัวอย่างเท่ากับ 30 50 100 ตามลำดับ พบว่าตัวแบบที่เหมาะสมในการพยากรณ์เมื่อพิจารณาที่ค่า MAPE คือ ตัวแบบเชิงเส้นทั่วไป โดยการแปลงค่า y ให้อยู่ในรูป $\frac{1}{y}$ การแปลงค่า y ให้อยู่ในรูป $\log(y)$ และแปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ เมื่อพิจารณาค่า RMSE กลับพบว่าตัวแบบที่เหมาะสม คือ ตัวแบบเชิงเส้นทั่วไปที่ไม่ได้แปลงค่า y รวมทั้งตัวแบบเชิงเส้นนัยทั่วไปที่มีฟังก์ชันเชื่อมโยงเป็น “log link” กับ “square root” และเมื่อพิจารณาขนาดตัวอย่างไม่มีผลต่อค่า MAPE กับ RMSE

กรณีที่**ตัวแปรอิสระ 1 ตัวแปร** มีความสัมพันธ์กับตัวแปรตาม**มากกว่าหรือเท่ากับ 0.7** ในกรณีที่ขนาดตัวอย่างเท่ากับ 30 พบว่าตัวแบบที่เหมาะสมในการพยากรณ์เมื่อพิจารณาที่ค่า MAPE คือ ตัวแบบเชิงเส้นทั่วไป โดยการแปลงค่า y ให้อยู่ในรูป $\frac{1}{y}$ การแปลงค่า y ให้อยู่ในรูป $\frac{1}{y^2}$ และแปลงค่า y ให้อยู่ในรูป $\log(y)$ ในกรณีที่ขนาดตัวอย่างเท่ากับ 50 และ 100 ค่า MAPE ที่เหมาะสมในการพยากรณ์ คือ ตัวแบบเชิงเส้นทั่วไป โดยการแปลงค่า y ให้อยู่ในรูป $\frac{1}{y}$ และแปลงค่า y ให้อยู่ในรูป $\log(y)$ เมื่อพิจารณาค่า RMSE กลับพบว่าตัวแบบที่เหมาะสม คือ ตัวแบบเชิงเส้นทั่วไปที่ไม่ได้แปลงค่า y และแปลงค่า y ให้อยู่ในรูป $\log(y)$ รวมทั้งตัวแบบเชิงเส้นนัยทั่วไปที่มีฟังก์ชันเชื่อมโยงเป็น “log link” และเมื่อพิจารณาขนาดตัวอย่างไม่มีผลต่อค่า MAPE กับ RMSE

กรณีที่**ตัวแปรอิสระ 2 ตัวแปร** มีความสัมพันธ์กัน**น้อยกว่าหรือเท่ากับ 0.3** และตัวแปรอิสระสัมพันธ์กับตัวแปรตามน้อยกว่าหรือเท่ากับ 0.3 กับ มากกว่าหรือเท่ากับ 0.7 ตามลำดับ และขนาดตัวอย่างเท่ากับ 30 50 100 ตามลำดับ พบว่าตัวแบบที่เหมาะสมในการพยากรณ์เมื่อพิจารณาที่ค่า MAPE คือ ตัวแบบเชิงเส้นทั่วไป โดยการแปลงค่า y ให้อยู่ในรูป $\frac{1}{y}$ การแปลงค่า y ให้อยู่ในรูป $\log(y)$ และแปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ เมื่อพิจารณาค่า RMSE กลับพบว่าตัวแบบที่เหมาะสม คือ ตัวแบบเชิงเส้นทั่วไป โดยการแปลงค่า y ให้อยู่ในรูป $\log(y)$ และแปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ รวมทั้งตัวแบบเชิงเส้นนัยทั่วไปที่มีฟังก์ชันเชื่อมโยงเป็น “log link” และเมื่อพิจารณาขนาดตัวอย่างไม่มีผลต่อค่า MAPE กับ RMSE

กรณีที่**ตัวแปรอิสระ 2 ตัวแปร** มีความสัมพันธ์กัน**มากกว่าหรือเท่ากับ 0.7** และตัวแปรอิสระสัมพันธ์กับตัวแปรตามน้อยกว่าหรือเท่ากับ 0.3 กับ มากกว่าหรือเท่ากับ 0.7 ตามลำดับ ในกรณีตัวอย่างเท่ากับ 30 พบว่าตัวแบบที่เหมาะสมในการพยากรณ์เมื่อพิจารณาที่ค่า MAPE คือ ตัวแบบเชิงเส้นทั่วไป โดยการแปลงค่า y ให้อยู่ในรูป $\frac{1}{y}$ การแปลงค่า y ให้อยู่ในรูป $\log(y)$ และแปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ ในกรณีตัวอย่างเท่ากับ 50 พบว่าตัวแบบที่เหมาะสมในการพยากรณ์เมื่อพิจารณาที่ค่า MAPE คือ ตัวแบบเชิงเส้นทั่วไป โดยการแปลงค่า y ให้อยู่ในรูป $\frac{1}{y^2}$ การแปลงค่า y ให้อยู่ในรูป $\log(y)$ และแปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ ในกรณีตัวอย่างเท่ากับ 100 พบว่าตัวแบบที่เหมาะสมในการพยากรณ์เมื่อพิจารณาที่ค่า MAPE คือ ตัวแบบเชิงเส้นทั่วไป โดยการแปลงค่า y ให้อยู่ในรูป $\log(y)$ และแปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ เมื่อพิจารณาค่า RMSE กลับพบว่าตัวแบบที่เหมาะสม คือ ตัวแบบเชิงเส้นทั่วไป โดยการแปลงค่า y ให้อยู่ในรูป $\log(y)$ และแปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ รวมทั้งตัวแบบเชิงเส้นนัยทั่วไปที่มีฟังก์ชันเชื่อมโยงเป็น “log link” และเมื่อพิจารณาขนาดตัวอย่างไม่มีผลต่อค่า MAPE กับ RMSE

5.1.3 สรุปผลตัวแบบที่เหมาะสมในการพยากรณ์และสมการพยากรณ์ปริมาณการใช้น้ำมันสำเร็จรูปของประเทศไทย

ชนิดของน้ำมันสำเร็จรูป	เกณฑ์การคัดเลือกตัวแบบ						สมการ
	MAPE			RMSE			
	ตัวแบบ	รูปแบบการแปลงค่า y	การแจกแจงของ y (ฟังก์ชันเชื่อมโยง)	ตัวแบบ	รูปแบบการแปลงค่า y	การแจกแจงของ y (ฟังก์ชันเชื่อมโยง)	
ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91	GLM	ไม่แปลง y	-	GLM	ไม่แปลง y	-	$\hat{G91} = 325.72772 - 1.84787(P_Ethanol) - 6.32264(EXC) + 3.14346(P_G95) - 2.68144(P_B95) + 0.20289(G95) + 1.74737(E20)$
ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95	GLM	ไม่แปลง y	-	GLM	ไม่แปลง y	-	$\hat{G95} = 192.5080 - 26.6521(P_G95) + 1.8926(P_Ethanol) + 10.6764(P_E20) + 12.3691(P_B95) - 0.9674(E85) + 2.3789(B95)$
ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20 *	GLMs	-	การแจกแจงแกมมา (link ="inverse")	GLMs	-	การแจกแจงแกมมา (link ="inverse")	$\frac{1}{\hat{\mu}_{E20}} = 0.1526 - 0.001220(P_Ethanol) + 0.002620(P_G95) - 0.002763(P_B95) - 0.0006810(G91) + 0.0003625(G95)$
ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20 **	GLMs	-	การแจกแจงแกมมา (link = "log")	GLMs	-	การแจกแจงแกมมา (link = "log")	$\log(\hat{\mu}_{E20}) = 2.7422814 - 0.0476477(P_E20) + 0.0109384(EXC) + 0.0042112(WCO) + 0.0391259(P_G95) + 0.0069715(G91) - 0.0152211(B95)$

* ข้อมูลตั้งแต่มกราคม พ.ศ.2553 ถึง กุมภาพันธ์ พ.ศ.2556 ** ข้อมูลตั้งแต่มีนาคม พ.ศ.2556 ถึง ธันวาคม พ.ศ.2558

5.1.3 สรุปผลตัวแบบที่เหมาะสมในการพยากรณ์และสมการพยากรณ์ปริมาณการใช้น้ำมันสำเร็จรูปของประเทศไทย (ต่อ)

ชนิดของน้ำมันสำเร็จรูป	เกณฑ์การคัดเลือกตัวแบบ						สมการ
	MAPE			RMSE			
	ตัวแบบ	รูปแบบการแปลงค่า y	การแจกแจงของ y (ฟังก์ชันเชื่อมโยง)	ตัวแบบ	รูปแบบการแปลงค่า y	การแจกแจงของ y (ฟังก์ชันเชื่อมโยง)	
ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E85	GLM	*q_normal*	-	GLM	*q_normal*	-	$\hat{E85} = -11.834081 + 0.664537(P_G91) - 0.887061(P_G95) + 0.428169(P_B95) + 0.090261(G91) - 0.074425(G95) + 0.154602(E20)$
ปริมาณการใช้น้ำมันเบนซิน 91	GLM	ไม่แปลง y	-	GLM	ไม่แปลง y	-	$\hat{B91} = 133.2913 - 14.6544(P_B91) + 0.9172(WCO) + 15.6165(P_G91) + 11.4914(P_G95) - 17.5601(P_E20) + 0.5483(G91) - 0.8934(E20) + 34.7601(P_LPG) - 15.9177(P_NGV)$
ปริมาณการใช้น้ำมันเบนซิน 95	GLM	$\log(y)$	-	-	-	-	$\log(\hat{B95}) = -0.6324247 - 0.2445370(P_B95) + 0.5792734(P_G95) - 0.2472373(P_E20) + 0.0171556(G95) + 0.0328272(E85) - 0.0064584(LPG)$
	-	-	-	GLM	ไม่แปลง y	-	$\hat{B95} = -40.44375 - 3.03114(P_B95) + 5.54552(P_G95) - 2.84934(P_E85) + 0.14041(G95) - 0.43299(E85) + 0.35024(G91) - 0.06663(LPG) - 0.13616(NGV) + 6.92681(P_LPG) - 5.66885(P_NGV)$

q_normal : แปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ

5.1.3 สรุปผลตัวแบบที่เหมาะสมในการพยากรณ์และสมการพยากรณ์ปริมาณการใช้น้ำมันสำเร็จรูปของประเทศไทย (ต่อ)

ชนิดของน้ำมันสำเร็จรูป	เกณฑ์การคัดเลือกตัวแบบ						สมการ
	MAPE			RMSE			
	ตัวแบบ	รูปแบบการแปลงค่า y	การแจกแจงของ y (ฟังก์ชันเชื่อมโยง)	ตัวแบบ	รูปแบบการแปลงค่า y	การแจกแจงของ y (ฟังก์ชันเชื่อมโยง)	
ปริมาณการใช้แก๊ส LPG	GLMs	-	การแจกแจงอินเวอร์สเกาส์เซียน (link ="$\frac{1}{\mu^2}$")	GLMs	-	การแจกแจงอินเวอร์สเกาส์เซียน (link ="$\frac{1}{\mu^2}$")	$\frac{1}{\hat{\mu}_{LPG}^2} = 0.0000005927 + 0.0000001664(EXC) + 0.000000008683(P_B95) - 0.0000003694(P_E20) + 0.0000003446(P_E85) + 0.0000005197(P_NGV) + 0.00000006312(B95) - 0.00000003466(G91) + 0.00000003693(E85)$
ปริมาณการใช้แก๊ส NGV	GLM	ไม่แปลง y	-	GLM	ไม่แปลง y	-	$\hat{NGV} = -34.1751 + 4.7870(P_G91) - 5.4784(P_E85) + 4.7439(P_LPG) - 0.7915(B95) + 1.3914(G91) - 0.2932(G95) - 1.0227(E20)$

จากตารางสรุปผลตัวแบบที่เหมาะสมในการพยากรณ์และสมการพยากรณ์ปริมาณการใช้น้ำมันสำเร็จรูปของประเทศไทยสามารถสรุปผลได้ ดังนี้

ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91 ตัวแบบที่เหมาะสมในการพยากรณ์เมื่อพิจารณาทั้งค่า MAPE และ RMSE คือ ตัวแบบเชิงเส้นทั่วไปเมื่อไม่แปลงค่า y โดยมีสมการพยากรณ์ คือ

$$\hat{G91} = 325.72772 - 1.84787(P_Ethanol) - 6.32264(EXC) + 3.14346(P_G95) - 2.68144(P_B95) + 0.20289(G95) + 1.74737(E20)$$

ปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95 ตัวแบบที่เหมาะสมในการพยากรณ์เมื่อพิจารณาทั้งค่า MAPE และ RMSE คือ ตัวแบบเชิงเส้นทั่วไปเมื่อไม่แปลงค่า y โดยมีสมการพยากรณ์ คือ

$$\hat{G95} = 192.5080 - 26.6521(P_G95) + 1.8926(P_Ethanol) + 10.6764(P_E20) + 12.3691(P_B95) - 0.9674(E85) + 2.3789(B95)$$

ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20 ข้อมูลตั้งแต่มกราคม พ.ศ.2553 ถึง กุมภาพันธ์ พ.ศ.2556 ตัวแบบที่เหมาะสมในการพยากรณ์เมื่อพิจารณาทั้งค่า MAPE และ RMSE คือ ตัวแบบเชิงเส้นทั่วไปเมื่อไม่แปลงค่า y โดยมีสมการพยากรณ์ คือ

$$\frac{1}{\hat{\mu}_{E20}} = 0.1526 - 0.001220(P_Ethanol) + 0.002620(P_G95) - 0.002763(P_B95) - 0.0006810(G91) + 0.0003625(G95)$$

ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20 ข้อมูลตั้งแต่มีนาคม พ.ศ.2556 ถึง ธันวาคม พ.ศ. 2558 ตัวแบบที่เหมาะสมในการพยากรณ์เมื่อพิจารณาค่า MAPE และ RMSE คือ ตัวแบบเชิงเส้นนัยทั่วไป เมื่อ y มีการแจกแจงแกมมาที่มีฟังก์ชันเชื่อมโยงเป็น “log link” โดยมีสมการพยากรณ์ คือ

$$\log(\hat{\mu}_{E20}) = 2.7422814 - 0.0476477(P_E20) + 0.0109384(EXC) + 0.0042112(WCO) + 0.0391259(P_G95) + 0.0069715(G91) - 0.0152211(B95)$$

ปริมาณการใช้น้ำมันแก๊สโซฮอล์ E85 ตัวแบบที่เหมาะสมในการพยากรณ์เมื่อพิจารณาค่า MAPE และ RMSE คือ ตัวแบบเชิงเส้นทั่วไป โดยการแปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติ โดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ โดยมีสมการพยากรณ์ คือ

$$\hat{E85} = -11.834081 + 0.664537(P_G91) - 0.887061(P_G95) + 0.428169(P_B95) + 0.090261(G91) - 0.074425(G95) + 0.154602(E20)$$

ปริมาณการใช้น้ำมันเบนซิน 91 ตัวแบบที่เหมาะสมในการพยากรณ์เมื่อพิจารณาทั้งค่า MAPE และ RMSE คือ ตัวแบบเชิงเส้นทั่วไปเมื่อไม่แปลงค่า y โดยมีสมการพยากรณ์ คือ

$$\begin{aligned}\hat{B91} &= 133.2913 - 14.6544(P_B91) + 0.9172(WCO) + 15.6165(P_G91) \\ &\quad + 11.4914(P_G95) - 17.5601(P_E20) + 0.5483(G91) - 0.8934(E20) \\ &\quad + 34.7601(P_LPG) - 15.9177(P_NGV)\end{aligned}$$

ปริมาณการใช้น้ำมันเบนซิน 95 ตัวแบบที่เหมาะสมในการพยากรณ์เมื่อพิจารณาทั้งค่า MAPE คือ ตัวแบบเชิงเส้นทั่วไปเมื่อแปลงค่า y อยู่ในรูป $\log(y)$ โดยมีสมการพยากรณ์ คือ

$$\begin{aligned}\log(\hat{B95}) &= -0.6324247 - 0.2445370(P_B95) + 0.5792734(P_G95) \\ &\quad - 0.2472373(P_E20) + 0.0171556(G95) + 0.0328272(E85) \\ &\quad - 0.0064584(LPG)\end{aligned}$$

แต่เมื่อพิจารณาค่า RMSE กลับพบว่าตัวแบบเชิงเส้นทั่วไปเมื่อไม่แปลงค่า y โดยมีสมการพยากรณ์ คือ

$$\begin{aligned}\hat{B95} &= -40.44375 - 3.03114(P_B95) + 5.54552(P_G95) - 2.84934(P_E85) \\ &\quad + 0.14041(G95) - 0.43299(E85) + 0.35024(G91) - 0.06663(LPG) \\ &\quad - 0.13616(NGV) + 6.92681(P_LPG) - 5.66885(P_NGV)\end{aligned}$$

ปริมาณการใช้แก๊ส LPG ตัวแบบที่เหมาะสมในการพยากรณ์เมื่อพิจารณาทั้งค่า MAPE และ RMSE คือ ตัวแบบเชิงเส้นนัยทั่วไป เมื่อ y มีการแจกแจงอินเวอร์สเกาส์เซียนที่มีฟังก์ชันเชื่อมโยงเป็น “ $\frac{1}{\mu^2}$ ”โดยมีสมการพยากรณ์ คือ

$$\begin{aligned}\frac{1}{\hat{\mu}^2_{LPG}} &= 0.0000005927 + 0.0000001664(EXC) + 0.000000008683(P_B95) \\ &\quad - 0.0000003694(P_E20) + 0.0000003446(P_E85) \\ &\quad + 0.0000005197(P_NGV) + 0.00000006312(B95) \\ &\quad - 0.00000003466(G91) + 0.00000003693(E85)\end{aligned}$$

ปริมาณการใช้แก๊ส NGV ตัวแบบที่เหมาะสมในการพยากรณ์เมื่อพิจารณาทั้งค่า MAPE และ RMSE คือ ตัวแบบเชิงเส้นทั่วไปเมื่อไม่แปลงค่า y โดยมีสมการพยากรณ์ คือ

$$\begin{aligned}\hat{NGV} &= -34.1751 + 4.7870(P_G91) - 5.4784(P_E85) + 4.7439(P_LPG) \\ &\quad - 0.7915(B95) + 1.3914(G91) - 0.2932(G95) - 1.0227(E20)\end{aligned}$$

5.2 อภิปรายผล

5.2.1 จากผลการจำลองข้อมูลตามสถานการณ์ต่างๆ ผลการเปรียบเทียบค่า MAPE กับ RMSE ไม่ได้สอดคล้องกันซึ่งโดยส่วนใหญ่การที่ใช้คนละเกณฑ์ที่เกี่ยวข้องกับค่าเฉลี่ยของความคลาดเคลื่อนในการคัดเลือกตัวแบบควรจะได้ตัวแบบที่ดีที่สุดเป็นตัวแบบเดียวกัน ทั้งนี้กรณีที่ไม่สอดคล้องกันอาจจะเป็นเพราะว่าค่าความคลาดเคลื่อนหรือ y มีค่าที่ผิดปรกติ (Robert N., 2018) ส่งผลให้ค่า MAPE กับ RMSE ไม่ได้สอดคล้องกัน อาจจะต้องพิจารณาเกณฑ์ในการตัดสินใจอย่างเช่น AIC หรือ Deviance ในการประกอบการตัดสินใจเพิ่มเติม

5.2.2 งานวิจัยเล่มนี้ในส่วนการจำลองข้อมูลตามสถานการณ์ต่างๆ พบว่าค่า MAPE ที่คำนวณได้มีค่าค่อนข้างสูง อาจจะเป็นเพราะสูตรในการคำนวณค่า MAPE ตัวหารเป็นข้อมูลจริงของตัวแปรตามที่มีการแจกแจงแกมมา เมื่อค่า y ที่ได้มีค่าเข้าใกล้ 0 จะส่งผลทำให้ค่า MAPE มีค่ามาก เมื่อพิจารณาตัวแปรตามที่มีการแจกแจงอินเวอร์สเกาส์เซียน ค่า MAPE ก็มีค่าที่ค่อนข้างสูงเช่นเดียวกัน แต่ไม่สูงมากเหมือนกับการจำลองตัวแปรตามที่มีการแจกแจงแกมมา

5.2.3 จากผลการนำข้อมูลมาประยุกต์ใช้กับข้อมูลปริมาณการใช้น้ำมันสำเร็จรูปของประเทศไทย พบว่าตัวแบบ GLM ที่ไม่ได้แปลงค่า y การแปลงค่า y ให้อยู่ในรูป $\frac{1}{y}$ การแปลงค่าให้อยู่ในรูป $\frac{1}{y^2}$ การแปลงค่า y ให้อยู่ในรูป $\log(y)$ และแปลงค่า y ให้อยู่ในรูปการแจกแจงปรกติโดยใช้ฟังก์ชันควอนไทล์ (quantile) ของการแจกแจงปรกติ มีค่า MAPE กับ RMSE ใกล้เคียงกันกับตัวแบบ GLMs เมื่อ y มีการแจกแจงแกมมาและการแจกแจงอินเวอร์สเกาส์เซียน ทำให้ตัวแบบเชิงเส้นนัยทั่วไปอาจจะเหมาะสมสำหรับในการพยากรณ์ เนื่องจากตัวแบบ GLM เมื่อมีการแปลงค่าให้อยู่ในรูปแบบต่างๆ ต้องมีการแปลงค่ากลับก่อนที่จะนำไปใช้งานซึ่งค่อนข้างยุ่งยากในการแปลงค่ามากกว่าการเลือกใช้ตัวแบบ GLMs

5.2.4 จากผลที่ได้จะสังเกตได้ว่าค่า MAPE กับ RMSE ในตัวแบบ GLM ที่ไม่ได้แปลงค่า y กับตัวแบบ GLMs ตัวแบบทั้งสองไม่มีความแตกต่างกัน ซึ่งจะสอดคล้องกับงานวิจัย (Myers and Montgomery, 1997) เรื่อง A Tutorial on Generalized Linear Models ที่นำเสนอขั้นตอนในการผลิต semiconductors ซึ่งมี 4 ปัจจัยที่มีผลต่อ resistivity โดยใช้ตัวแบบ GLM ในการแปลงค่า y ให้อยู่ในรูป log และคัดเลือกตัวแบบโดยใช้เกณฑ์ AIC และลองใช้ตัวแบบ GLMs ที่มีการแจกแจงแกมมา ที่มีฟังก์ชันเชื่อมโยงเป็น “log link” เพื่อให้สอดคล้องกับตัวแบบ GLM และคัดเลือกตัวแบบโดยใช้เกณฑ์ AIC พบว่าค่าสัมประสิทธิ์ทั้งตัวแบบ GLM และตัวแบบ GLMs มีความคล้ายคลึงกัน แม้แต่ค่าความคลาดเคลื่อนก็เกือบจะเหมือนกัน และค่า square root ของ dispersion สอดคล้องกับ residual standard error ของตัวแบบ GLM เนื่องจากค่า $\alpha = \frac{1}{\phi}$; ϕ คือ dispersion

parameter ที่มีขนาดใหญ่ การแจกแจงแกมมานั้นใกล้เคียงกับการแจกแจงปรกติ ในทำนองเดียวกันกับตัวแปรตอบสนองในตัวแบบเชิงเส้น การแจกแจงแบบ log normal ที่มีความแปรปรวนขนาดเล็กก็สามารถประมาณค่าที่ได้เช่นเดียวกับการแจกแจงแบบปรกติ ด้วยเหตุนี้ทำให้ตัวแบบ GLM กับ GLMs จึงไม่มีความแตกต่างกัน ซึ่งจะสอดคล้องกับงานวิจัยของ (Hallin and Ingenbleek, 1983) เรื่อง The Swedish automobile portfolio in 1977 ข้อมูลเกี่ยวกับการจ่ายค่าสินไหมทดแทนสำหรับพื้นที่ต่างๆ ในประเทศสวีเดนในปี 1977 ข้อมูลจะถูกแบ่งย่อยโดยระยะทางที่รถวิ่ง เงินจากการไม่ใช้สิทธิ์การเคลมครั้งก่อน และประเภทของรถยนต์ มีการวิเคราะห์ในปีกรมธรรม์ภายในแต่ละกลุ่ม เนื่องจากคาดว่ายอดรวมของค่าสินไหมทดแทนสำหรับกลุ่มจะเป็นสัดส่วนกับจำนวนของผู้เอาประกัน โดยใช้ตัวแบบ GLMs ที่มีการแจกแจงแบบแกมมาใช้ฟังก์ชันเชื่อมโยงเป็น “log link”

5.3 ข้อเสนอแนะ

5.3.1 ข้อเสนอแนะจากการวิจัย

5.3.1.1 จากผลการจำลองข้อมูลตามสถานการณ์ที่กำหนดให้ เมื่อเปรียบเทียบประสิทธิภาพความแม่นยำของตัวแบบเชิงเส้นทั่วไปกับตัวแบบเชิงเส้นนัยทั่วไปให้ค่า MAPE ไม่ค่อยแตกต่างกันทั้ง 2 ตัวแบบ เมื่อพิจารณาที่ค่า RMSE กลับพบว่าค่าที่ได้ก็ไม่ต่างกันมากเช่นเดียวกัน นั้นหมายความว่าสามารถที่จะใช้ตัวแบบเชิงเส้นนัยทั่วไปในการหาสมการที่เหมาะสมในการพยากรณ์ซึ่งง่ายกว่าการใช้ตัวแบบเชิงเส้นทั่วไปที่ต้องแปลงข้อมูลและต้องแปลงค่ากลับเมื่อจะนำไปใช้ และต้องตรวจสอบข้อมูลให้เป็นไปตามข้อตกลงเบื้องต้น

5.3.1.2 ข้อมูลที่ใช้ในการพยากรณ์ปริมาณการใช้น้ำมันสำเร็จรูปของประเทศไทยตั้งแต่ เดือนมกราคม 2553 ถึง เดือนธันวาคม 2558 จำนวน 72 เดือน ซึ่งข้อมูลที่ได้บางเดือนหรือบางปีจะมีปัญหาเพราะข้อมูลที่ได้มาตัวเลขค่อนข้างมีการกระโดดทำให้ข้อมูลผิดปรกติอาจเป็นเพราะการนำเข้าน้ำมันดิบจากต่างประเทศซึ่งราคาที่ซื้อ - ขาย เป็นราคาที่ถูกกำหนดโดยกลไกลในตลาด โดยตลาดกลางซื้อ - ขาย น้ำมันที่ใกล้ประเทศไทยมากที่สุดและเป็นศูนย์กลางการซื้อ - ขายในภูมิภาคเอเชียก็คือตลาดสิงคโปร์ และยังมีการอุดหนุนราคาน้ำมันจาก รัฐบาล และบางชนิดน้ำมันมีการยกเลิกการผลิตอย่างเช่นน้ำมันเบนซิน 91 ที่มีการยกเลิกใช้เมื่อวันที่ 1 มกราคม 2556 เพื่อสนับสนุนและส่งเสริมให้ผู้ใช้รถหันมาใช้พลังงานทดแทนประเภทน้ำมันแก๊สโซฮอล์ให้มากขึ้น ซึ่งจะทำให้มีการใช้เอทานอลที่ทำมาจากพืชผลทางการเกษตรเพิ่มมากขึ้น (ข้อมูลจากสำนักงานนโยบายและแผนพลังงานกระทรวงพลังงาน) ทำให้ค่าที่พยากรณ์ที่ได้มานำมาใช้ประโยชน์ได้ไม่ดีเท่าที่ควร

5.3.2 ข้อเสนอแนะเพื่อการวิจัยครั้งต่อไป

5.3.2.1 จากผลการจำลองข้อมูลตัวแบบเชิงเส้นทั่วไปกับตัวแบบเชิงเส้นนัยทั่วไปให้ค่าที่ได้ไม่แตกต่างกันมากเมื่อข้อมูลไม่ได้มีการแจกแจงปรกติ ดังนั้นในงานวิจัยครั้งต่อไปแนะนำศึกษากรณีที่ Y มีความแปรปรวนไม่คงที่ซึ่งตัวแบบเชิงเส้นนัยทั่วไปอาจให้ผลที่ดีกว่า

5.3.2.2 ในการประยุกต์ใช้ข้อมูลโดยการนำปริมาณการใช้น้ำมันสำเร็จรูปในประเทศไทยมาพยากรณ์ ในงานครั้งต่อไปแนะนำศึกษาตัว แปรอิสระเพิ่มเติมที่ส่งผลต่อปริมาณการใช้งานน้ำมันสำเร็จรูป เช่น ราคาน้ำมันดิบสิงคโปร์ ราคาหน้าโรงกลั่น ราคาขายส่ง ราคาขายปลีก อัตราภาษี สรรพสามิต กองทุนน้ำมันเชื้อเพลง เป็นต้น

พหุน ปณ ฑิโต ชีเว

บรรณานุกรม

บรรณานุกรม

กมลวรรณ สารพานิช. (2555). การพยากรณ์ราคาน้ำมันดิบล่วงหน้าในตลาดฟิวเจอรส์ไนแม็กซ์โดยวิธีอารีมาและอารีแม็กซ์. มหาวิทยาลัยธุรกิจบัณฑิตย์.

ฐิติรัตน์ แพทย์มงคล. (2552). ปัจจัยที่มีผลต่อการปรับตัวของราคาน้ำมันดีเซลในประเทศไทย. กรุงเทพฯ มหาวิทยาลัยหอการค้าไทย.

ณัฐภัทร แก้วรัตนภัทร์ และคณะ. (2555). การเปรียบเทียบประสิทธิภาพทางเทคนิคเหมืองข้อมูลเพื่อแทนค่าสูญหาย. 561–567.

ทรงศิริ แต้สมบัติ. (2549). การพยากรณ์เชิงปริมาณ. กรุงเทพฯ: สำนักพิมพ์เกษตรศาสตร์.

ธิดาเดียว มยุรีสวรรค์. (2559). การวิเคราะห์การถดถอย. ขอนแก่น: บริษัทเพ็ญพรินติ้ง จำกัด.

ธีระศักดิ์ อุรัจนานนท์. (2546). ความน่าจะเป็นและสถิติประยุกต์. กรุงเทพฯ: สกายบุ๊กส์.

บังอร กุมพล. (2544). การวิเคราะห์การถดถอย. มหาสารคาม: มหาวิทยาลัยมหาสารคาม.

พรพิมล ศรีประเสริฐรัตน์. (2549). ผลกระทบของการเปลี่ยนแปลงราคาน้ำมัน ต่อต้นทุนการผลิตภาคอุตสาหกรรมของไทย. มหาวิทยาลัยธุรกิจบัณฑิตย์.

มารู้จักน้ำมันเชื้อเพลิง. (2560). คลังความรู้SciMath. Retrieved from https://www.scimath.org/article-science/item/7473-2017-09-08-03-40-23

มุกดา แม้นมินทร์. (2549). การเลือกตัวแบบเมื่อมีตัวแบบที่เหมาะสมมากกว่าหนึ่งตัวแบบ. กรุงเทพฯ: สำนักพิมพ์ประกายพรึก.

รจนาถ จันทร์มี. (2555). การเปรียบเทียบช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยของประชากรที่มีการแจกแจงแบบเบ้ขวา โดยใช้ค่ามัธยฐานของตัวอย่าง. มหาวิทยาลัยบูรพา.

รามนรี ภูดีบุตร. (2550). การพยากรณ์การใช้ปิโตเลียมของประเทศไทย. มหาวิทยาลัยขอนแก่น.

เรณู นิยมเดชา. (2554). การคำนวณอัตราเบี้ยประกันอัคคีภัยที่อยู่อาศัยด้วยตัวแบบเชิงเส้นนัยทั่วไป. สถาบันบัณฑิตพัฒนบริหารศาสตร์.

วรางคณา กีรติวิบูลย์. (2556). ตัวแบบพยากรณ์ราคาขายปลีกน้ำมันแก๊สโซฮอล์ 95 ในเขตกรุงเทพมหานครและปริมณฑล. วารสารวิจัยและพัฒนา มจธ, *36*(4), 423–438.

ศูนย์ข้อมูล & ข่าวสืบสวนเพื่อสิทธิพลเมือง. (2557). เหตุใดราคาน้ำมันของไทย ต้องอิงกับราคาตลาดสิงคโปร์. Retrieved from http://tcijthai.com/tcijthainews/view.php?ids=5196

ศูนย์พยากรณ์และสารสนเทศพลังงาน. (2556). สถานการณ์พลังงานไทยปี 2556. Retrieved from http://doc-eppo.eppo.go.th/EnergySituation/EnerSituation_YF.htm

สัณฐิติ ทองช่วง และคณะ. (2553). การวิเคราะห์ปริมาณการใช้แก๊สโซฮอล์ E20 ในประเทศไทย. *BU Academic Review*, *9*(1685–4322), 47–55.

สำรวม จงเจริญ. (2551). ตัวแบบเชิงเส้น. กรุงเทพฯ: โครงการส่งเสริมและพัฒนาเอกสารวิชาการ สถาบันพัฒนบริหารศาสตร์.

สุจิตตา สุระภี. (2560). การวิเคราะห์ข้อมูลเชิงกลุ่ม. มหาสารคาม: มหาวิทยาลัยมหาสารคาม.

สุนีย์ สัมมาทัต. (2557). การวิเคราะห์เชิงพื้นที่ทางคณิตศาสตร์สำหรับสร้างแผนที่โรคหลายชนิดประยุกต์ใช้กับโรคความดันโลหิตสูงและโรคหัวใจขาดเลือด. มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร.

Agresti A. (2009). An introduction to categorical data analysis. In *Statistics in Medicine* (Vol. 28). https://doi.org/10.1002/sim.3564

Faraway J. (2016). *Extending the Linear Model with R*. Taylor & Francis Inc.

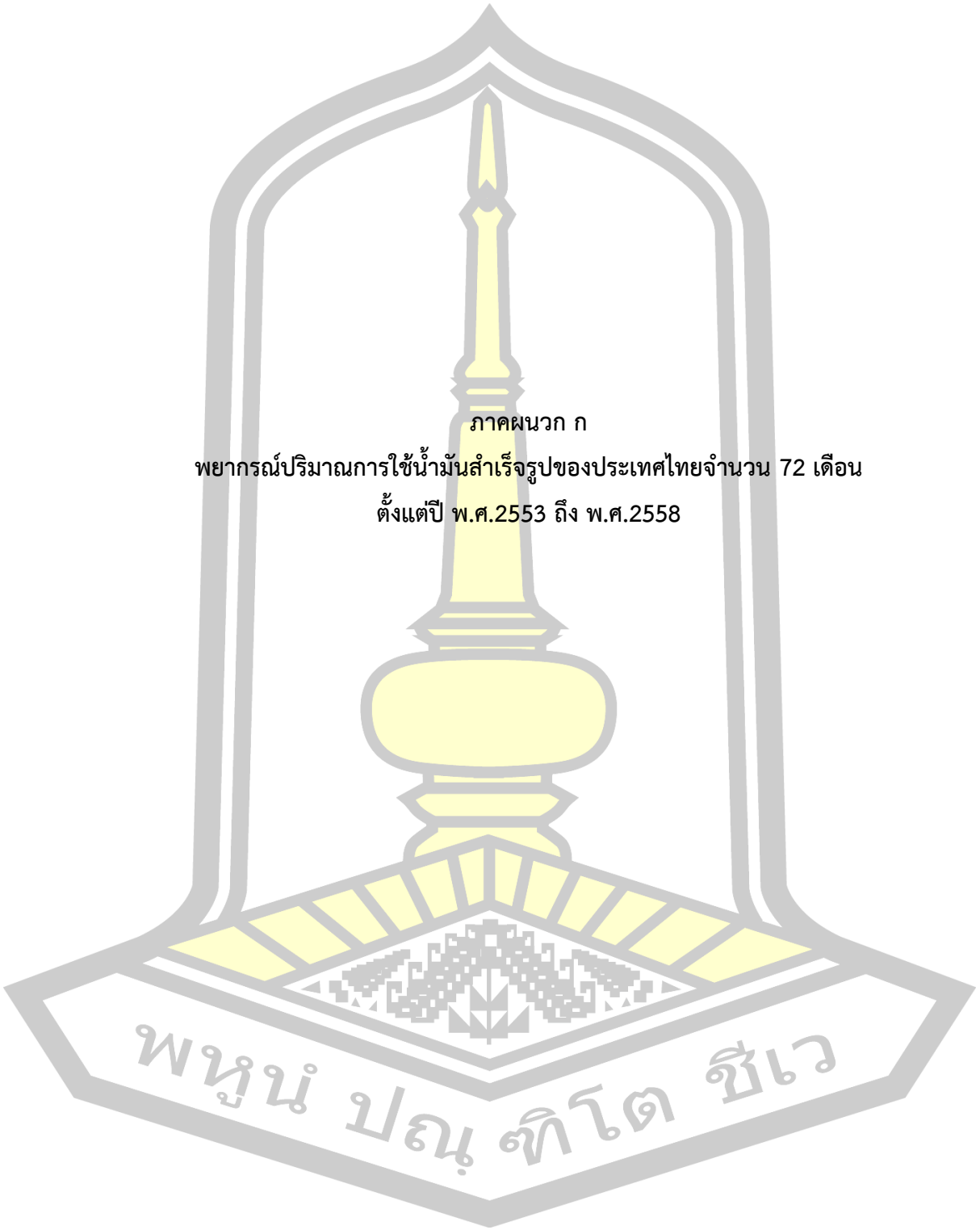
Fatemeh R.A. and Youset A.G. (2011). Artificial Neural Network and stepwise multiple range regression methods for prediction of tractor fuel consumption. *Measurement: Journal of the International Measurement Confederation*, *44*(10), 2104–2111. https://doi.org/10.1016/j.measurement.2011.08.006

Hinkle, D.E, William, W. and S. G. . (1998). *Applied Statistics for the Behavior Sciences*. New York: Houghton Mifflin.

Lewis, C. D. (1982). International and Business Forecasting Methods. *JOURNAL ARTICLE*, *6*(3), 2–3.

Zong W.G. and William E. (2010). Energy demand estimation of South Korea using artificial neural network. *Energy Policy*, *38*(10), 6379–6380. https://doi.org/10.1016/j.enpol.2009.08.070

ภาคผนวก



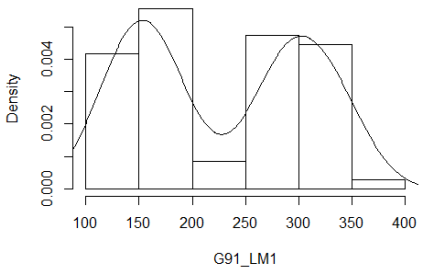
ภาคผนวก ก

พยากรณ์ปริมาณการใช้น้ำมันสำเร็จรูปของประเทศไทยจำนวน 72 เดือน

ตั้งแต่ปี พ.ศ.2553 ถึง พ.ศ.2558

อิสโตแกรมและสัมประสิทธิ์ความเบ้แสดงการแจกแจงของ y และการแจกแจงของความคลาดเคลื่อนรวมทั้งสถิติทดสอบความเป็นอิสระของความคลาดเคลื่อนของปริมาณการใช้น้ำมันแก๊สโซฮอล์ 91 โดยใช้วิธีการวิเคราะห์ตัวแบบเชิงเส้นทั่วไป (General Linear Model) และวิธีการวิเคราะห์ตัวแบบเชิงเส้นนัยทั่วไป (Generalized Linear Model)

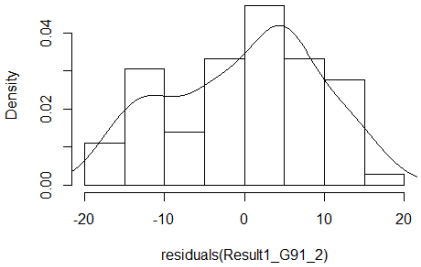
Consumption of Gasohol 91 (y_{G91})



Skewness = 0.0615

GLM

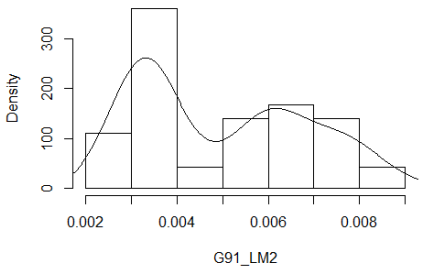
Residual Consumption of Gasohol 91 (y_{G91})



Skewness = - 0.2043

Durbin-Watson test = 0.51986

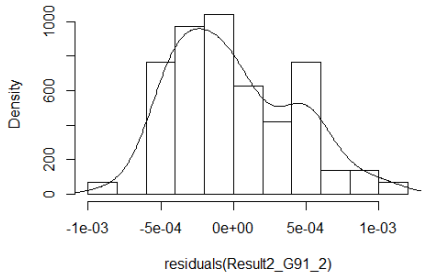
Consumption of Gasohol 91 $\left(\frac{1}{y_{G91}}\right)$



Skewness = 0.3850

GLM

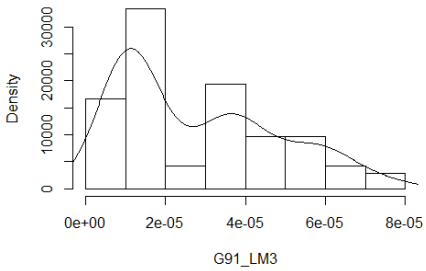
Residual Consumption of Gasohol 91 $\left(\frac{1}{y_{G91}}\right)$



Skewness = 0.5220

Durbin-Watson test = 0.61339

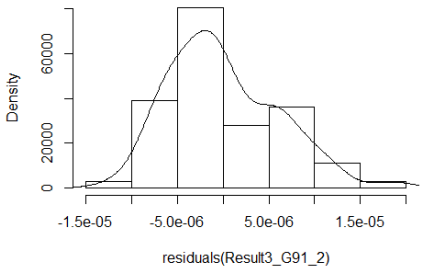
Consumption of Gasohol 91 $\left(\frac{1}{y_{G91}^2}\right)$



Skewness = 0.6822

GLM

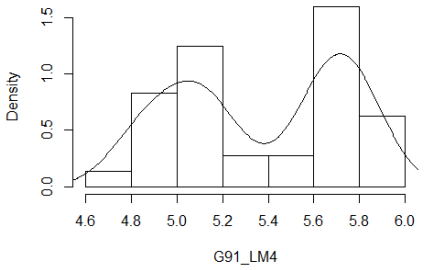
Residual Consumption of Gasohol 91 $\left(\frac{1}{y_{G91}^2}\right)$



Skewness = 0.6601

Durbin-Watson test = 0.5791

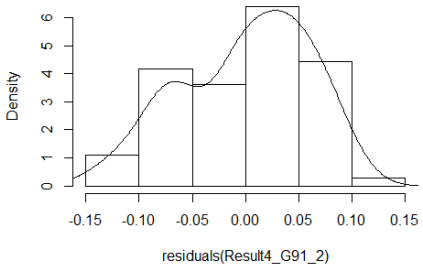
Consumption of Gasohol 91 $\left(\log\left(y_{G91}\right)\right)$



Skewness = - 0.1416

GLM

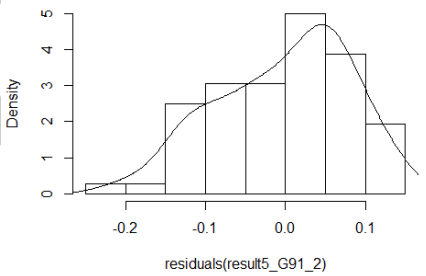
Residual Consumption of Gasohol 91 $\left(\log\left(y_{G91}\right)\right)$



Skewness = - 0.3940

Durbin-Watson test = 0.6171

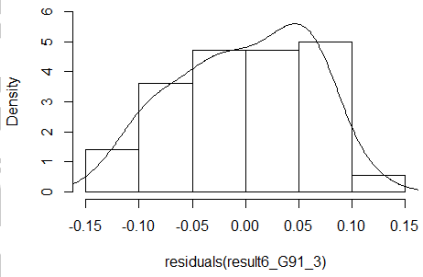
Residual Consumption of Gasohol 91 (y_{G91})
GLMs , family = Gamma(link = "inverse")



Skewness = - 0.3968

Durbin-Watson test = 0.45854

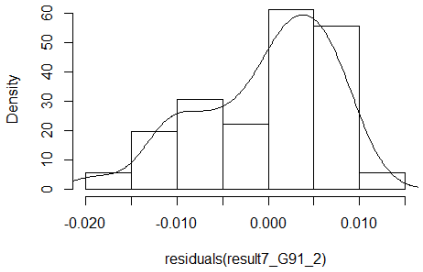
Residual Consumption of Gasohol 91 (y_{G91})
GLMs , family = Gamma(link = "log")



Skewness = - 0.2472

Durbin-Watson test = 0.53737

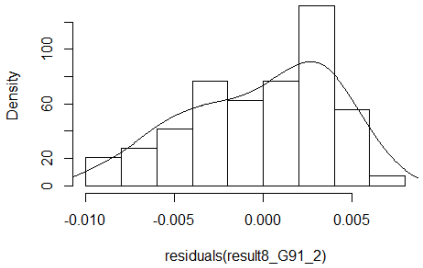
Residual Consumption of Gasohol 91 (y_{G91})
GLMs , family = inverse.gaussian(link = $\frac{1}{\mu^2}$)



Skewness = - 0.6931

Durbin-Watson test = 0.49911

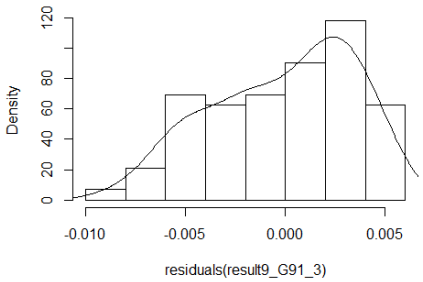
Residual Consumption of Gasohol 91 (y_{G91})
GLMs , family = inverse.gaussian(link = "log")



Skewness = - 0.3830

Durbin-Watson test = 0.5747

Residual Consumption of Gasohol 91 (y_{G91})
GLMs , family = inverse.gaussian(link = "sqrt")

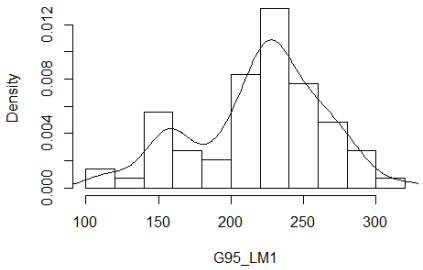


Skewness = - 0.4047

Durbin-Watson test = 0.5747

อิสโตแกรมและสัมประสิทธิ์ความเบ้แสดงการแจกแจงของ y และการแจกแจงของความคลาดเคลื่อนรวมทั้งสถิติทดสอบความเป็นอิสระของความคลาดเคลื่อนของปริมาณการใช้น้ำมันแก๊สโซฮอล์ 95 โดยใช้วิธีการวิเคราะห์ตัวแบบเชิงเส้นทั่วไป (General Linear Model) และวิธีการวิเคราะห์ตัวแบบเชิงเส้นนัยทั่วไป (Generalized Linear Model)

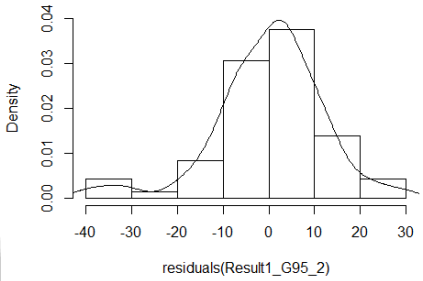
Consumption of Gasohol 95 (y_{G95})



Skewness = - 0.441

GLM

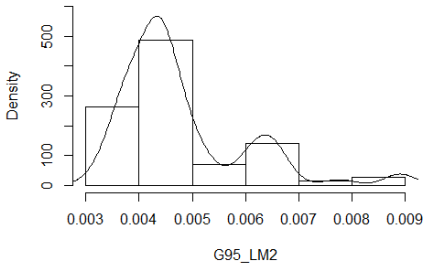
Residual Consumption of Gasohol 95 (y_{G95})



Skewness = - 0.7305

Durbin-Watson test = 1.169

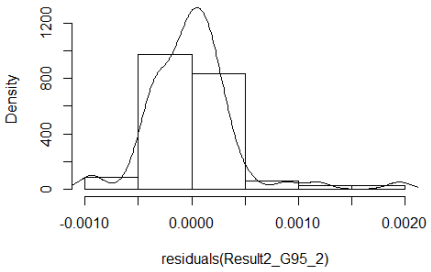
Consumption of Gasohol 95 $\left(\frac{1}{y_{G95}}\right)$



Skewness = 1.5533

GLM

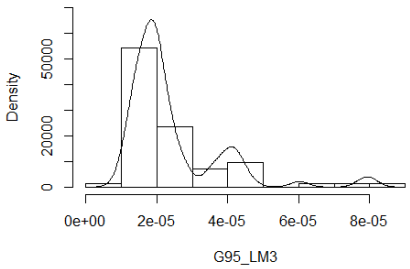
Residual Consumption of Gasohol 95 $\left(\frac{1}{y_{G95}}\right)$



Skewness = 1.5693

Durbin-Watson test = 1.3646

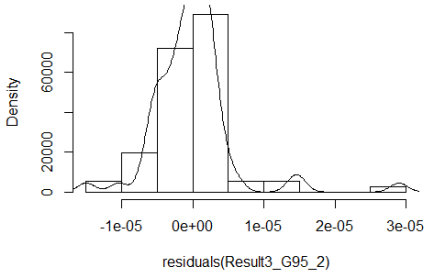
Consumption of Gasohol 95 $\left(\frac{1}{y_{G95}^2}\right)$



Skewness = 2.2096

GLM

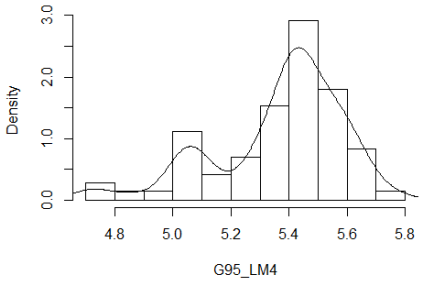
Residual Consumption of Gasohol 95 $\left(\frac{1}{y_{G95}^2}\right)$



Skewness = 2.0481

Durbin-Watson test = 1.4946

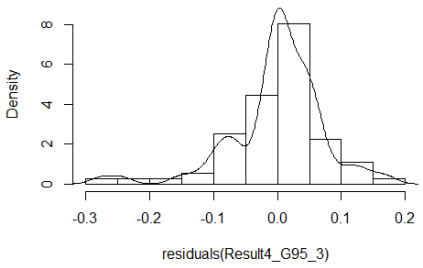
Consumption of Gasohol 95 $\left(\log\left(y_{G95}\right)\right)$



Skewness = - 0.9713

GLM

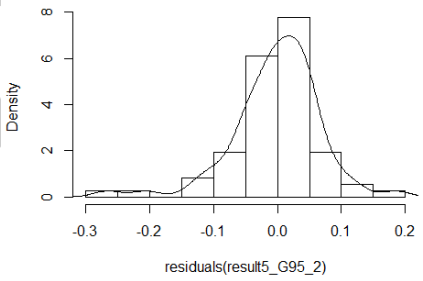
Residual Consumption of Gasohol 95 $\left(\log\left(y_{G95}\right)\right)$



Skewness = - 1.0756

Durbin-Watson test = 0.96203

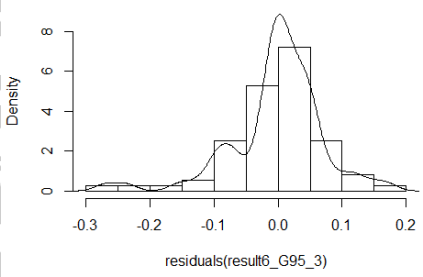
Residual Consumption of Gasohol 95 (y_{G95})
GLMs , family = Gamma(link = "inverse")



Skewness = - 0.9103

Durbin-Watson test = 1.1361

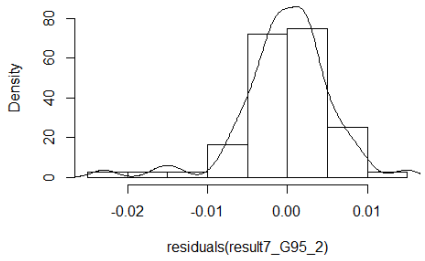
Residual Consumption of Gasohol 95 (y_{G95})
GLMs , family = Gamma(link = "log")



Skewness = - 1.009

Durbin-Watson test = 0.83339

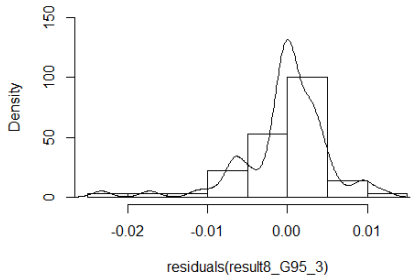
Residual Consumption of Gasohol 95 (y_{G95})
GLMs , family = inverse.gaussian(link = $\frac{1}{\mu^2}$)



Skewness = - 1.0011

Durbin-Watson test = 1.1361

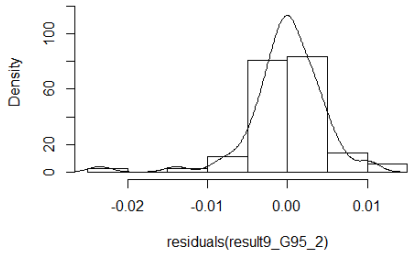
Residual Consumption of Gasohol 95 (y_{G95})
GLMs , family = inverse.gaussian(link = "log")



Skewness = - 1.3401

Durbin-Watson test = 0.83339

Residual Consumption of Gasohol 95 (y_{G95})
GLMs , family = inverse.gaussian(link = "sqrt")

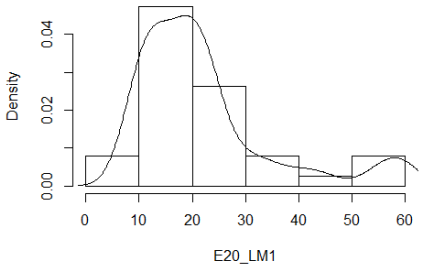


Skewness = - 1.689

Durbin-Watson test = 1.169

อิสโตแกรมและสัมประสิทธิ์ความเบ้แสดงการแจกแจงของ y และการแจกแจงของความคลาดเคลื่อนรวมทั้งสถิติทดสอบความเป็นอิสระของความคลาดเคลื่อนของปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20 โดยใช้วิธีการวิเคราะห์ตัวแบบเชิงเส้นทั่วไป (General Linear Model) และวิธีการวิเคราะห์ตัวแบบเชิงเส้นนัยทั่วไป (Generalized Linear Model) ข้อมูลตั้งแต่ มกราคม พ.ศ.2553 ถึง กุมภาพันธ์ พ.ศ.2556

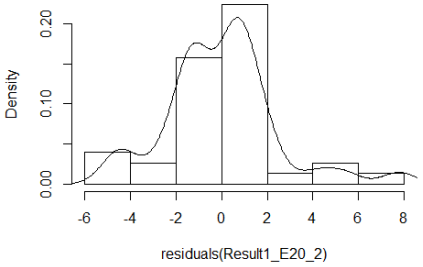
Consumption of Gasohol E20 (y_{E20})



Skewness = 1.6189

GLM

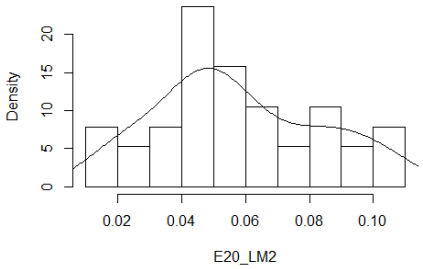
Residual Consumption of Gasohol E20 (y_{E20})



Skewness = 0.7209

Durbin-Watson test = 1.2333

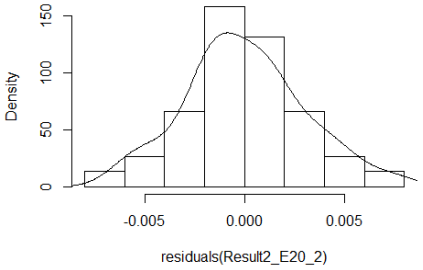
Consumption of Gasohol E20 $\left(\frac{1}{y_{E20}}\right)$



Skewness = 0.3282

GLM

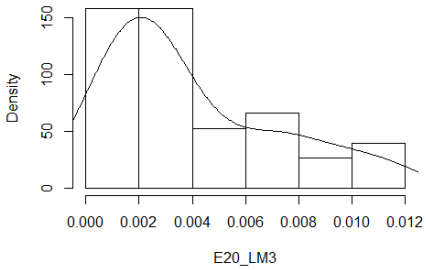
Residual Consumption of Gasohol E20 $\left(\frac{1}{y_{E20}}\right)$



Skewness = 0.2319

Durbin-Watson test = 1.6876

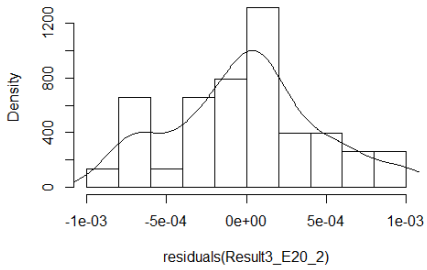
Consumption of Gasohol E20 $\left(\frac{1}{y_{E20}^2}\right)$



Skewness = 0.9386

GLM

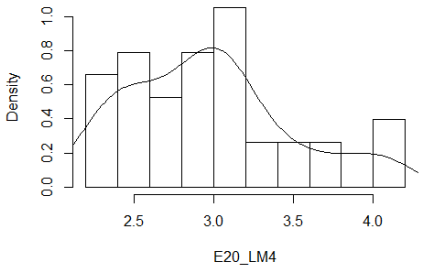
Residual Consumption of Gasohol E20 $\left(\frac{1}{y_{E20}^2}\right)$



Skewness = 0.1597

Durbin-Watson test = 1.7735

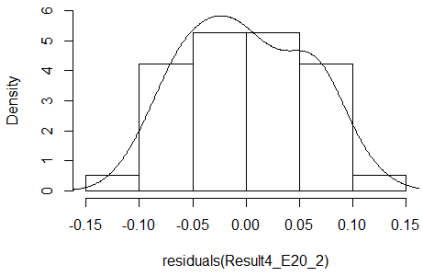
Consumption of Gasohol E20 $\left(\log\left(y_{E20}\right)\right)$



Skewness = 0.5999

GLM

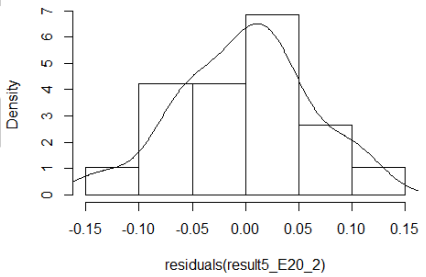
Residual Consumption of Gasohol E20 $\left(\log\left(y_{E20}\right)\right)$



Skewness = 0.1118

Durbin-Watson test = 0.8084

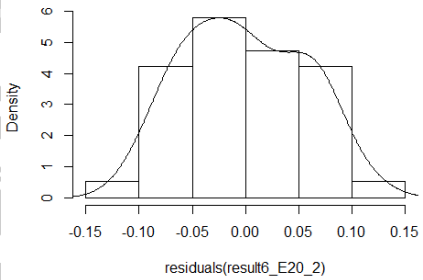
Residual Consumption of Gasohol E20 (y_{E20})
GLMs , family = Gamma(link = "inverse")



Skewness = - 0.1713

Durbin-Watson test = 0.9305

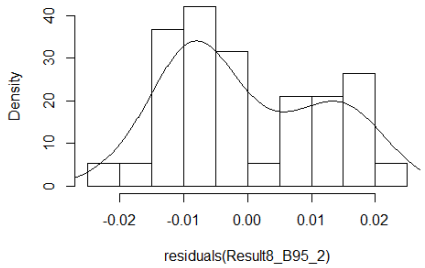
Residual Consumption of Gasohol E20 (y_{E20})
GLMs , family = Gamma(link = "log")



Skewness = 0.1179

Durbin-Watson test = 0.7658

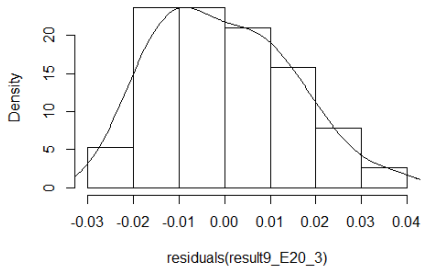
Residual Consumption of Gasohol E20 (y_{E20})
GLMs , family = inverse.gaussian(link = "log")



Skewness = 0.3437

Durbin-Watson test = 0.76581

Residual Consumption of Gasohol E20 (y_{E20})
GLMs , family = inverse.gaussian(link = "sqrt")

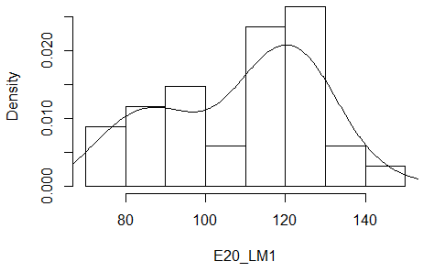


Skewness = 0.4301

Durbin-Watson test = 1.177

อิสโตแกรมและสัมประสิทธิ์ความเบ้แสดงการแจกแจงของ y และการแจกแจงของความคลาดเคลื่อนรวมทั้งสถิติทดสอบความเป็นอิสระของความคลาดเคลื่อนของปริมาณการใช้น้ำมันแก๊สโซฮอล์ E20 โดยใช้วิธีการวิเคราะห์ตัวแบบเชิงเส้นทั่วไป (General Linear Model) และวิธีการวิเคราะห์ตัวแบบเชิงเส้นนัยทั่วไป (Generalized Linear Model) ข้อมูลตั้งแต่ มีนาคม พ.ศ.2556 ถึง ธันวาคม พ.ศ.2558

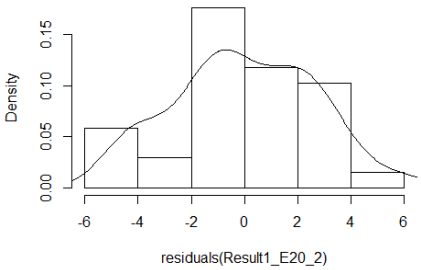
Consumption of Gasohol E20 (y_{E20})



Skewness = - 0.3449

GLM

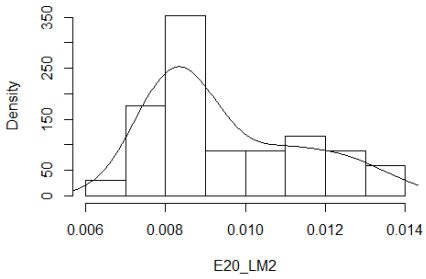
Residual Consumption of Gasohol E20 (y_{E20})



Skewness = 0.0305

Durbin-Watson test = 1.1859

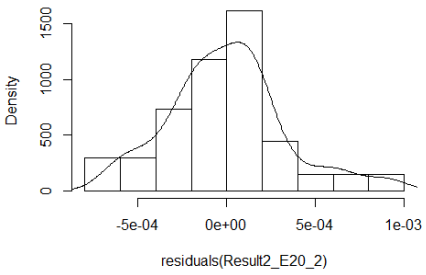
Consumption of Gasohol E20 $\left(\frac{1}{y_{E20}}\right)$



Skewness = 0.8003

GLM

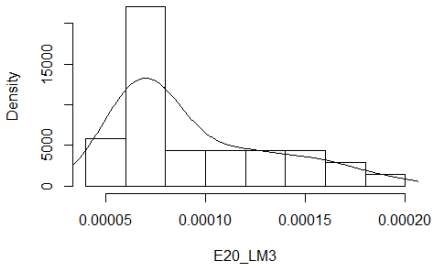
Residual Consumption of Gasohol E20 $\left(\frac{1}{y_{E20}}\right)$



Skewness = 0.4835

Durbin-Watson test = 1.1464

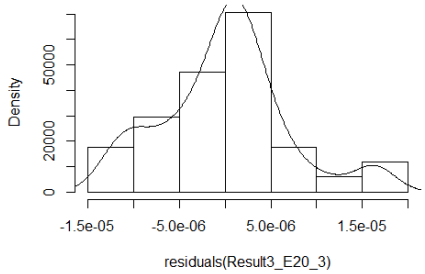
Consumption of Gasohol E20 $\left(\frac{1}{y_{E20}^2}\right)$



Skewness = 1.0217

GLM

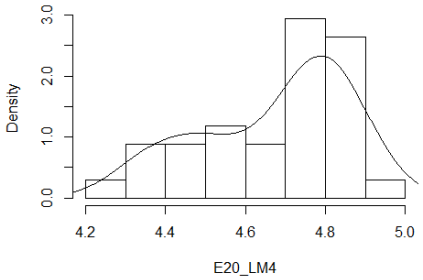
Residual Consumption of Gasohol E20 $\left(\frac{1}{y_{E20}^2}\right)$



Skewness = 0.3351

Durbin-Watson test = 1.5819

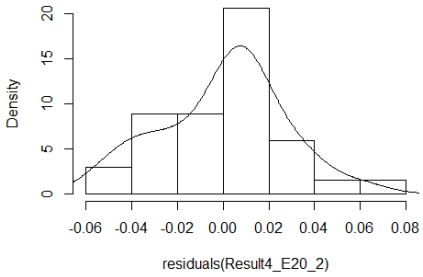
Consumption of Gasohol E20 $\left(\log\left(y_{E20}\right)\right)$



Skewness = - 0.5769

GLM

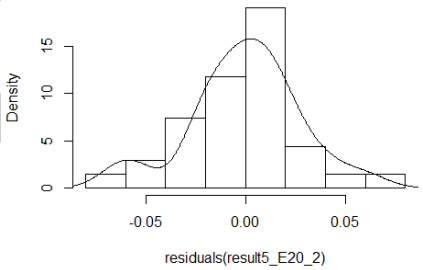
Residual Consumption of Gasohol E20 $\left(\log\left(y_{E20}\right)\right)$



Skewness = - 0.15

Durbin-Watson test = 1.1905

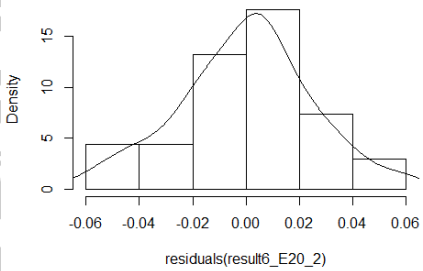
Residual Consumption of Gasohol E20 (y_{E20})
GLMs , family = Gamma(link = "inverse")



Skewness = - 0.2865

Durbin-Watson test = 1.1859

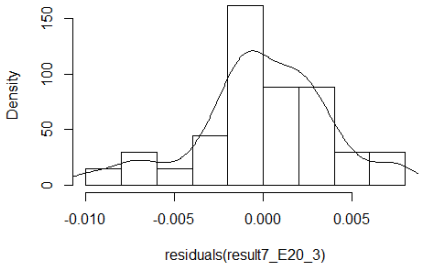
Residual Consumption of Gasohol E20 (y_{E20})
GLMs , family = Gamma(link = "log")



Skewness = - 0.068

Durbin-Watson test = 1.1859

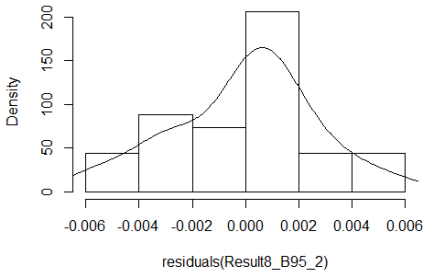
Residual Consumption of Gasohol E20 (y_{E20})
GLMs , family = inverse.gaussian(link = $\frac{1}{\mu^2}$)



Skewness = - 0.4262

Durbin-Watson test = 0.7965

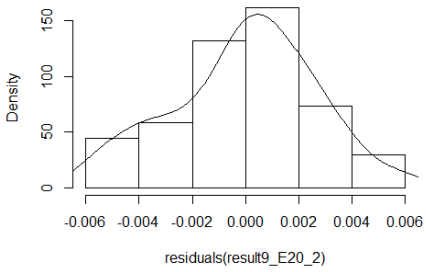
Residual Consumption of Gasohol E20 (y_{E20})
GLMs , family = inverse.gaussian(link = "log")



Skewness = - 0.2309

Durbin-Watson test = 1.0451

Residual Consumption of Gasohol E20 (y_{E20})
GLMs , family = inverse.gaussian(link = "sqrt")

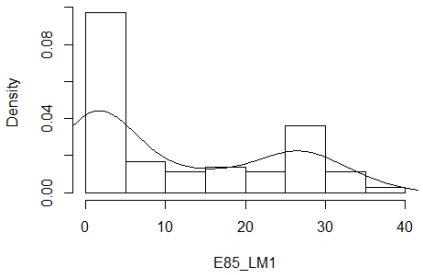


Skewness = - 0.2294

Durbin-Watson test = 1.0451

อิสโตแกรมและสัมประสิทธิ์ความเบ้แสดงการแจกแจงของ y และการแจกแจงของความคลาดเคลื่อนรวมทั้งสถิติทดสอบความเป็นอิสระของความคลาดเคลื่อนของปริมาณการใช้น้ำมันแก๊สโซฮอล์ E85 โดยใช้วิธีการวิเคราะห์ตัวแบบเชิงเส้นทั่วไป (General Linear Model) และวิธีการวิเคราะห์ตัวแบบเชิงเส้นนัยทั่วไป (Generalized Linear Model)

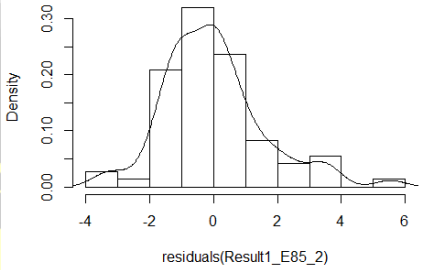
Consumption of Gasohol E85 (y_{E85})



Skewness = 0.5365

GLM

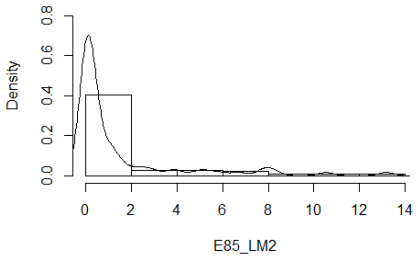
Residual Consumption of Gasohol E85 (y_{E85})



Skewness = 0.8256

Durbin-Watson test = 0.8225

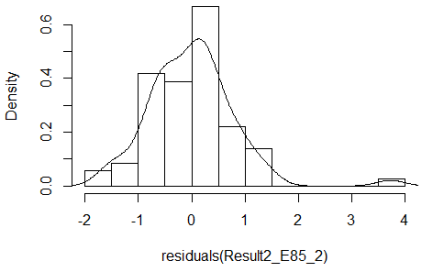
Consumption of Gasohol E85 $\left(\frac{1}{y_{E85}}\right)$



Skewness = 2.423

GLM

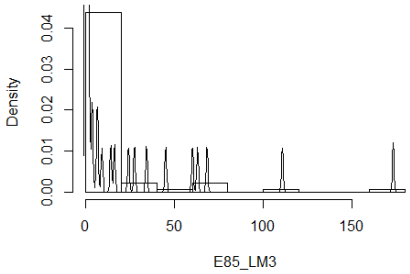
Residual Consumption of Gasohol E85 $\left(\frac{1}{y_{E85}}\right)$



Skewness = 1.1849

Durbin-Watson test = 1.0949

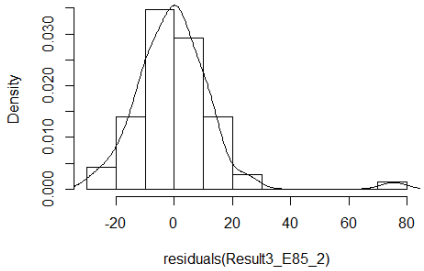
Consumption of Gasohol E85 $\left(\frac{1}{y_{E85}^2}\right)$



Skewness = 4.0985

GLM

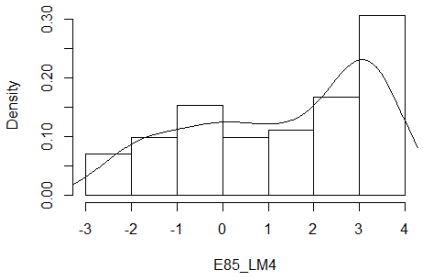
Residual Consumption of Gasohol E85 $\left(\frac{1}{y_{E85}^2}\right)$



Skewness = 2.0825

Durbin-Watson test = 0.92246

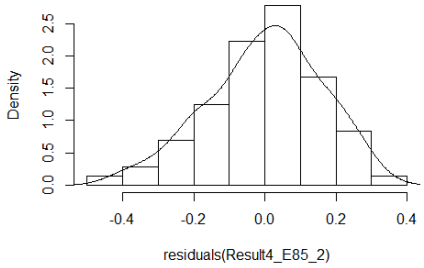
Consumption of Gasohol E85 $\left(\log\left(y_{E85}\right)\right)$



Skewness = - 0.4584

GLM

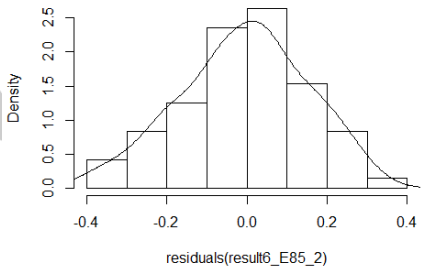
Residual Consumption of Gasohol E85 $\left(\log\left(y_{E85}\right)\right)$



Skewness = - 0.3211

Durbin-Watson test = 1.2415

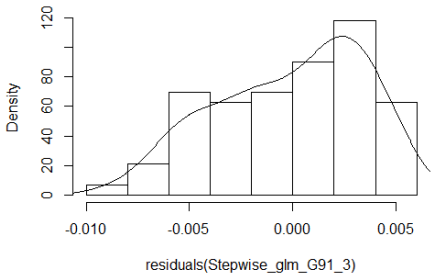
Residual Consumption of Gasohol E85 (y_{E85})
GLMs , family = Gamma(link = "log")



Skewness = - 0.2177

Durbin-Watson test = 0.1328

Residual Consumption of Gasohol E85 (y_{E85})
GLMs , family = inverse.gaussian(link = "sqrt")



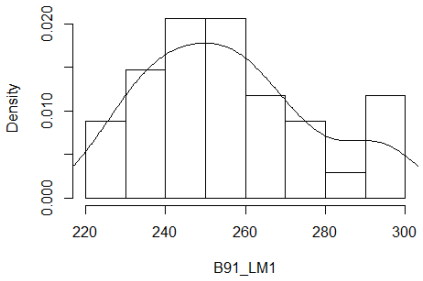
Skewness = 0.0369

Durbin-Watson test = 0.93748

พหุน ปณ ฑิโต ชีเว

อิสโตแกรมและสัมประสิทธิ์ความเบ้แสดงการแจกแจงของ y และการแจกแจงของความคลาดเคลื่อนรวมทั้งสถิติทดสอบความเป็นอิสระของความคลาดเคลื่อนของปริมาณการใช้น้ำมันเบนซิน 91 โดยใช้วิธีการวิเคราะห์ตัวแบบเชิงเส้นทั่วไป (General Linear Model) และวิธีการวิเคราะห์ตัวแบบเชิงเส้นนัยทั่วไป (Generalized Linear Model)

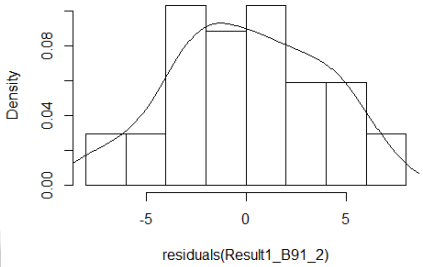
Consumption of Gasoline 91 (y_{B91})



Skewness = 0.4727

GLM

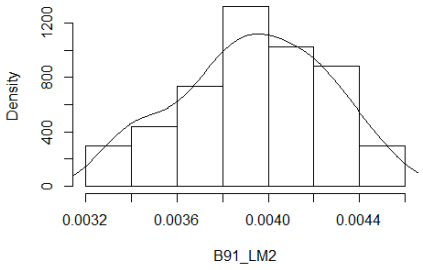
Residual Consumption of Gasoline 91 (y_{B91})



Skewness = - 0.1471

Durbin-Watson test = 2.0792

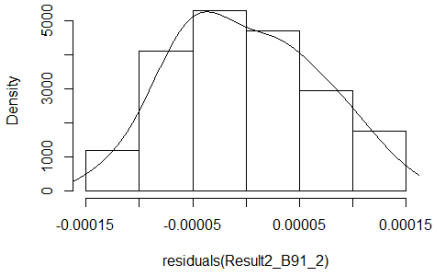
Consumption of Gasoline 91 $\left(\frac{1}{y_{B91}}\right)$



Skewness = - 0.189

GLM

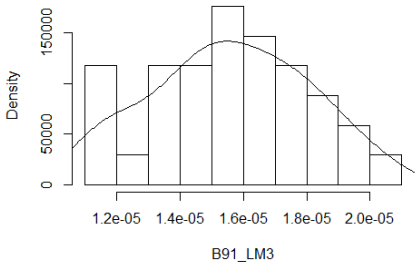
Residual Consumption of Gasoline 91 $\left(\frac{1}{y_{B91}}\right)$



Skewness = 0.1984

Durbin-Watson test = 2.2286

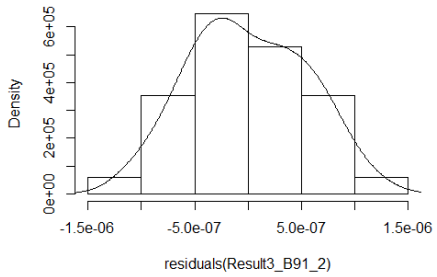
Consumption of Gasoline 91 $\left(\frac{1}{y_{B91}^2}\right)$



Skewness = - 0.0453

GLM

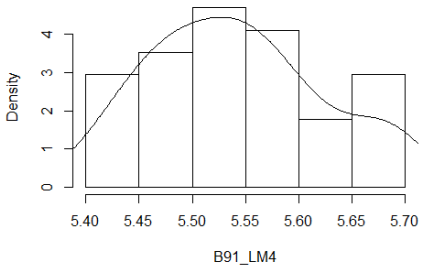
Residual Consumption of Gasoline 91 $\left(\frac{1}{y_{B91}^2}\right)$



Skewness = 0.0766

Durbin-Watson test = 2.2259

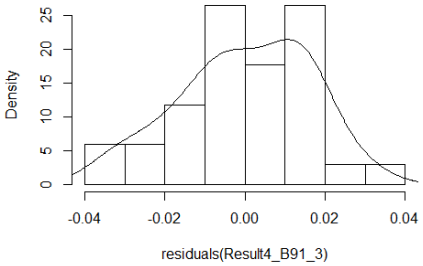
Consumption of Gasoline 91 $\left(\log\left(y_{B91}\right)\right)$



Skewness = 0.3318

GLM

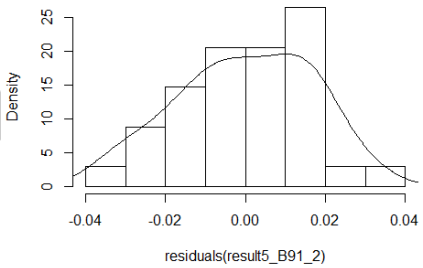
Residual Consumption of Gasoline 91 $\left(\log\left(y_{B91}\right)\right)$



Skewness = - 0.3395

Durbin-Watson test = 2.2407

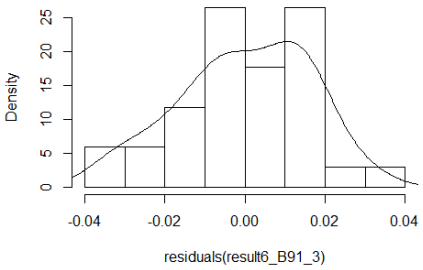
Residual Consumption of Gasoline 91 (y_{B91})
GLMs , family = Gamma(link = "inverse")



Skewness = - 0.2418

Durbin-Watson test = 2.2599

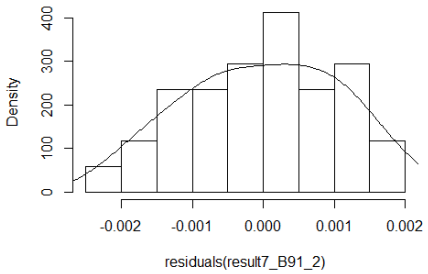
Residual Consumption of Gasoline 91 (y_{B91})
GLMs , family = Gamma(link = "log")



Skewness = - 0.3362

Durbin-Watson test = 2.2599

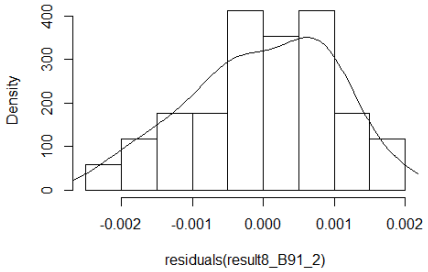
Residual Consumption of Gasoline 91 (y_{B91})
GLMs , family = inverse.gaussian(link = $\frac{1}{\mu^2}$)



Skewness = - 0.1078

Durbin-Watson test = 2.2599

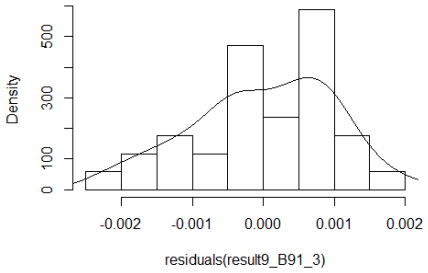
Residual Consumption of Gasoline 91 (y_{B91})
GLMs , family = inverse.gaussian(link = "log")



Skewness = - 0.301

Durbin-Watson test = 2.2599

Residual Consumption of Gasoline 91 (y_{B91})
GLMs , family = inverse.gaussian(link = "sqrt")



Skewness = - 0.333

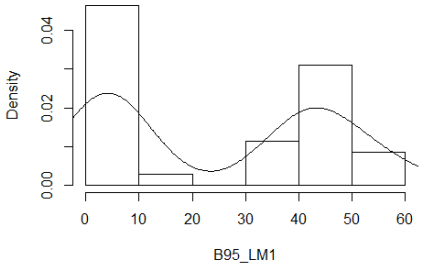
Durbin-Watson test = 2.2599

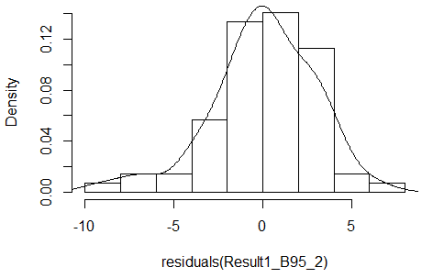
อิสโตแกรมและสัมประสิทธิ์ความเบ้แสดงการแจกแจงของ y และการแจกแจงของความคลาดเคลื่อนรวมทั้งสถิติทดสอบความเป็นอิสระของความคลาดเคลื่อนของปริมาณการใช้น้ำมันเบนซิน 95 โดยใช้วิธีการวิเคราะห์ตัวแบบเชิงเส้นทั่วไป (General Linear Model) และวิธีการวิเคราะห์ตัวแบบเชิงเส้นนัยทั่วไป (Generalized Linear Model)

Consumption of Gasoline 95 (y_{B95})

GLM

Residual Consumption of Gasoline 95 (y_{B95})





Skewness = 0.1136

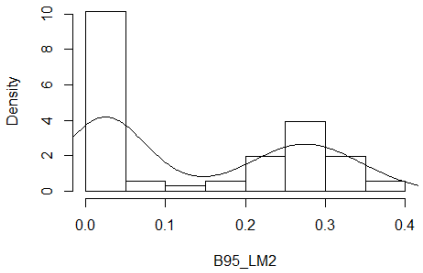
Skewness = - 0.6362

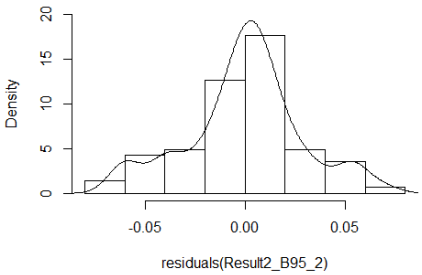
Durbin-Watson test = 1.4956

GLM

Consumption of Gasoline 95 $\left(\frac{1}{y_{B95}}\right)$

Residual Consumption of Gasoline 95 $\left(\frac{1}{y_{B95}}\right)$



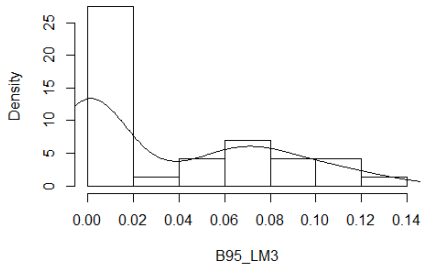


Skewness = 0.3223

Skewness = - 0.1092

Durbin-Watson test = 1.1966

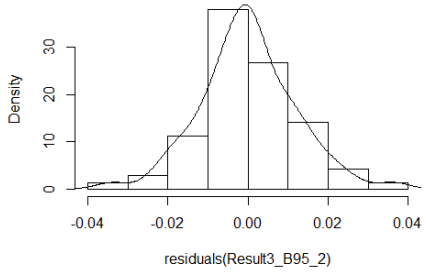
Consumption of Gasoline 95 $\left(\frac{1}{y_{B95}^2}\right)$



Skewness = 0.6738

GLM

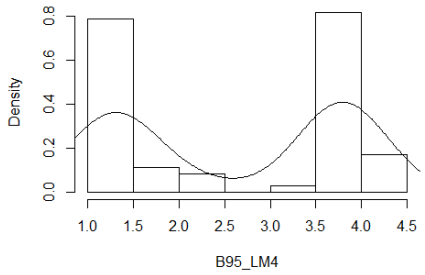
Residual Consumption of Gasoline 95 $\left(\frac{1}{y_{B95}^2}\right)$



Skewness = 0.1552

Durbin-Watson test = 1.481

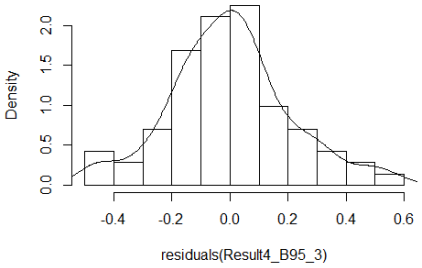
Consumption of Gasoline 95 $\left(\log\left(y_{B95}\right)\right)$



Skewness = - 0.0905

GLM

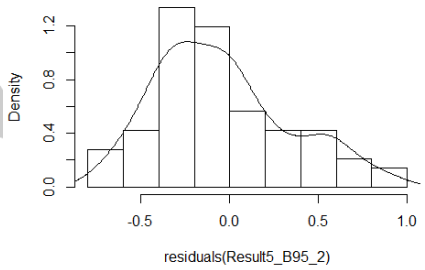
Residual Consumption of Gasoline 95 $\left(\log\left(y_{B95}\right)\right)$



Skewness = 0.2376

Durbin-Watson test = 1.0573

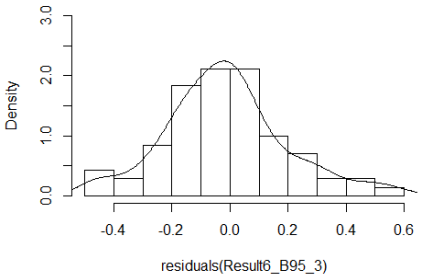
Residual Consumption of Gasoline 95 (y_{B95})
GLMs , family = Gamma(link = "inverse")



Skewness = 0.5858

Durbin-Watson test = 1.1009

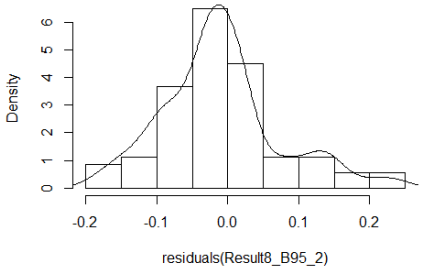
Residual Consumption of Gasoline 95 (y_{B95})
GLMs , family = Gamma(link = "log")



Skewness = 0.3964

Durbin-Watson test = 0.97422

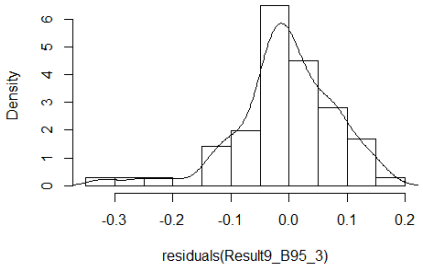
Residual Consumption of Gasoline 95 (y_{B95})
GLMs , family = inverse.gaussian(link = "log")



Skewness = 0.6309

Durbin-Watson test = 0.78788

Residual Consumption of Gasoline 95 (y_{B95})
GLMs , family = inverse.gaussian(link = "sqrt")

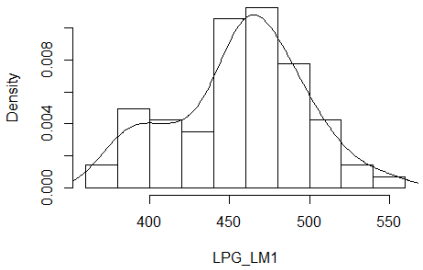


Skewness = - 0.8779

Durbin-Watson test = 0.9379

อิสโตแกรมและสัมประสิทธิ์ความเบ้แสดงการแจกแจงของ y และการแจกแจงของความคลาดเคลื่อนรวมทั้งสถิติทดสอบความเป็นอิสระของความคลาดเคลื่อนของปริมาณการใช้แก๊ส LPG โดยใช้วิธีการวิเคราะห์ตัวแบบเชิงเส้นทั่วไป (General Linear Model) และวิธีการวิเคราะห์ตัวแบบเชิงเส้นนัยทั่วไป (Generalized Linear Model)

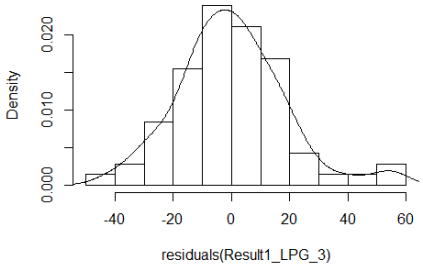
Consumption of LPG (y_{LPG})



Skewness = - 0.2153

GLM

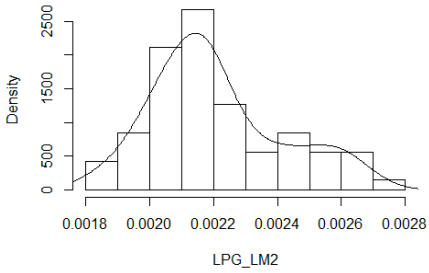
Residual Consumption of LPG (y_{LPG})



Skewness = 0.5238

Durbin-Watson test = 1.4705

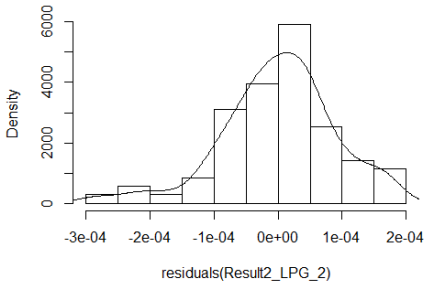
Consumption of LPG $\left(\frac{1}{y_{LPG}}\right)$



Skewness = 0.6182

GLM

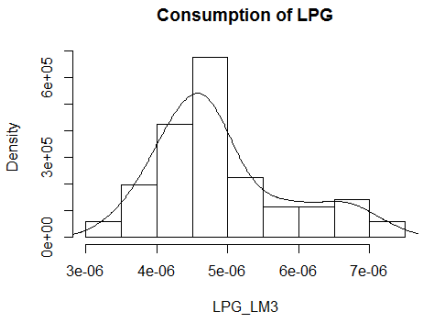
Residual Consumption of LPG $\left(\frac{1}{y_{LPG}}\right)$



Skewness = - 0.5639

Durbin-Watson test = 1.4617

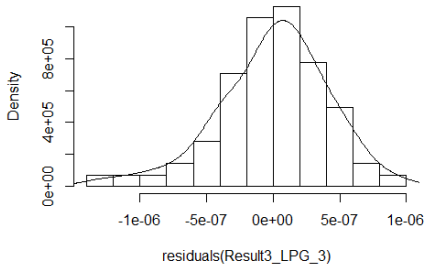
Consumption of LPG $\left(\frac{1}{y_{LPG}^2}\right)$



Skewness = 0.8014

GLM

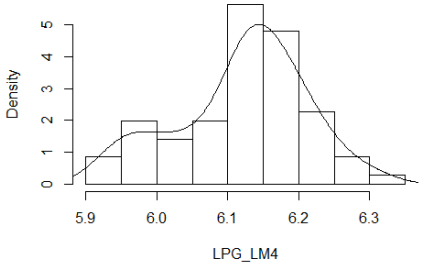
Residual Consumption of LPG $\left(\frac{1}{y_{LPG}^2}\right)$



Skewness = - 0.5717

Durbin-Watson test = 1.3923

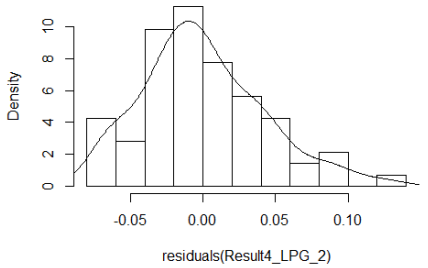
Consumption of LPG $\left(\log\left(y_{LPG}\right)\right)$



Skewness = - 0.4229

GLM

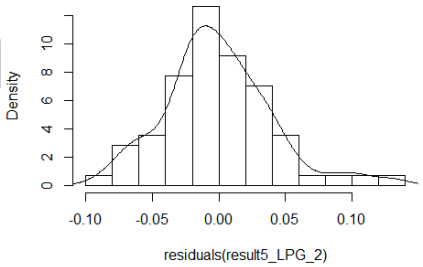
Residual Consumption of LPG $\left(\log\left(y_{LPG}\right)\right)$



Skewness = 0.5914

Durbin-Watson test = 1.3766

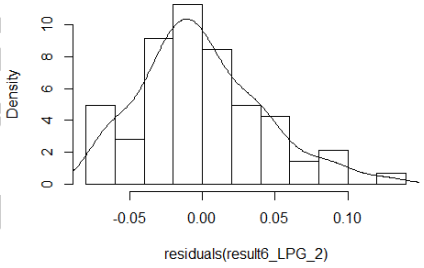
Residual Consumption of LPG (y_{LPG})
GLMs , family = Gamma(link = "inverse")



Skewness = 0.6064

Durbin-Watson test = 1.4705

Residual Consumption of LPG (y_{LPG})
GLMs , family = Gamma(link = "log")

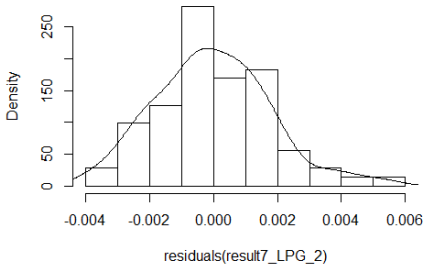


Skewness = 0.6173

Durbin-Watson test = 1.3683

Residual Consumption of LPG (y_{LPG})

GLMs , family = inverse.gaussian(link = $\frac{1}{\mu^2}$)

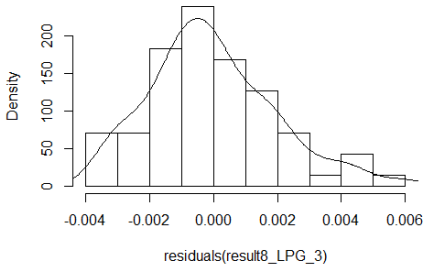


Skewness = 0.3648

Durbin-Watson test = 1.6069

Residual Consumption of LPG (y_{LPG})

GLMs , family = inverse.gaussian(link = "log")

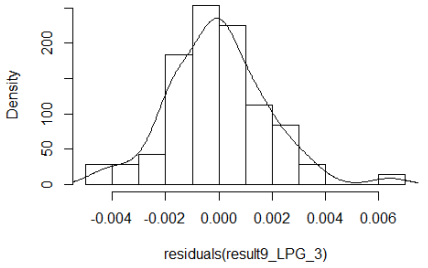


Skewness = 0.6187

Durbin-Watson test = 1.3683

Residual Consumption of LPG (y_{LPG})

GLMs , family = inverse.gaussian(link = "sqrt")

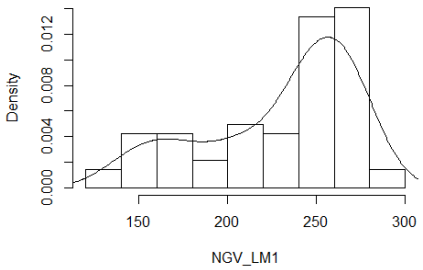


Skewness = 0.4117

Durbin-Watson test = 1.6744

อิสโตแกรมและสัมประสิทธิ์ความเบ้แสดงการแจกแจงของ y และการแจกแจงของความคลาดเคลื่อนรวมทั้งสถิติทดสอบความเป็นอิสระของความคลาดเคลื่อนของปริมาณการใช้แก๊ส NGV โดยใช้วิธีการวิเคราะห์ตัวแบบเชิงเส้นทั่วไป (General Linear Model) และวิธีการวิเคราะห์ตัวแบบเชิงเส้นนัยทั่วไป (Generalized Linear Model)

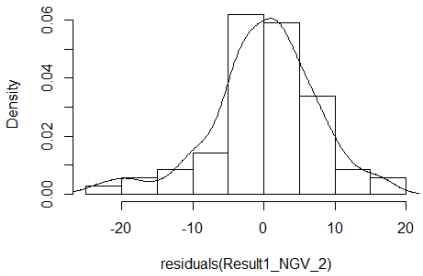
Consumption of Gasohol 91 (y_{G91})



Skewness = - 0.8092

GLM

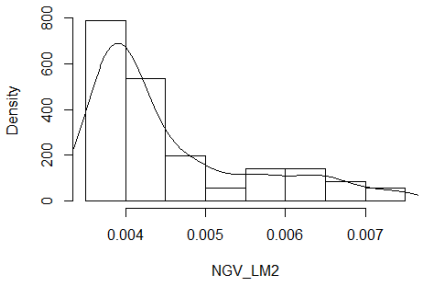
Residual Consumption of Gasohol 91 (y_{G91})



Skewness = - 0.5846

Durbin-Watson test = 2.0245

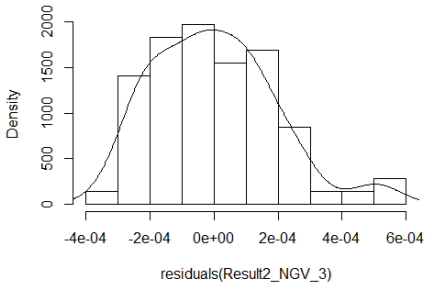
Consumption of Gasohol 91 $\left(\frac{1}{y_{G91}}\right)$



Skewness = 1.2956

GLM

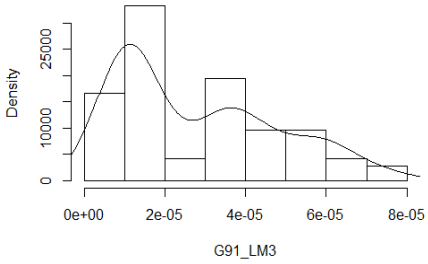
Residual Consumption of Gasohol 91 $\left(\frac{1}{y_{G91}}\right)$



Skewness = 0.6288

Durbin-Watson test = 1.7128

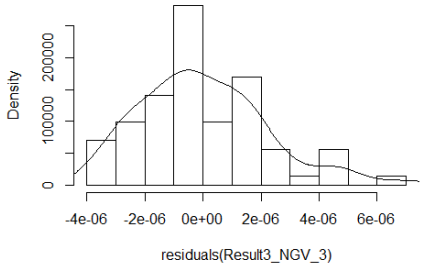
Consumption of Gasohol 91 $\left(\frac{1}{y_{G91}^2}\right)$



Skewness = 1.5686

GLM

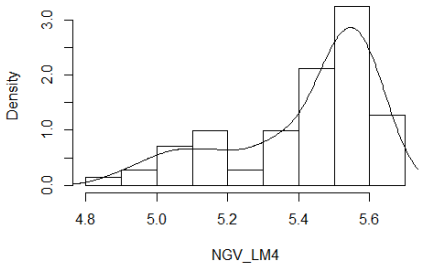
Residual Consumption of Gasohol 91 $\left(\frac{1}{y_{G91}^2}\right)$



Skewness = 0.7096

Durbin-Watson test = 1.8209

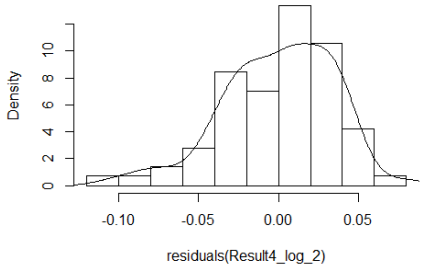
Consumption of Gasohol 91 $\left(\log\left(y_{G91}\right)\right)$



Skewness = - 1.0432

GLM

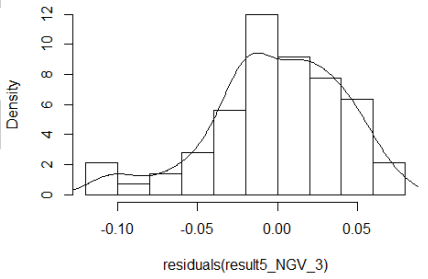
Residual Consumption of Gasohol 91 $\left(\log\left(y_{G91}\right)\right)$



Skewness = - 0.5649

Durbin-Watson test = 1.7965

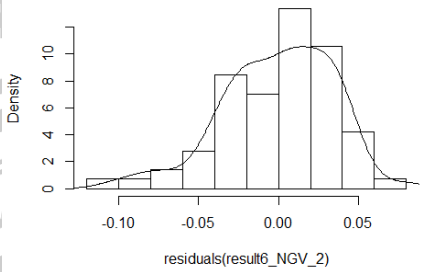
Residual Consumption of Gasohol 91 (y_{G91})
GLMs , family = Gamma(link = "inverse")



Skewness = - 0.6058

Durbin-Watson test = 1.8505

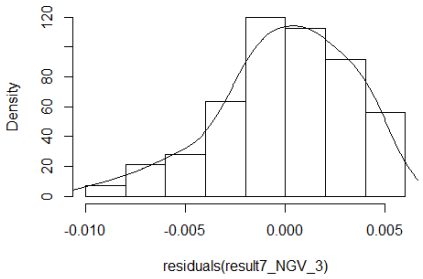
Residual Consumption of Gasohol 91 (y_{G91})
GLMs , family = Gamma(link = "log")



Skewness = - 0.5466

Durbin-Watson test = 2.037

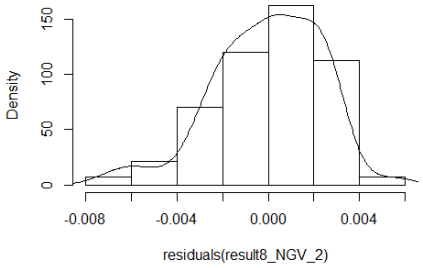
Residual Consumption of Gasohol 91 (y_{G91})
GLMs , family = inverse.gaussian(link = $\frac{1}{\mu^2}$)



Skewness = - 0.6587

Durbin-Watson test = 1.8505

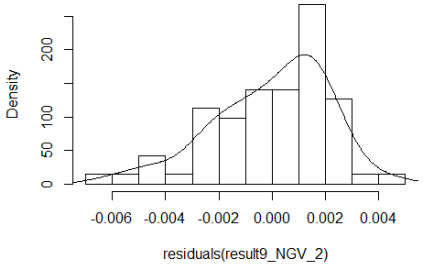
Residual Consumption of Gasohol 91 (y_{G91})
GLMs , family = inverse.gaussian(link = "log")



Skewness = - 0.5849

Durbin-Watson test = 2.037

Residual Consumption of Gasohol 91 (y_{G91})
GLMs , family = inverse.gaussian(link = "sqrt")



Skewness = - 0.6369

Durbin-Watson test = 2.0211

ประวัติผู้เขียน

ชื่อ	นางสาวอัมพิกา ก้อนแก้ว
วันเกิด	วันที่ 20 ตุลาคม พ.ศ. 2533
สถานที่เกิด	อำเภอเมือง จังหวัดเชียงราย
สถานที่อยู่ปัจจุบัน	บ้านเลขที่ 252 หมู่ 3 ตำบลจอมหมอกแก้ว อำเภอแม่ลาว จังหวัดเชียงราย รหัสไปรษณี 57250
ประวัติการศึกษา	พ.ศ. 2552 มัธยมศึกษาตอนปลาย โรงเรียนดาวรุ่งวิทยา จังหวัดภูเก็ต พ.ศ. 2556 ปริญญาวิทยาศาสตรบัณฑิต (วท.บ.) สาขาวิชา สถิติ มหาวิทยาลัยมหาสารคาม พ.ศ. 2562 ปริญญาวิทยาศาสตรมหาบัณฑิต (วท.ม.) สาขาวิชา วิทยาการจัดการสถิติ มหาวิทยาลัยมหาสารคาม

พหูน ปณ ฑิโต ชีเว