

ปีที่ 22 ฉบับที่ 3 [2566] Vol.22, No.3 [2023]

กันยายน – ธันวาคม 2566

ISSN 2822-1451 [Print]

ISSN 2822-1508 [Online]



Approved by TCI during
2020-2024



ศูนย์ดัชนีการอ้างอิงวารสารไทย
Thai Journal Citation Index Centre

Advisory Board

from King Mongkut's University of Technology North Bangkok, Thailand

Professor Dr. Saowanit Sukparungsee

Professor Dr. Vilai Rungsardthong

Associate Professor Dr. Surapun Yimman

Associate Professor Dr. Pianpool Kamoljitprapa

Associate Professor Dr. Kitti Bodhipadma

Editor-in-Chief

from King Mongkut's University of Technology North Bangkok, Thailand

Associate Professor Narumol Kreua-ongarjnukool

Editor

from King Mongkut's University of Technology North Bangkok, Thailand

Associate Professor Dr. Nonchanutt Chudpooti

Associate Professor Dr. Sekson Sirisubtawee

Co-Editor

from King Mongkut's University of Technology North Bangkok, Thailand

Professor Dr. Yupaporn Areepong

Associate Professor Dr. Anuchit Jitpattanakul

Assistant Professor Dr. Chatchawan Singhapol

Dr. Wannarak Nopcharoenkul

Editorial Team

from King Mongkut's University of Technology North Bangkok, Thailand

Associate Professor Dr. Orathai Polsen

Associate Professor Dr. Suvimol Phanyaem

Associate Professor Dr. Walaiporn Prissanaroon Ouajai

Associate Professor Dr. Tanapat Anusas-amornkul

Assistant Professor Dr. Anusara Srisrual

Assistant Professor Dr. Theerawut Phusantisampan

Assistant Professor Dr. Jintawat Tanamatayarat

Assistant Professor Dr. Sukanya Thepwatee

Assistant Professor Dr. Thidarat Wangkham

Assistant Professor Dr. Saowapa Thumsing Niyomthai

Editorial Board

Professor Dr. Paul Pigram

La Trobe University, Australia

Professor Dr. Melvin Pascall

The Ohio State University, USA

Prof. Dr. Andrea De Gaetano

Óbuda University, Budapest, Hungary and CNR-IASI Rome and CNR-IRIB Palermo, Italy

Assoc.Prof. Dr. Nutapong Somjit

University of Leeds, United Kingdom

Dr. Nutsuda Sumonsiri

Teesside University, United Kingdom

Professor Dr. Sergey Meleshko

Suranaree University of Technology, Thailand

Editorial Board (cont.)

Emeritus Professor Dr. Yongwimon Lenbury

Mahidol University, Thailand

Professor Dr. Sutthisak Phongthanapanich

King Mongkut's University of Technology North Bangkok, Thailand

Professor Dr. Anuvat Sirivat

Chulalongkorn University, Thailand

Prof. Dr. Supa Hannongbua

Kasetsart University, Thailand

Associate Professor Dr. Yaowadee Temtanapat

Thammasat University, Thailand

Associate Professor Dr. Khongsak Srikaeo

Pibulsongkram Rajabhat University, Thailand

Associate Professor Dr. Weeradej Meeinkuirt

Mahidol University, Thailand

Associate Professor Dr. Wararit Panichkitkosolkul

Thammasat University, Thailand

Associate Professor Dr. Montip Tiensuwan

Mahidol University, Thailand

Associate Professor Dr. Piyapatr Busababodhin

Maharakham University, Thailand

Associate Professor Dr. Wantida Chaiyana

Chiang Mai University, Thailand

Assoc.Prof. Dr. Pranee Phinyocheep

Mahidol University, Thailand

Assistant Professor Dr. Kwan Arayathanitkul

Mahidol University, Thailand

Assistant Professor Dr. Narumon Emarat

Mahidol University, Thailand

Assistant Professor Dr. Kitiporn Plaimas

Chulalongkorn University, Thailand

Assistant Professor Dr. Chatchai Khunboa

Khon Kaen University, Thailand

Assistant Professor Dr. Julaluk Khemacheewakul

Chiang Mai University, Thailand

Assistant Professor Dr. Jutarat Iewkittayakorn

Prince of Songkla University, Thailand

Assistant Professor Dr. Yanisa Laoong-u-thai

Burapha University, Thailand

Assistant Professor Dr. Sararat Mahasaronon

Naresuan University, Thailand

Dr. Worawut Srisukkhram

Chiang Mai University, Thailand

Management Staff

from King Mongkut's University of Technology North Bangkok, Thailand

Miss Wilawan Sae-jeng

Miss Kitsiya Chuchuaysuwan

JOURNAL POLICY

Journal of Applied Science and Emerging Technology (JASET) is an academic journal published Triannually by the Faculty of Applied Science, King Mongkut's University of Technology North Bangkok. The JASET publishes papers in four types: (1) original research articles in Thai or English, (2) academic articles in Thai or English, (3) review articles in Thai or English, and (4) editorial corner/invitation articles in English, encompassing all areas of applied science and technology. The journal will not accept articles, which have been published or are being considered for publication by another journal, nor should papers published here be submitted to other journals.

"Journal of Applied Science and Emerging Technology does not have policy to collect publication fee"

"Journal of Applied Science and Emerging Technology published both as hard-copies [ISSN 2822-1451 (Print)] and electronic journal [ISSN 2822-1508 (Online)] available on ThaiJO system"

Scope of the Journal

To publish academic, research and review articles covering all areas of both basic and applied science and technology including pure and applied mathematics, statistics, chemistry and applied chemistry, physics and industrial physics, environmental science and technology, biotechnology, agro-industrial and food technology, medical science and applications, health and beauty technology, computer and informatic science and materials science

Peer Review Process

Each article must be double blind peer reviewed by at least 3 reviewers from the related field.

Human and Animal Ethics in Research

The authors are requested for the appropriate disclosures and declarations if their research involves human participants and animals having an impact on their right, prestige, safety, and health. Hence, they must receive approval from the appropriate ethics committees for research involving humans and/or animals and submit the evidence of approval with the manuscript. Starting from May, 1st 2022)

Language

Both Thai and English

Publication Frequency

Once every four months (Triannual)

First Issue: January to April

Second Issue: May to August

Third Issue: September to December

Remarks

Authors have to be responsible for any legal effects that may occur due to their opinions expressed in the articles.

Website for submissions: www.tci-thaijo.org/index.php/JASCI

Contact: jaset@sci.kmutnb.ac.th

Copyright©2023 by Faculty of Applied Science, King Mongkut's University of Technology North Bangkok. All rights reserved.

No part of this publication may be reproduced stored or transmitted in any material form or by any methods including electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the publisher, except in accordance with the provisions of the Copyright Design and Patents Act 1988.

TABLE OF CONTENT

Editorial Corner (invitation article)

From Waste to Wealth: Challenges in Producing Value-Added Biochemicals from Lignocellulose Biorefinery

Malinee Sriariyanun* and Babu Dharmalingame900001

Research Articles

Forecasting of Thai International Imports and Exports Using Holt-Winters' and Autoregressive Integrated

Moving Average Models

Pianpool Kamoljitprapa*, Orathai Polsen and Umar Kabir Abdullahie252955

Confidence Intervals using Bootstrap Methods for the Parameter of the zero-truncated Poisson-Shanker Distribution
and their Application

Panicha Kaskasamkul and Wararit Panichkitkosolkule251640

Exact solutions of the fractional Landau-Ginzburg-Higgs equation and the (3+1)-dimensional space-time fractional
modified KdV-Zakharov-Kuznetsov equation using the simple equation method

Aungkanaporn Chankaew, Sirasrete Phoosree and Jiraporn Sanjun*e253149

Comparison of Methods on Asymptotic and Exact Confidence Intervals for a Binomial Proportion with Application

การเปรียบเทียบวิธีการประมาณค่าของช่วงความเชื่อมั่นเชิงเส้นกำกับและช่วงความเชื่อมั่นที่แท้จริงสำหรับสัดส่วนของประชากร
ที่มีการแจกแจงทวินามและการประยุกต์

Chanakan Sungboonchoo*, Piyaphon Paichite251337

Confidence Interval for the Population Mean of Rayleigh Distribution with Application to Shear wave Velocity
of Soils in Thailand

ช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยประชากรของการแจกแจงเรย์ลีย์: กรณีศึกษาความเร็วคลื่นเฉือนของดินในประเทศไทย

Thanakit Teawpongpan and Patarawan Sangnawakij*e251542

A Comparison of Forecasting Models of Stock Price using ARIMA and GVM Models

การเปรียบเทียบตัวแบบการพยากรณ์ราคาหุ้นโดยใช้ตัวแบบอาร์มาและจีวีเอ็ม

Boonkong Dhakonlayodhine253493

Development of Kha Nom Tong Muan by using Paka-umpoul Defatted Bice Bran Flour Substituted Wheat Flour (*Oryza sativa* L.)

การพัฒนาผลิตภัณฑ์ขนมทองม้วนโดยใช้แป้งรำข้าวปกาอ์ปัสสัถนํ้ามันทดแทนแป้งสาลี

Phattaraporn Kaewkool*, Apichargorn Dangkhapensirikul and Nutpaphat Chaladmake248526

From Waste to Wealth: Challenges in Producing Value-Added Biochemicals from Lignocellulose Biorefinery

Malinee Sriariyanun^{1*} and Babu Dharmalingam²

¹Biorefinery and Process Automation Engineering Center, Department of Chemical and Process Engineering,

The Sirindhorn International Thai-German Graduate School of Engineering,

King Mongkut's University of Technology North Bangkok, Bangkok, Thailand

²Department of Mechanical Engineering, College of Engineering Guindy, Anna University, Chennai, Tamil Nadu, India

KEYWORDS: Biochemical, Biofuel, Biomass, Biorefinery, Circular economy

*Corresponding Author: malinee.s@tggs.kmutnb.ac.th

Lignocellulose biorefinery is a sustainable and integrated approach to convert lignocellulosic biomass into a variety of valuable products, including biofuels, biochemicals, platform chemicals, and biomaterials (Ruensodsai & Sriariyanun, 2022). Lignocellulosic biomass is composed of three main components: cellulose, hemicellulose, and lignin, which are tightly interconnected in the plant cell wall. These components are abundantly found in various renewable resources such as grass, wood, municipal wastes, agricultural residues, and dedicated energy crops. This concept supports the circular economy concept and the sustainable development goal (SDG), especially SDG 12, Responsible Consumption and Production. The main benefit of lignocellulose biorefinery is not only to gain the economic benefit of converting waste to products, it also reduces the release of waste and pollution to the environment (Panakkal & Sriariyanun, 2023).

Due to the recalcitrant structure and chemical properties of lignocellulose biomass, the biorefining process to convert biomass to a final product has low efficiency. The main bottleneck of this process is the hydrolysis of lignocellulose to monomers or oligomers of cellulose hemicellulose and lignin derivatives. Therefore,

the main steps in lignocellulose biorefinery generally include pretreatment, hydrolysis, fermentation and product recovery (Figure 1) (Phakeenuya & Kitiborwornkul, 2024). Pretreatment methods are employed to break down the physical and chemical barriers, making cellulose, hemicellulose, and lignin more accessible for subsequent processing, e.g. hydrolysis. Various pretreatment methods have been developed from lab scale to industrial scale operation that are mainly categorized as chemical, physical, biological, and combined methods. During hydrolysis, enzymatic (such as cellulases and hemicellulases) or chemical hydrolysis is used to break down cellulose into glucose and hemicellulose into various sugars. The resulting sugars, as oligomers or monomers, serve as the building blocks for the production of biofuels, biochemical and other value-added products. Then microorganisms, such as bacteria or yeast, are employed to ferment the sugars obtained from hydrolysis into biofuels (such as ethanol) or high-value biochemicals. Finally, the produced biofuels or biochemicals are separated from the fermentation broth and purified by various separation techniques, such as distillation, chromatography, and membrane filtration.

Challenges in Lignocellulose biorefinery

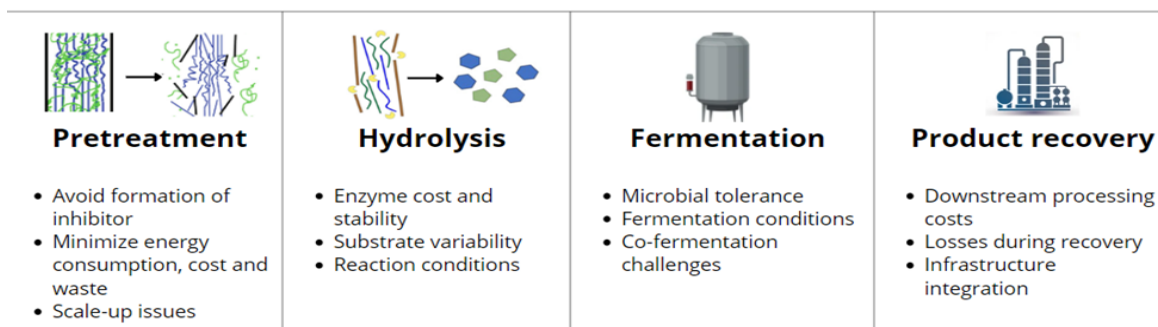


Figure 1 Challenges in lignocellulose biorefinery for the production of value-added biochemicals.

Based on this scenario, a biorefining process of lignocellulose is therefore multistep process, in which each step should support each other to promote the process efficiency. Therefore, the design of each step to optimize efficiency should not interfere with the downstream process. Ongoing research aims to improve efficiency, reduce costs, and expand the range of products obtained from lignocellulosic biomass and various challenges have been indicated. For example, some chemical pretreatment methods have a negative impact on enzymatic hydrolysis. The chemical residues of ionic liquid used in pretreatment impair cellulase activities. Furthermore, during harsh acid pretreatment, several byproducts, such as furfural, 5-HMF and acetic acid, are formed that suppress the activities of cellulases and microbes (Phusantisampan & Kitiborwornkul, 2022). In the case of enzymatic hydrolysis, it is challenging to formulate cellulase and hemicellulase to be compatible with various lignocellulose substrates. Currently, commercial enzymes for biorefinery are available, such as CelluClast 1.5L, CTec2, and Accellerase 1000 (Panakkal et al, 2023). To maximize the utilization of glucans and xylans obtained from cellulose and hemicellulose, co-fermentation is offered by using *Saccharomyces cerevisiae* and *Pichia kudriavzevii*, which are yeast strains that have strict preferences on glucans and xylans, respectively (Kirdponpattara et al, 2022). Furthermore, to make the product recovery step economically feasible, the process design should be done based on existing infrastructures, facilities and

technologies to accelerate the commercialization. Addressing these challenges requires ongoing research, innovation, and collaboration across various disciplines to make lignocellulose biorefinery more efficient, economically viable, and environmentally sustainable.

REFERENCES

- Kirdponpattara, S., Chueter, S., Sririyanun, M., & Phisalaphong, M. (2022). Bioethanol production by *pichia stipitis* immobilized on water hyacinth and Thin-Shell silk cocoon. *Applied Science and Engineering Progress*, 15(3), 4662–4662.
- Panakkal, E., & Sririyanun, M. (2023). Valorization of Lignocellulosic Biomass to Value Added Products. *The Journal of KMUTNB*, 33(1), 1–3.
- Phakeenuya, V., & Kitiborwornkul, N. (2024). Recent progress in biorefining process for production of biofuels, biochemicals and biomaterials from lignocellulosic biomass. *The Journal of KMUTNB*, 34(4), 1–8. <https://doi.org/10.14416/j.kmutnb.2023.03.002>
- Phusantisampan, T., & Kitiborwornkul, N. (2022). Progress in chemical pretreatment of lignocellulose biomass for applications in biorefinery. *The Journal of KMUTNB*, 32(4), 1087–1101.
- Ruensodsai, T., & Sririyanun, M. (2022). Sustainable development and progress of lignocellulose conversion to platform chemicals. *The Journal of KMUTNB*, 32(4), 815–818.
- Sririyanun, M., Gundupalli, M. P., Phakeenuya, V., Phusantisampan, T., Cheng, Y. S., & Venkatachalam, P. (2023). Biorefinery approaches for production of cellulosic ethanol fuel using recombinant engineered microorganisms. *Journal of Applied Science and Engineering*, 27(2), 1985–2005.

Forecasting of Thai International Imports and Exports Using Holt–Winters’ and Autoregressive Integrated Moving Average Models

Pianpool Kamoljitprapa^{1*}, Orathai Polsen¹, Umar Kabir Abdullahi²

¹Department of Applied Statistics, Faculty of Applied Science, King Mongkut’s University of Technology North Bangkok, Thailand

²Department of Statistics, Ahmadu Bello University, Zaria, Nigeria.

ABSTRACT

Imports and exports are two vital issues of the Thai economy, which were ranked as the 23rd largest economy in the world in 2020. Therefore, this research aimed to examine the Thai import and export volume trends (in million Baht) from January, 2010 to December, 2022. To conduct the analysis, the research employed Holt–Winters’ additive and multiplicative models, as well as various seasonal ARIMA models. The models were evaluated using different selection measure criteria, and the results indicated that the Holt–Winters’ multiplicative forecasting model was optimal for predicting both import and export volumes, as it produced the least mean absolute error (MAE) and root mean square error (RMSE). Thus, it is recommended for future Thai imports and exports forecasts. However, it is notable that the COVID–19 pandemic had a significant impact on the country’s imports and exports during the 2019–2020 period.

KEYWORDS: Imports, Exports, ARIMA model, Holt–Winters’ model, Seasonal, Forecasting

*Corresponding Author: pianpool.k@sci.kmutnb.ac.th

Received: 07/06/2023; Revised: 08/08/2023; Accepted: 16/08/2023

1. INTRODUCTION

Imports and exports are two crucial issues of the Thai economy. In 2020, Thailand’s GDP makes it 23rd largest economy in the world. Thailand was both a major exporter and the 25th largest importer. Its imports and exports constituted more than 130% of its GDP, according to the world bank, 2022 (World Bank, 2022). Thailand’s top 10 exports comprised vehicles (including parts and equipment), precious minerals and gems, and computers (including components), while Thai top 10 imports included crude oil, machinery and components, and electrical machinery and components, among others, as per Thailand trade policy and strategy (Trade Policy and Strategy office, 2022). To stimulate economic growth, the Thai government has actively encouraged exports through various policies and initiatives, including the “Thailand 4.0” policy, which seeks to transform the country’s economy from a manufacturing-based to a value-based one by emphasizing innovation, creativity, and technology. As

part of this strategy, the government has invested in infrastructure and logistics to enhance the efficiency of imports and exports, according to the OECD (OECD, 2017). Moreover, to promoting exports, the Thai government has also been striving to reduce the country’s reliance on imports, especially for energy, through the development of renewable energy sources such as wind, solar, and biomass, as well as the promotion of energy conservation and efficiency measures, as stated by the Ministry of Energy, 2021 (Ministry of Energy, 2021).

Forecasting is a widely used method for predicting uncertain future events using existing information. The Holt–Winters’ method developed by (Winters, 1960) is a seasonal time series model used to study the trend, pattern and forecasting future events from time series data. The model is an automatic forecasting framework based on exponential smoothing where more weight is assigned to recent values. It is either additive or multiplicative based on the pattern shown by the time

series plot. Many researchers have used Holt–Winters’ method to studied and forecast various events. For example, (Mladenović et al., 2016), analysed the export trends of the Republic of Serbia during the period of 2004 to 2014. Holt–Winters’ and ARIMA models were the two models used in the study. The forecast accuracy of the two models were compared for a period of 12 months. In (Nieto et al., 2021), studied the cargo volume trends of the ports of San Pedro Bay, California, United States of America (USA) using data from 2008 to 2016. Multiplicative Holt–Winters’ with two other time-series models and one machine-learning model were the models considered in the study. The models forecast cargo through the ports of results show that the multiplicative Holt–Winters’ model is the best method to forecast imports and exports of bulk cargo. In (Wongoutong, 2021) studied the effect on forecasting accuracy of multiplicative and additive Holt–Winters’ method using ten simulated datasets of different seasonality trends. Five actual datasets, in which it was difficult to distinguish the type of seasonal component, were also considered in the study. The results are in line with the significance of the correct identification of the type of seasonality before applying Holt–Winters’ method.

Autoregressive Integrated Moving Average (ARIMA) model, developed by (Box & Jenkins, 1976), is the most commonly used model in statistical analysis for forecasting. The models are especially useful for analyzing non-stationary time series data that can be explained by its past observations or lagged values. It utilizes autoregression (AR), moving average (MA), and seasonal differencing to identify trends and generate forecasts. The model building procedure involves finding the differences in the series by generating changes in the set $\{p, d, q, P, D, Q\}$ to ensure that stationarity is satisfied. Therefore, it is essential to test the overall stationery to determine the appropriate orders of the AR and MA components. This testing procedure requires a powerful algorithm for iterative estimation to reduce the computation burden. After obtaining the estimated parameters of a tentatively identified ARIMA model, diagnostic checking is

necessary to confirm that the model is adequate. To select the optimal model, various statistical criteria such as *MAE*, *MAPE*, *RMSE*, and *MSE* are used. These measures provide information about the accuracy of the forecasts generated by different models. The goal is to choose the set of orders $\{p, d, q, P, D, Q\}$ that produces the smallest values of *MAE*, *MAPE*, *RMSE*, and *MSE*. This approach is based on the statistical principle of minimizing the error between the actual and forecasted values. It involves identifying, estimating, and diagnosing predictions using lagged moving averages to smooth time series data. (Benvenuto et al., 2021).

The ARIMA can be modified to suit different models based on different time series. Many researchers have used ARIMA to forecast various events. For example, (Benvenuto et al., 2021), studied the dynamics of COVID-19 in India using ARIMA to determine both the prevalence and incidence of the disease. In (Fattah et al., 2018) adopted the ARIMA approach to forecast the demand for finished products in food manufacturing. In (Kamoljitprapa & Sookkhee, 2022), employed an alternative technique, the Box–Jenkins time series procedure, to forecast the total amount of CO₂ emissions in Thailand from 2001 to 2020. In (Kaur & Rakshit, 2019), used suitable ARIMA methods for forecasting the analysis of rainfall. In a context of the exports and imports, ARIMA models are also play a vital role in these area. For example, (Panday & Dhakal, 2020), studied the use of ARIMA models to forecast exports and imports in Nepal’s agricultural sector and found that ARIMA models provided more accurate forecasts compared to other techniques. In (Oghenekevwe & Mercy, 2021), used ARIMA models to forecast Nigeria’s export and import performance in the agricultural sector and found that ARIMA models could accurately predict export and import trends. Therefore, forecasting techniques such as the ARIMA and Holt–Winters’ methods can significantly contribute to a country’s economy by providing a roadmap for sustainable trade growth and development. Policymakers and stakeholders can make informed decisions about their imports and exports, which can significantly contribute to a country’s economy.

2. MATERIALS AND METHODS

The time series data used in this study was collected by the Information Communication Technology Center, Office of the Permanent Secretary, Ministry of Commerce, spanning from January 2010 to December 2022 (a total of 132 months) (Office of the Permanent Secretary, 2023). The models used 120 months for training data and the last 12 months were used for the test set. The R software was utilized to estimate the model parameters and perform diagnostic tests to validate the models.

2.1 Holt-Winters' Method

The Holt-Winters' method (Winters, 1960), is a seasonal time series model used to study the trend, pattern and forecasting future events from time series data where recent values are assigned more weight (Chatfield & Yar, 2014). The model is either additive or multiplicative based on the pattern shown by the time series plot.

The additive model is generally stated as;

$$J_t = \delta(y_t - S_{t-p}) + (1 - \delta)(J_{t-1} + H_{t-1}),$$

$$H_t = \beta(J_t - J_{t-1}) + (1 - \beta)H_{t-1},$$

$$S_t = \lambda(y_t - J_t) + (1 - \lambda)S_{t-p}.$$

The multiplicative model is generally stated as;

$$j_t = \delta \frac{y_t}{S_{t-p}} + (1 - \delta)(j_{t-1} + h_{t-1}),$$

$$h_t = \beta(j_t - j_{t-1}) + (1 - \beta)h_{t-1},$$

$$s_t = \lambda \frac{y_t}{j_t} + (1 - \lambda)s_{t-p},$$

where $0 \leq \delta \leq 1$, $0 \leq \beta \leq 1$, $0 \leq \lambda \leq 1$: δ , β and λ are the smoothing parameters, J_t is the smoothed level at time t , H_t is the change in the trend at time t , S_t is the seasonal smooth at time t , p is the number of seasons per year. The Holt-Winters' algorithm requires starting value for each of the component values. Most commonly are:

$$j_p = \frac{1}{p}(y_1 + y_2 + \dots + y_p),$$

$$h_p = \frac{1}{p} \left[\frac{y_{p+1} - y_1}{p} + \frac{y_{p+2} - y_2}{p} + \dots + \frac{y_{p+p} - y_p}{p} \right].$$

The formulas for linear and trend component are the same for both additive and multiplicative models as shown:

$$s_1 = Y_1 - j_p, \quad s_2 = Y_2 - j_p, \quad s_p = Y_p - j_p$$

(seasonal component for additive model),

$$s_1 = \frac{Y_1}{j_p}, \quad s_2 = \frac{Y_2}{j_p}, \dots, \quad s_p = \frac{Y_p}{j_p}$$

(seasonal component for multiplicative model).

So, the forecast model for time period is shown in Eq. (1),

$$F_{t+p} = J_t + pH_t + S_{t+p-1}. \quad (1)$$

2.2 ARIMA (Box and Jenkins) Model

The model combines the autoregressive (AR) and moving average (MA) concepts and can be denoted by ARIMA (p, d, q), where p represents the order of the AR component, d is the order of differencing, and q represents the order of the MA component. Eq. (2) presents the general form of the ARIMA model:

$$\phi_p(B)(1-B)^d x_t = \theta_q(B)\varepsilon_t, \quad (2)$$

where ϕ_p and θ_q are the AR and MA characteristic operators, and are represented by:

$$\phi_p(B) = 1 - \sum_{i=1}^p \phi_i B^i = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p,$$

$$\theta_q(B) = 1 - \sum_{i=1}^q \theta_i B^i = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q,$$

respectively. Here, B is a backward shift operator that shifts the time series data back by one time period, $B(y_t) = y_{t-1}$, and ε_t is the forecasting error term.

For seasonal time series data, a seasonal ARIMA model, denoted by ARIMA (p, d, q) (P, D, Q) is commonly used. Eq. (3) shows the seasonal ARIMA model, which includes the seasonal AR and MA components, as well as the seasonal differencing parameter, D :

$$\phi_p(B)\omega_p(B^s)(1-B^s)^D y_t = \theta_q(B)\gamma_q(B^s)\varepsilon_t \quad (3)$$

where $\omega_p(B^s)$ and $\gamma_q(B^s)$ are the seasonal AR and MA characteristic operators respectively,

$$\omega_p(B^l) = 1 - \omega_1 B^s - \omega_2 B^{2s} - \dots - \omega_p B^{ps},$$

$$\gamma_q(B^l) = 1 - \gamma_1 B^s - \gamma_2 B^{2s} - \dots - \gamma_q B^{qs},$$

and l is the sub index that stands for the seasonal period. The orders of the seasonal AR, MA, and differencing components are denoted by P , D and Q , respectively.

In summary, the ARIMA model provides a useful framework for modeling non-stationary time series data, and the inclusion of seasonal components in the model can further improve its accuracy and predictive power.

2.3 Model Evaluation

In the context of forecasting, the occurrence of errors is inevitable. To minimize these errors and ensure a high level of accuracy, several measures, such as the mean absolute error (MAE) and root mean squared error (RMSE) are commonly used and are used in this study.

The MAE evaluates the average magnitude of the errors in a set of predictions, without considering their direction. It is calculated as the average of the absolute differences between the predicted values ($\hat{y}_i$) and actual values (y_i) across the number of observed values (n). The weights assigned to each individual difference are equal. The MAE can be calculated as shown in Eq. (4),

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4)$$

RMSE, on the other hand, is a quadratic scoring rule that also assesses the average magnitude of the error. It is computed as the square root of the average of the squared differences between the predicted and actual values. The RMSE can be calculated as shown in Eq. (5),

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5)$$

3. RESULTS AND DISCUSSION

As depicted in Figure 1 the trend of Thailand's imports and exports over the past 12 years has been characterized by an overall upward trajectory punctuated by fluctuations. Specifically, Thailand's imports exhibited a downward trend from 2012 to

2016, followed by an upward trend from 2016 to 2019. Conversely, Thailand's exports have shown a consistent increase over time, albeit with occasional downturns. It is noteworthy that both imports and exports experienced a sharp increase in value around mid-year 2020. Again, the pandemic had a noticeable effect, with a decline in imports during the 2019–2020 period. Therefore, the graphs of both imports and exports series exhibited the non-stationarity at a level.

Moreover, Figure 2 and Figure 3 show that the magnitude of the seasonal component does not change with time; thus, a multiplicative decomposition is more appropriate. Hence, multiplicative Holt–Winters' model handles the data efficiently. Furthermore, Table 1 shows that the MAE and RMSE values obtained using this model satisfy the forecasting error criteria outlined in Section 2.

Table 1 The computed forecasting errors for the imports and exports using additive and multiplicative Holt–Winters' models

Model	MAE ($\times 10^4$)	RMSE ($\times 10^4$)
Imports:		
Additive seasonal	3.55	4.71
Multiplicative seasonal	2.65	3.38
Exports:		
Additive seasonal	3.54	4.21
Multiplicative seasonal	2.57	3.33

The comparison of the fitted values using the additive and multiplicative Holt–Winters' models with the actual values are shown in Figure 4 and Figure 5 for imports and exports, respectively. Both graphs satisfy the assumption of constant variance and independence normality, yet it is clear that Holt–Winters' multiplicative model (red line) follows the actual trend more closely than Holt–Winters' additive model (blue line). The analysis also involves looking at the Autocorrelation Function (ACF).

Table 2 The computed of forecasting errors for the imports and exports using ARIMA model

Model	MAE ($\times 10^4$)	RMSE ($\times 10^4$)
Imports: ARIMA(3,1,0)(0,0,1) ₁₂	3.58	4.86
Exports: ARIMA(0,1,1)(0,0,2) ₁₂	3.30	4.23

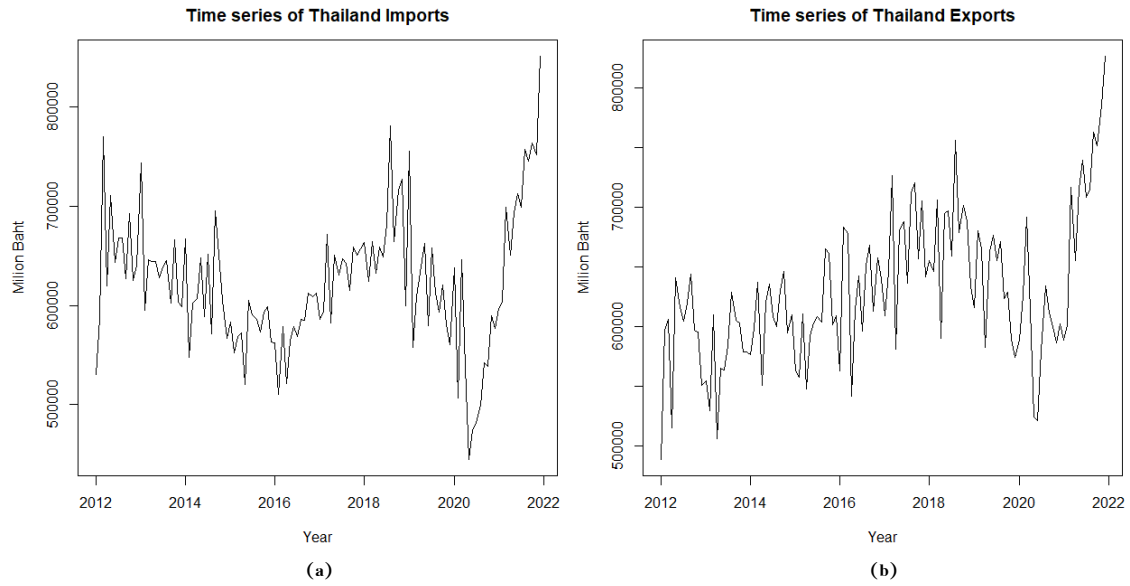


Figure 1 Trend of Thailand's (a) imports and (b) exports over the years 2012 to 2022

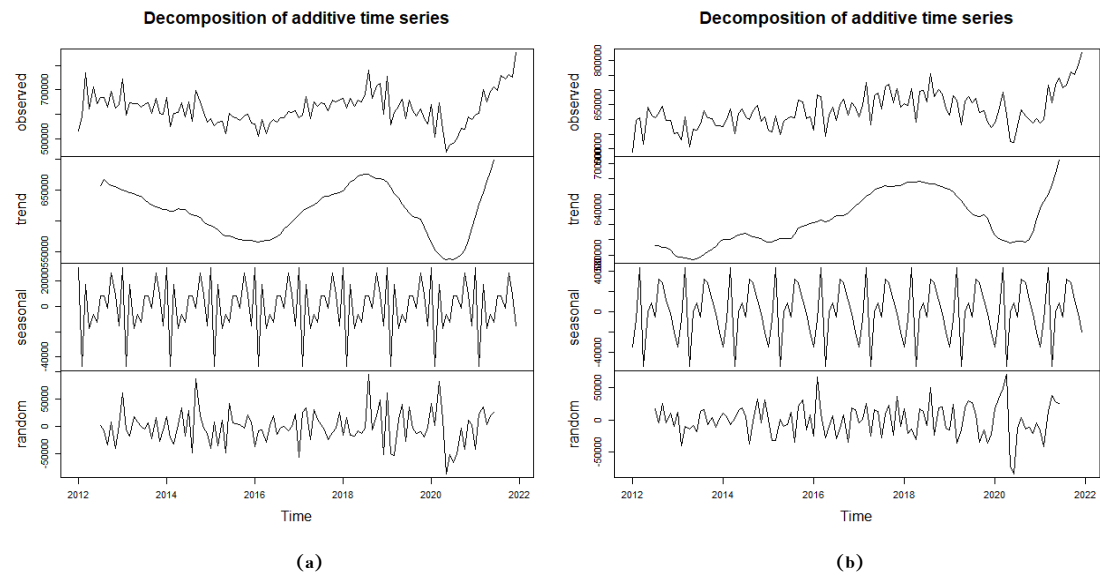


Figure 2 Additive decomposition plots for Thailand's (a) imports and (b) exports

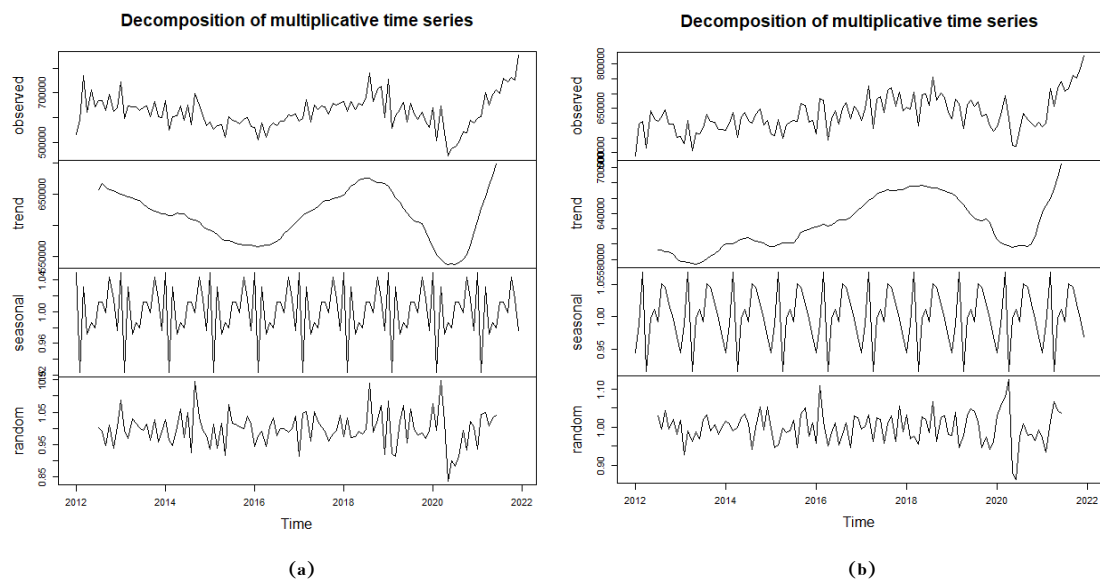


Figure 3 Multiplicative decomposition plots for Thailand's (a) imports and (b) exports

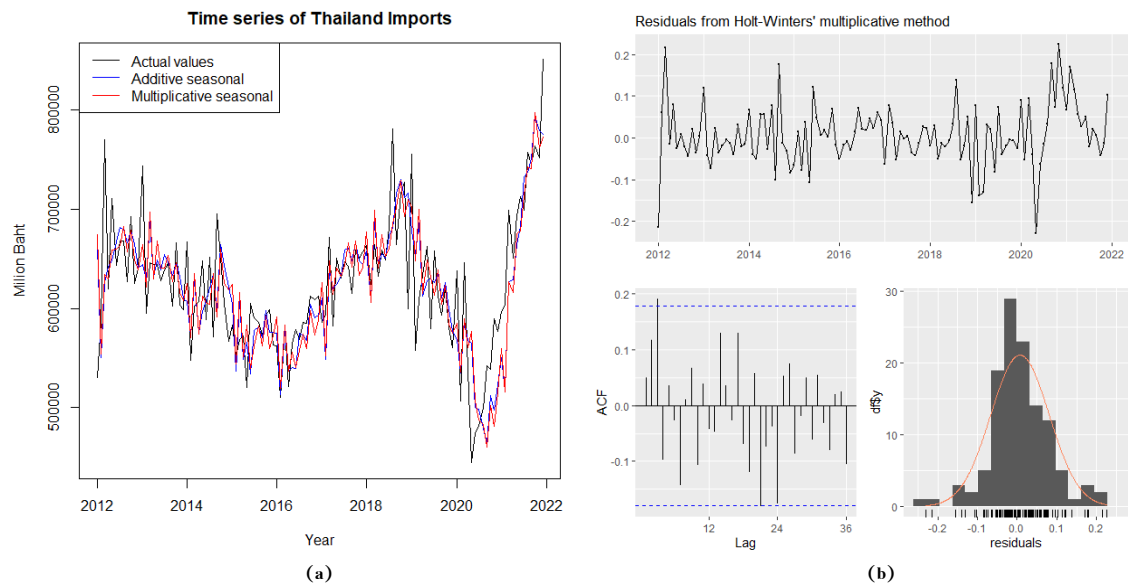


Figure 4 Comparison of Holt-Winters' additive and multiplicative models for imports (a) and the evaluation of the multiplicative model accuracy (b)

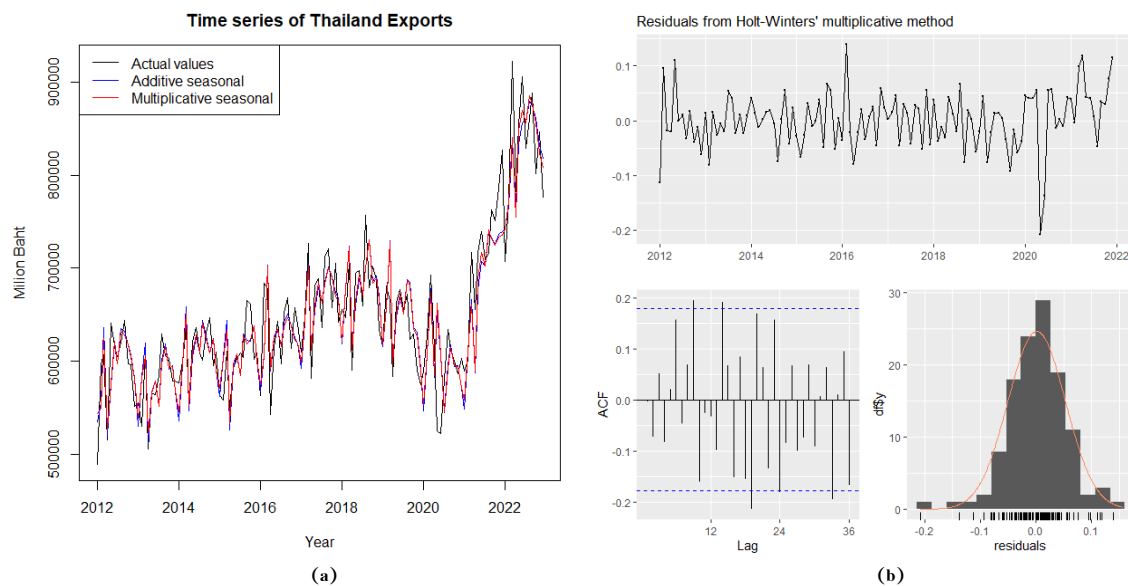


Figure 5 Comparison of Holt-Winters' additive and multiplicative models for exports (a) and the evaluation of the multiplicative model accuracy (b)

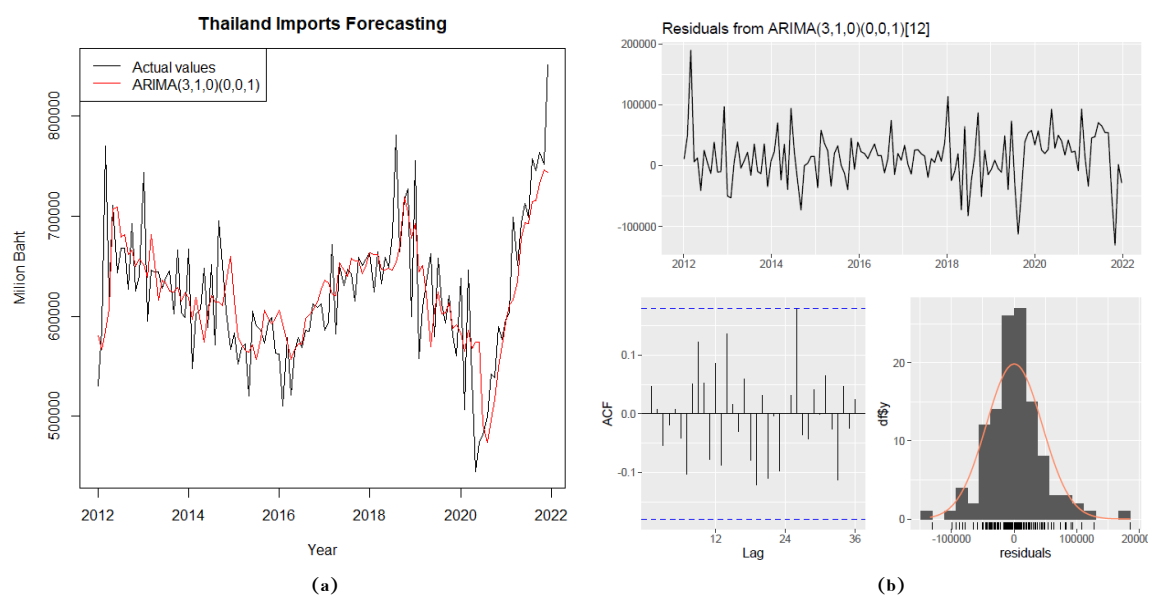


Figure 6 Actual and predicted values of Thailand's imports using ARIMA(3, 1, 0)(0, 0, 1)₁₂ model (a) and the evaluation of the model accuracy (b)

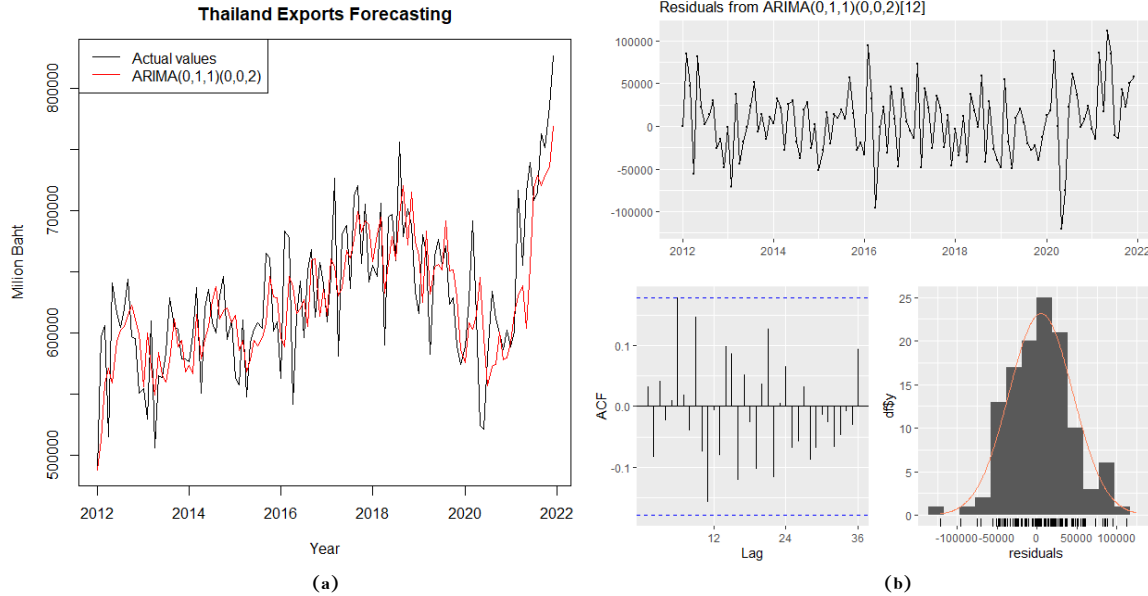


Figure 7 Actual and predicted values of Thailand's exports using ARIMA(0, 1, 1)(0, 0, 2)₁₂ model (a) and the evaluation of the model accuracy (b)

We also applied the ARIMA method by first stationarized the time series data and tested various combinations of the parametric orders $\{p, d, q, P, D, Q\}$ to determine the best-fitted models. Our findings indicate that the recommended models are ARIMA(3,1,0)(0,0,1)₁₂ and ARIMA(0,1,1)(0,0,2)₁₂ for imports and exports, respectively. The predicted values align closely with the actual values, as shown in Figure 6 and Figure 7, and the proposed models exhibit small residuals that satisfy the assumptions of independence, normality, and homoscedasticity. However, upon comparing the MAE and RMSE values presented in Table 2 for the ARIMA models with those obtained from the Holt-Winters' method, we observed that the multiplicative model produced lower error values. Thus, based on our results, the multiplicative Holt-Winters' model is the optimal choice for forecasting Thailand's imports and exports in this study.

Regarding the discussion, the optimal model selected, the multiplicative Holt-Winters' model, was used to predict monthly imports and exports values for the year 2022 and compare them with the actual values, which are presented in Table 3. Again, the overall predicted values by the models are consistent with the actual values. The models confirm their accurate prediction and successfully captures the changing trend

of imports and exports. Additionally, Table 4 and Figure 8 display the forecasting values (blue line) for 12 periods of the year 2023. The forecasts show rapid fluctuations, with both decreasing and increasing trends. While the point forecasts indicate a downward trend, the prediction intervals allow for potential downward movement in the data during the forecast period for both imports and exports. Also shown in Figure 8, the shade represents the 80% confidence interval (dark grey) and the 95% confidence interval (light grey).

4. CONCLUSION

In conclusion, based on the analysis and visual plots, multiplicative Holt-Winters' forecasting model exhibits the least MAE and RMSE for both Thailand's imports (2.65×10^4 and 3.38×10^4) and exports (2.57×10^4 and 3.33×10^4) datasets from January, 2010 to December, 2022, compared to additive Holt-Winters' model and seasonal ARIMA model. This model considers not only the dependence on the time series but also the interference of seasonal component, resulting in a relatively high prediction accuracy of long-term trends, which is often used in the field of economic forecasting (Hyndman & Athanasopoulos, 2021). Therefore, the model is recommended for future Thai import and export forecasts.

Table 3 Comparison of actual and predicted import and export values (in Million Bath) using optimal model, multiplicative Holt–Winters’ model, for the year 2022

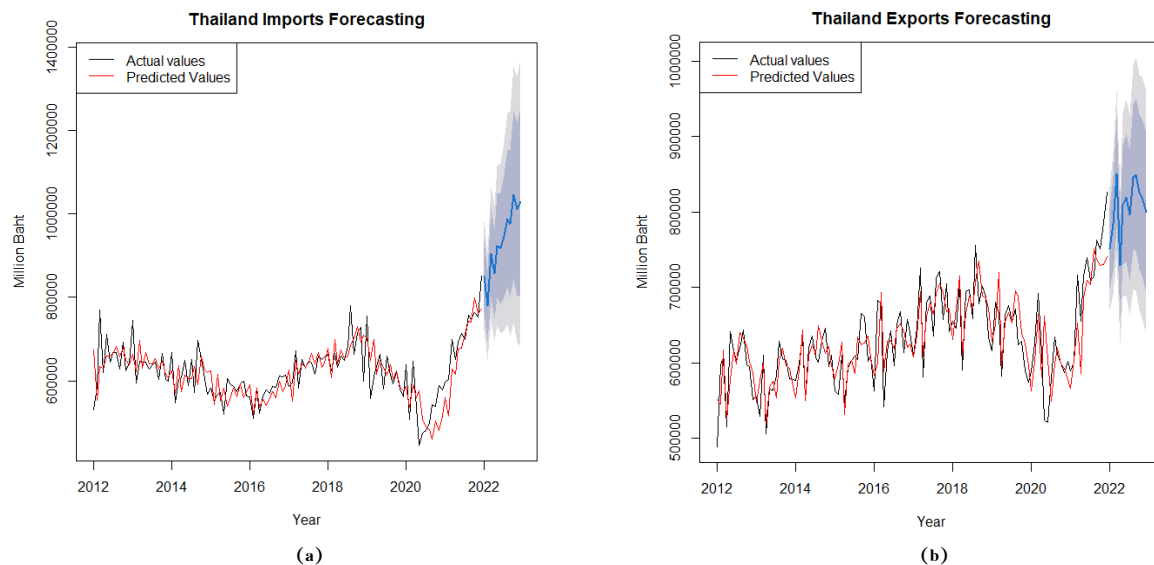
Month	Imports			Exports		
	Actual value	Predictive value	Residuals	Actual value	Predictive value	Residuals
January	796,295.6	851,788.8	-0.0651	706,715.7	741,745.4	-0.0472
February	772,456.5	763,243.3	0.0121	771,032.5	770,822.2	0.0003
March	869,554.1	890,302.4	-0.0233	922,277.2	840,113.8	0.0978
April	842,640.8	837,099.4	0.0066	781,925.0	754,351.1	0.0366
May	919,987.4	894,090.4	0.0290	854,264.7	848,356.8	0.0070
June	957,595.3	907,489.6	0.0552	905,744.9	869,202.1	0.0420
July	968,939.5	963,008.0	0.0062	829,028.8	855,766.2	-0.0312
August	1,026,653.5	1,012,311.9	0.0142	861,169.2	885,671.2	-0.0277
September	929,731.9	1,003,353.6	-0.0734	888,371.1	876,639.3	0.0134
October	832,874.8	989,734.3	-0.1585	801,273.4	856,394.9	-0.0644
November	907,142.7	926,844.8	-0.0213	846,190.5	823,913.6	0.0270
December	823,081.5	892,417.6	-0.0777	776,323.6	808,601.7	-0.0400

Table 4 Forecasting import and export values (in Million Bath) using optimal model, multiplicative Holt–Winters’ model, for the year 2023

Month	Imports		Exports	
	Forecasting	95% CI.	Forecasting	95% CI.
January	904,933.5	766,776.5, 1,043,090.5	753,002.3	672,088.5, 833,916.1
February	821,210.0	680,902.6, 961,517.3	801,268.8	705,144.8, 897,392.8
March	936,871.4	757,884.9, 1,115,857.9	873,162.0	758,532.1, 987,791.9
April	876,538.5	689,815.0, 1,063,262.0	747,752.5	641,802.7, 853,702.3
May	919,255.8	701,763.6, 1,136,748.1	825,971.6	700,935.3, 951,007.9
June	906,163.3	669,103.0, 1,143,223.6	843,371.4	70,8026.0, 978,716.9
July	921,168.7	655,931.2, 1,186,406.1	813,393.4	675,858.4, 950,928.5
August	947,892.7	648,857.6, 1,246,927.8	855,085.5	703,502.5, 1,006,668.5
September	916,201.6	600,902.9, 1,231,500.4	858,153.2	699,317.0, 1,016,989.4
October	920,504.1	576,369.0, 1,264,639.2	832,814.5	672424.7, 993,204.3
November	921,810.5	548,874.8, 1,294,746.1	827,651.2	662286.9, 993,015.4
December	891,906.9	502,832.3, 1,280,981.6	801,518.3	635799.8, 967,236.9

Notably, the COVID–19 pandemic had a significant impact on the country’s imports and exports, resulting in a modest decline in the years 2019 to 2020 and a subsequent sharp increase in 2021. While Thailand’s

trade patterns have been generally favorable over the past decade, the COVID–19 pandemic has posed significant challenges to the country’s import and export industries, as evidenced by the fluctuations in the data.

**Figure 8** Forecasting values for Thailand’s (a) imports and (b) exports using optimal model, multiplicative Holts–Winters’ model, for the year 2023

ACKNOWLEDGMENT

Funding was provided by the Faculty of Applied Science, King Mongkut's University of Technology North Bangkok for the fiscal year 2023, Contract no: 652112. This support is gratefully acknowledged.

REFERENCES

- Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., & Ciccozzi, M. (2020). Application of the ARIMA model on the COVID-2019 epidemic dataset. *Data in Brief*, 29, 105340.
- Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: forecasting and control*. San Francisco, CA: Holden-Day.
- Chatfield, C., & Yar, M. (2014). Holt-Winters forecasting: some practical issues. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 34(2-3), 215-225.
- Fattah, J., Ezzine, L., Aman, Z., El Moussami, H., & Lachhab, A. (2018). Forecasting of demand using ARIMA model. *International Journal of Engineering Business Management*, 10, 1-9. <https://doi.org/10.1177/1847979018808673>
- Hyndman, R. J., & Athanasopoulos G. (2021). *Forecasting: Principles and Practice* (3rd ed.). <https://otexts.com/fpp3/decomposition.html>
- Kamoljitprapa, P., & Sookkhee, S. (2022). Forecasting Models for Carbon Dioxide Emissions in Major Economic Sectors of Thailand. *Journal of Physics: Conference Series* 2346, 012001.
- Kaur, P., & Rakshit, M. (2019). Seasonal and Periodic Autoregressive Time Series Models Used for Forecasting Analysis of Rainfall Data. *International Journal of Advanced Research in Engineering and Technology*, 10(1), 230-242.
- Ministry of Energy. (2021). *Thailand's energy situation*. <https://www.energy.go.th/en/energy-situation/energy-situation-in-thailand/>
- Mladenović, J., Lepojević, V., & Janković-Milić V. (2016). Modelling and Prognosis of the Export of The Republic of Serbia by using Seasonal Holt-Winters and ARIMA Method. *Economic Themes*, 54(2), 233-260.
- Nieto, M. R., Carmona-Benitez, R. B., & Martinez, J. N. (2021). Comparing models to forecast cargo volume at port terminals. *Journal of Applied Research and Technology*, 19(3), 238-249.
- OECD. (2017). *Economic survey of Thailand 2017*. <https://www.oecd.org/economy/surveys/economic-survey-thailand.htm>
- Office of the Permanent Secretary. (2023). *Ministry of Commerce*. <https://tradereport.moc.go.th/Report/ReportEng.aspx?Report=TradeEnBalanceYearly>
- Oghenekevwe, O. A., & Mercy, O. O. (2021). Forecasting Export and Import Performance in the Agricultural Sector of Nigeria Using ARIMA Models. *International Journal of Agricultural and Biological Engineering*, 14(4), 145-152.
- Panday, P. K., & Dhakal, R. (2020). Forecasting Exports and Imports of Nepal's Agricultural Sector: A Comparison of ARIMA, ANN and ETS Models. *International Journal of Applied Sciences and Biotechnology*, 8(1), 15-20.
- Trade Policy and Strategy office. (2022). *Ministry of Commerce*. <https://www.tpsoc.moc.go.th/th/node/12413>
- Winters, P. R. (1960). Forecasting Sales by Exponentially Weighted Moving Averages. *Management Science*, 6(3), 324-342.
- Wongoutong, C. (2021). The Effect of Forecasting Accuracy of the Holt-Winter Method When Using the Incorrect Model on a Non-Stationary Time Series. *Thailand Statistician*, 19(3), 565-582.
- World Bank. (2022). *World development indicators*. <https://databank.worldbank.org/source/world-development-indicators>

Confidence Intervals using Bootstrap Methods for the Parameter of the zero-truncated Poisson-Shanker Distribution and their Application

Panicha Kaskasamkul¹ and Wararit Panichkitkosolkul^{2*}

¹Department of Mathematics, Faculty of Science, Naresuan University, Thailand

²Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, Thailand

ABSTRACT

Count data with non-zero values can be modeled using the zero-truncated Poisson-Shanker distribution in a variety of situations. However, the confidence interval estimation for the parameter has not yet been examined. In this paper, confidence interval estimation based on percentile bootstrap, simple bootstrap, biased-corrected and accelerated bootstrap, and bootstrap-t methods was examined in terms of coverage probability, lower and upper error rate, and average interval length through Monte Carlo simulation. The simulation results indicate that attaining the nominal confidence level using the bootstrap methods were not possible for small sample sizes, regardless of the other settings. Furthermore, in cases when the sample size was large, there was no significant difference observed in the performances of the all methods. In several scenarios, the bias-corrected and accelerated bootstrap method exhibited superior performance compared to the other methods. Last, we used the bootstrap methods to calculate the confidence interval for the parameter of the zero-truncated Poisson-Shanker distribution via three numerical examples, and the results of these examples matched those from the simulation study.

KEYWORDS: interval estimation, count data, Shanker distribution, bootstrap interval, simulation

*Corresponding Author: wararit@mathstat.sci.tu.ac.th

Received: 11/02/2023; Revised: 17/08/2023; Accepted: 25/08/2023

1. INTRODUCTION

The Poisson distribution is a discrete probability distribution that quantifies the likelihood of a certain number of events taking place within specified time or spaces. (Kissell and Poserina, 2017; Siegel and Wanger, 2022). This distribution is used to model data like the number of orders a firm will receive tomorrow, the number of people who will apply for a job tomorrow, the number of defects on a finished product, the number of confirmed COVID-19 cases per day, the number of customers entering a post office on a given day, etc. (Siegel, 2016; Ross, 2006).

The probability mass function (pmf) of a Poisson distribution is defined as

$$p(x; \lambda) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, 2, \dots; \lambda > 0 \quad (1)$$

where e is a constant approximately equal to 2.71828 and λ is the mean number of events within a given interval of time or space. The Poisson distribution is

frequently employed to model the number of occurrences under the assumption that events are independent and occur at a constant rate. In real situations, however, this assumption may not always hold true. The typical Poisson distribution may be inadequate if the underlying rate of occurrence is heterogeneous or variable. The mixed Poisson distribution lets the rate parameter be random, dealing with this constraint. Instead of having a constant rate parameter, the rate parameter is assumed to follow a certain distribution.

Shanker (2016) introduced the mixed Poisson distribution known as the Poisson-Shanker (PS) distribution and investigated its mathematical and statistical properties. Combining the Poisson distribution and the Shanker distribution (Shanker, 2015) yields this distribution. When the Poisson parameter λ follows a Shanker distribution, the Poisson distribution gives rise to the PS distribution

(Shanker, 2016). The pmf of the PS distribution is given by

$$p_0(x; \theta) = \frac{\theta^2}{\theta^2 + 1} \frac{x + (\theta^2 + \theta + 1)}{(\theta + 1)^{x+2}}, x = 0, 1, 2, \dots; \theta > 0. \quad (2)$$

The mathematical and statistical properties of the PS distribution for modeling count data were established by Shanker (2016). When applied to several real-world data sets, the PS distribution proved to be more appropriate than the Poisson and Poisson-Lindley (Sankaran, 1970) distributions.

The Shanker distribution was introduced by Shanker (2015) as a lifetime continuous distribution with a probability density function (pdf) defined as

$$f(\lambda; \theta) = \frac{\theta^2}{\theta^2 + 1} (\theta + \lambda) e^{-\theta \lambda}, \lambda > 0, \theta > 0. \quad (3)$$

It is a composite of an exponential distribution with a scale parameter θ , and a gamma distribution with a shape parameter 2 and a scale parameter θ , with respective proportions $\theta^2 / (\theta^2 + 1)$ and $1 / (\theta^2 + 1)$. Shanker (2015) showed that the pdf in Eq.(3) provides a better model than the exponential and Lindley (Lindley, 1958) distributions for modeling lifetime data.

The probability model of PS distribution can be utilized to analyze data that contains both zero and positive values with low occurrence probabilities within a predefined time or area range (Sangnawakij, 2021). However, probability models can become truncated when a range of possible values for the variables is either disregarded or impossible to observe. Zero truncation is commonly employed when analyzing count data without zeros. David and Johnson (1952) developed the zero-truncated (ZT) Poisson (ZTP) distribution, which has been applied to datasets such as the length of hospital stays, the number of published journal articles across various disciplines, the number of children ever born to mothers over 40 years old, and the number of passengers in cars (Hussain, 2020). A ZT distribution's pmf can be derived as

$$p(x; \theta) = \frac{p_0(x; \theta)}{1 - p_0(0; \theta)}, x = 1, 2, 3, \dots \quad (4)$$

where $p_0(x; \theta)$ is the pmf of the un-truncated distribution. Shanker (2017) proposed the zero-truncated Poisson-Shanker (ZTPS) distribution and its properties, such as the moment, coefficient of variation, skewness, kurtosis and the dispersion index. The method of moments and the maximum likelihood have also been derived for estimating its parameter. Furthermore, when the ZTPS distribution was applied to real data set, it was more suitable than the ZTP and zero-truncated Poisson-Lindley (ZTPL) (Shanker et al., 2015) distributions.

To our knowledge, no research has been conducted to estimate the confidence interval for the ZTPS distribution parameter θ . Bootstrap methods provide a way to estimate confidence intervals by quantifying uncertainties in statistical inferences based on a sample of data. The concept is to run a simulation study based on the actual data for estimating the likely extent of sampling error (Wood, 2004). Therefore, the objective of the current study is to assess the efficiencies of four bootstrap methods, namely the percentile bootstrap (PB), the simple bootstrap (SB), the bias-corrected and accelerated (BCa) bootstrap, and bootstrap-t (B-t) to estimate the confidence interval for the parameter (θ) of the ZTPS distribution. Due to a theoretical comparison is not possible, a simulation study was conducted to compare their performance and used the results to determine the best performing method based on the coverage probability, lower and upper error rate and the average length.

2. THEORETICAL BACKGROUND

Compounding probability distributions is a robust and innovative technique for obtaining new probability distributions that better fit datasets not adequately modeled by common parametric distributions. Shanker (2017) proposed a novel compounding distribution by combining the Poisson distribution with the Shanker distribution, aiming to find a more flexible model for analyzing statistical data. The pmf of the PS distribution is given in Eq.(2).

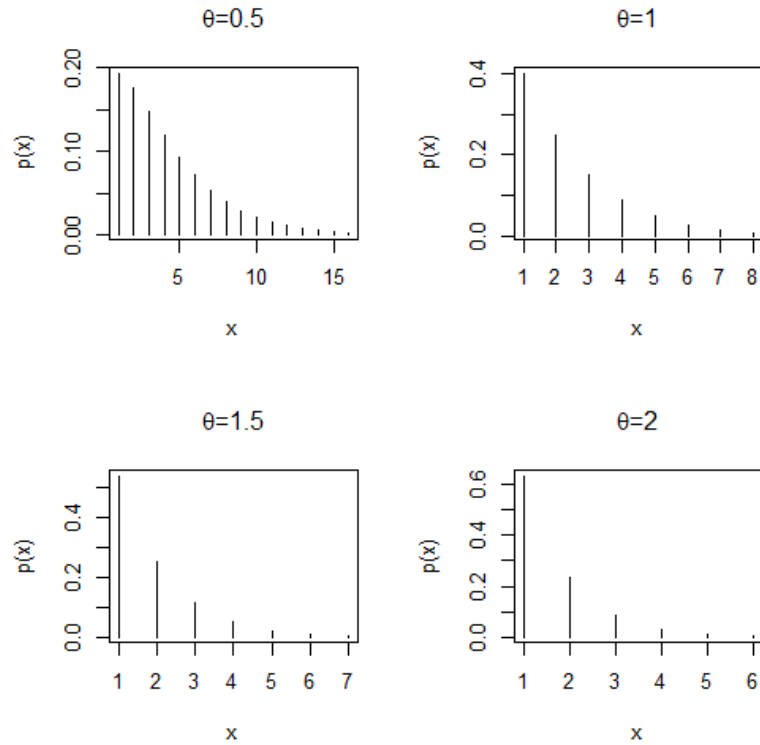


Figure 1 The plots of the pmf of the ZTPS distribution with $\theta=0.5, 1, 1.5,$ and 2 .

Table 1 Empirical coverage probability and average length of the 95% bootstrap confidence intervals for the parameter of the ZTPS distribution

n	θ	Coverage probability				Average length			
		PB	SB	BCa	B-t	PB	SB	BCa	B-t
10	0.25	0.900	0.885	0.912	0.900	0.2795	0.2801	0.2552	0.2328
	0.5	0.882	0.879	0.881	0.867	0.7057	0.7049	0.6318	0.5155
	1	0.879	0.858	0.883	0.786	2.2385	2.2524	2.0565	1.2764
	1.5	0.920	0.846	0.928	0.750	3.9184	3.9368	3.9010	2.9430
	2	0.961	0.833	0.953	0.716	5.1095	5.1023	5.0118	4.7787
30	0.25	0.923	0.921	0.921	0.920	0.1453	0.1450	0.1400	0.1379
	0.5	0.939*	0.918	0.944*	0.933	0.3184	0.3183	0.3056	0.2979
	1	0.912	0.898	0.923	0.907	0.8177	0.8211	0.7607	0.7073
	1.5	0.926	0.903	0.933	0.909	1.6302	1.6218	1.4655	1.2649
	2	0.897	0.899	0.903	0.919	3.2230	3.2036	2.7655	2.2254
50	0.25	0.944*	0.929	0.940*	0.932	0.1112	0.1112	0.1089	0.1081
	0.5	0.938	0.933	0.940*	0.932	0.2438	0.2437	0.2377	0.2344
	1	0.928	0.918	0.931	0.923	0.5870	0.5875	0.5643	0.5429
	1.5	0.934	0.907	0.941*	0.922	1.1121	1.1118	1.0438	0.9744
	2	0.941*	0.908	0.934	0.924	1.8518	1.8520	1.7044	1.5345
100	0.25	0.948*	0.942*	0.943*	0.948*	0.0777	0.0776	0.0769	0.0767
	0.5	0.939*	0.940*	0.939*	0.939*	0.1699	0.1696	0.1677	0.1667
	1	0.936	0.943*	0.941*	0.941*	0.3998	0.3999	0.3922	0.3855
	1.5	0.960*	0.949*	0.958*	0.952*	0.7267	0.7271	0.7061	0.6844
	2	0.938	0.921	0.936	0.929	1.1647	1.1686	1.1229	1.0735
500	0.25	0.941*	0.947*	0.934	0.946*	0.0346	0.0345	0.0344	0.0344
	0.5	0.944*	0.943*	0.948*	0.945*	0.0747	0.0746	0.0745	0.0743
	1	0.960*	0.950*	0.955*	0.955*	0.1740	0.1742	0.1736	0.1733
	1.5	0.948*	0.959*	0.948*	0.953*	0.3109	0.3114	0.3089	0.3087
	2	0.949*	0.938	0.948*	0.949*	0.4705	0.4709	0.4659	0.4638

* represents the empirical coverage probability is greater than or equal to the nominal confidence level.

ZTPS distribution (Shanker, 2017) with parameter θ , it is denoted as $X \sim \text{ZTPS}(\theta)$. Using Eqs. (2) and (4), the pmf of the ZTPS distribution can be obtained as

$$p(x; \theta) = \frac{\theta^2}{\theta^3 + \theta^2 + 2\theta + 1} \frac{x + (\theta^2 + \theta + 1)}{(\theta + 1)^x}, \quad x = 1, 2, 3, \dots; \theta > 0.$$

Figure 1 shows the graphs of the pmf of the ZTPS distribution with some specified parameter values θ . We noticed that the parameter values θ affect the form of Figure 1. When the parameter value is larger, the $p(1)$ is larger and $p(x) \rightarrow 0$ for the large value of x . The expected value and variance of X are as follows:

$$E(X) = \mu = \frac{(\theta + 1)(\theta^3 + \theta^2 + 2\theta + 2)}{\theta(\theta^3 + \theta^2 + 2\theta + 1)},$$

and

$$\text{Var}(X) = \sigma^2 = \frac{(\theta + 1)(\theta^6 + 2\theta^5 + 6\theta^4 + 9\theta^3 + 10\theta^2 + 8\theta + 2)}{\theta^2(\theta^3 + \theta^2 + 2\theta + 1)^2}.$$

The point estimator of θ is obtained by maximizing the log-likelihood function $\log L(x_i; \theta)$ or the logarithm of joint pmf of $X_1, \dots, X_n$. Therefore, the maximum likelihood (ML) estimator for θ of the ZTPS distribution is derived by the following processes:

$$\begin{aligned} \frac{\partial}{\partial \theta} \log L(x_i; \theta) &= \frac{\partial}{\partial \theta} \left[n \log \left(\frac{\theta^2}{\theta^3 + \theta^2 + 2\theta + 1} \right) - \sum_{i=1}^n x_i \log(\theta + 1) + \sum_{i=1}^n \log [x_i + (\theta^2 + \theta + 1)] \right] \\ &= \frac{2n}{\theta} - \frac{n(3\theta^2 + 2\theta + 2)}{(\theta^3 + \theta^2 + 2\theta + 1)} - \frac{n\bar{x}}{\theta + 1} + \sum_{i=1}^n \frac{2\theta + 1}{x_i + (\theta^2 + \theta + 1)}. \end{aligned}$$

Solving the equation $\frac{\partial}{\partial \theta} \log L(x_i; \theta) = 0$ for θ ,

we have the non-linear equation

$$\frac{2n}{\theta} - \frac{n(3\theta^2 + 2\theta + 2)}{(\theta^3 + \theta^2 + 2\theta + 1)} - \frac{n\bar{x}}{\theta + 1} + \sum_{i=1}^n \frac{2\theta + 1}{x_i + (\theta^2 + \theta + 1)} = 0.$$

where $\bar{x} = \sum_{i=1}^n x_i / n$ denotes the sample mean. Since the ML estimator for θ does not provide the closed-form solution, the non-linear equation can be solved by numerical iteration methods such as Newton-Raphson method, bisection method and Ragula-Falsi method. In this research, we use maxLik package (Henningsen and Toomet, 2011) with Newton-Raphson method for ML estimation in the statistical software R. The R code for estimating the parameter of the ZTPS distribution is shown in Appendix.

3. BOOTSTRAP METHODS

Confidence intervals are obtained from a parametric estimator of the standard errors of a quantity of interest ϕ . Then, the $(1 - \alpha)100\%$ confidence interval for ϕ is obtained by adding or subtracting the standard error multiplied by a critical value (for example, $\hat{\phi} \pm z_{1-(\alpha/2)} SE(\hat{\phi})$). This calculation assumes that the distribution of the estimator of ϕ is approximately normal (Flowers-Cano et al., 2018). However, there are several situations in which the assumption of normality is violated. In these cases, or when the standard error is very difficult to estimate, one alternative is to use techniques based on the bootstrap method (Van den Boogaard and Hall, 2004). The computer-intensive bootstrap methods described in this study provide an alternative for constructing approximate confidence intervals without assuming the underlying distribution (Meeker, 2007). Furthermore, it is essential to note that the score function of ZTPS is complicated, and the maximum likelihood has no closed form. Therefore, likelihood-based, score, and Wald-type confidence intervals have no closed forms. In such cases, finding these confidence intervals can be challenging; alternative methods, such as numerical techniques or resampling methods like the bootstrap, can be utilized.

In this paper, we focus on the four bootstrap methods for estimating confidence interval for the parameter of the ZTPS distribution. In practice, the popular bootstrap methods are the percentile bootstrap, the simple bootstrap, the bias-corrected and accelerated bootstrap and bootstrap-t methods. See the details of some bootstrap methods in DiCiccio and Efron (1996) and Manoharan et al. (2017).

3.1 Percentile bootstrap (PB) method

The PB two-sided confidence interval is the range between the $(\alpha/2) \times 100$ and $(1 - (\alpha/2)) \times 100$ percentiles of the resampled distribution of θ estimations or the distribution of $\hat{\theta}^*$, where θ

Table 2 Lower and upper error rates of the 95% bootstrap confidence intervals for the parameter of the ZIPS distribution

n	θ	Lower error rates				Upper error rates			
		PB	SB	BCa	B-t	PB	SB	BCa	B-t
10	0.25	0.077	0.005	0.057	0.044	0.023	0.110	0.031	0.056
	0.5	0.101	0.002	0.085	0.079	0.017	0.119	0.034	0.054
	1	0.100	0.000	0.089	0.159	0.021	0.142	0.029	0.055
	1.5	0.060	0.000	0.040	0.180	0.020	0.154	0.032	0.070
	2	0.002	0.000	0.001	0.207	0.038	0.167	0.045	0.077
30	0.25	0.056	0.008	0.045	0.042	0.021	0.071	0.034	0.038
	0.5	0.038	0.004	0.023	0.027	0.023	0.078	0.033	0.040
	1	0.067	0.000	0.047	0.051	0.021	0.102	0.030	0.042
	1.5	0.057	0.000	0.042	0.054	0.017	0.097	0.025	0.037
	2	0.089	0.000	0.069	0.042	0.014	0.101	0.028	0.039
50	0.25	0.035	0.007	0.032	0.024	0.021	0.064	0.028	0.044
	0.5	0.041	0.002	0.030	0.031	0.021	0.065	0.030	0.037
	1	0.048	0.000	0.035	0.034	0.024	0.082	0.034	0.043
	1.5	0.049	0.000	0.033	0.044	0.017	0.093	0.026	0.034
	2	0.044	0.000	0.038	0.041	0.015	0.092	0.028	0.035
100	0.25	0.031	0.013	0.032	0.025	0.021	0.045	0.025	0.027
	0.5	0.039	0.010	0.030	0.029	0.022	0.050	0.031	0.032
	1	0.049	0.003	0.035	0.036	0.015	0.054	0.024	0.023
	1.5	0.030	0.000	0.022	0.026	0.010	0.051	0.020	0.022
	2	0.044	0.000	0.034	0.039	0.018	0.079	0.030	0.032
500	0.25	0.040	0.024	0.042	0.035	0.019	0.029	0.024	0.019
	0.5	0.030	0.019	0.027	0.028	0.026	0.038	0.025	0.027
	1	0.023	0.015	0.020	0.022	0.017	0.035	0.025	0.023
	1.5	0.042	0.017	0.040	0.036	0.010	0.024	0.012	0.011
	2	0.018	0.003	0.015	0.015	0.033	0.059	0.037	0.036

Table 3 The descriptive measures of the number of unrest events

Min	Mean	Median	Var	Skewness	Kurtosis	Q1	Q3	Max
1.000	6.714	6.000	17.915	0.546	2.183	3.750	9.500	15.000

Table 4 The the log-likelihood (Log L), AIC and BIC for the number of unrest events

Distribution	Log L	AIC	BIC
ZIPS	-77.328	156.657	157.989
ZTPL	-77.552	157.104	158.436
Poisson	-86.529	175.057	176.390

Table 5 The number of unrest events in the southern border area of Thailand

Number of unrest events	1-3	4-5	6-7	≥ 8
Observed frequency	7	6	7	8
Expected frequency	8.7702	5.3505	4.2138	9.6656

Table 6 The 95% confidence intervals and their corresponding lengths for the parameter in the unrest events example, using all intervals

Methods	Confidence intervals	Lengths
PB	(0.2463, 0.4023)	0.1560
SB	(0.2102, 0.3651)	0.1549
BCa	(0.2430, 0.3946)	0.1516
B-t	(0.2375, 0.3818)	0.1443

Table 7 The descriptive measures of the number of fertile mothers

Min	Mean	Median	Var	Skewness	Kurtosis	Q1	Q3	Max
1.000	1.607	1.000	1.091	1.941	6.446	1.000	2.000	6.000

significance (e.g., $\alpha = 0.05$ for 95% confidence intervals) (Efron, 1982). Obtaining a PB confidence interval for θ is as follows:

- 1) With a replacement, B random bootstrap samples of the underlying distribution are created, where B is the number of bootstrap replications,
- 2) From each bootstrap sample, a parameter estimate $\hat{\theta}^*$ is determined,
- 3) All parameter estimates from the B bootstrap are ranked from lowest to highest, and
- 4) The $(1-\alpha)100\%$ PB two-sided confidence interval for θ is created as follows:

$$CI_{PB} = [\hat{\theta}_{(r)}^*, \hat{\theta}_{(s)}^*], \quad (5)$$

where $\hat{\theta}_{(r)}^*$ is the r^{th} quantile of a set of quantiles ordered from lowest to highest, $\hat{\theta}_{(s)}^*$ is the s^{th} quantile of the same set, $r = [(\alpha/2)B]$, $s = [(1-(\alpha/2))B]$, and $1-\alpha$ is the confidence level. This study utilized $B = 1,000$ and $\alpha = 0.05$; the quantile corresponding to the lower limit of the confidence interval was $\hat{\theta}_{(r)}^* = \hat{\theta}_{(25)}^*$, and the quantile corresponding to the upper limit was $\hat{\theta}_{(s)}^* = \hat{\theta}_{(975)}^*$.

3.2 Simple bootstrap (SB) method

The SB method is as straightforward to implement as the PB method and is sometimes referred to as the simple bootstrap method. Consider the quantity of interest to be θ and the estimator of θ to be $\hat{\theta}$. The straightforward bootstrap method implies that the distributions of $\hat{\theta} - \theta$ and $\hat{\theta}^* - \hat{\theta}$ are roughly equivalent (Meeker et al., 2017). The $(1-\alpha)100\%$ SB two-sided confidence interval for θ is

$$CI_{SB} = [2\hat{\theta} - \hat{\theta}_{(s)}^*, 2\hat{\theta} - \hat{\theta}_{(r)}^*], \quad (6)$$

where the quantiles $\hat{\theta}_{(r)}^*$ and $\hat{\theta}_{(s)}^*$ represent the same percentile of the empirical distribution of bootstrap estimates $\hat{\theta}^*$ that are utilized in Eq. (5) to calculate the PB confidence interval.

3.3 Bias-corrected and accelerated (BCa) bootstrap method

The calculation of BCa bootstrap confidence intervals commonly involves the utilization of influence statistics

derived from jackknife simulations. However, incorporating jackknife simulation alongside ordinary bootstrapping is computationally expensive for the intended purposes. The BCa bootstrap confidence interval uses a bias-correction factor and an acceleration factor to correct for the bias and skewness of the bootstrap parameter estimates, mitigating the over-coverage problems seen with the PB confidence interval (Efron, 1987; Efron and Tibshirani, 1993). Davison and Hinkley (1997) and Chernick and LaBudde (2011) described the mathematical particulars of the BCa adjustment. The bias-correction factor $\hat{z}_0$ calculated as the proportion of bootstrap estimates that are smaller than the original parameter estimate $\hat{\theta}$,

$$\hat{z}_0 = \Phi^{-1} \left(\frac{\#\{\hat{\theta}^* \leq \hat{\theta}\}}{B} \right),$$

where Φ^{-1} is the inverse function of a standard normal cumulative distribution function (e.g., $\Phi^{-1}(0.975) \approx 1.96$). The acceleration factor $\hat{a}$ is estimated through jackknife resampling, also known as ‘leave-one-out’ resampling. This technique involves generating n replicates of the original sample, where n represents the number of observations in the sample. The initial jackknife replicate is derived by excluding the first case ($i=1$) from the original sample. Subsequently, the second replicate is obtained by excluding the second case ($i=2$), and this process continues until n samples of size $n-1$ are generated. The value of $\hat{\theta}_{(-i)}$ is obtained for each of the jackknife resamples. The mean of these estimates is denoted by $\hat{\theta}_{(\cdot)} = \sum_{i=1}^n \hat{\theta}_{(-i)} / n$. The acceleration factor $\hat{a}$ is then determined as follows:

$$\hat{a} = \frac{\sum_{i=1}^n (\hat{\theta}_{(\cdot)} - \hat{\theta}_{(-i)})^3}{6 \left\{ \sum_{i=1}^n (\hat{\theta}_{(\cdot)} - \hat{\theta}_{(-i)})^2 \right\}^{3/2}}.$$

With the values of $\hat{z}_0$ and $\hat{a}$, the values α_1 and α_2 are calculated,

Table 8 The the log-likelihood (Log L), AIC and BIC for the number of fertile mothers

Distribution	Log L	AIC	BIC
ZTPS	-143.976	289.951	292.856
ZTPL	-144.091	290.182	293.087
Poisson	-191.055	384.110	387.015

Table 9 The number of fertile mothers who have experienced at least one child death

Number of child deaths	1	2	3	≥ 4
Observed frequency	89	25	11	10
Expected frequency	83.4756	32.3839	12.2451	6.8953

Table 10 The 95% confidence intervals and their corresponding lengths for the parameter in the demographic example, using all intervals

Methods	Confidence intervals	Lengths
PB	(1.5810, 2.5710)	0.9900
SB	(1.3739, 2.3301)	0.9562
BCa	(1.5567, 2.4995)	0.9428
B-t	(1.5527, 2.4537)	0.9010

Table 11 The descriptive measures of the number of flower heads as per the number of fly eggs

Min	Mean	Median	Var	Skewness	Kurtosis	Q1	Q3	Max
1.000	3.034	3.000	3.344	0.812	3.135	1.750	4.000	9.000

Table 12 The the log-likelihood (Log L), AIC and BIC for the number of flower heads as per the number of fly eggs

Distribution	Log L	AIC	BIC
ZTPS	-167.154	336.308	338.785
ZTPL	-167.381	336.763	339.240
Poisson	-171.281	344.561	347.039

Table 13 The number of flower heads as per the number of fly eggs

Number of fly eggs	1	2	3	4	5	6	≥ 7
Observed frequency	22	18	18	11	9	6	4
Expected frequency	24.9287	19.7204	14.6526	10.2922	6.9078	4.4711	7.0272

Table 14 The 95% confidence intervals and corresponding lengths using all methods for the parameter in the flower heads example

Methods	Confidence intervals	Lengths
PB	(0.6342, 0.8684)	0.2342
SB	(0.6045, 0.8312)	0.2267
BCa	(0.6323, 0.8597)	0.2274
B-t	(0.6253, 0.8525)	0.2272

$$\alpha_1 = \Phi \left\{ \hat{z}_0 + \frac{\hat{z}_0 + z_{\alpha/2}}{1 - \hat{a}(\hat{z}_0 + z_{\alpha/2})} \right\} \quad \text{and}$$

$$\alpha_2 = \Phi \left\{ \hat{z}_0 + \frac{\hat{z}_0 + z_{1-\alpha/2}}{1 - \hat{a}(\hat{z}_0 + z_{1-\alpha/2})} \right\},$$

where $z_{\alpha/2}$ is the $\alpha/2$ quantile of the standard normal distribution (e.g. $z_{0.05/2} \approx -1.96$). Then, the $(1-\alpha)100\%$ BCa bootstrap two-sided confidence interval for θ is as follows

$$CI_{BCa} = [\hat{\theta}_{(j)}^*, \hat{\theta}_{(k)}^*], \quad (7)$$

where $j = [\alpha_1 B]$ and $k = [\alpha_2 B]$. When the value $\hat{z}_0$ is set to 0 and $\hat{a}$ is also set to 0, it can be observed that the BCa confidence interval is equal to the PB confidence interval.

3.4 Bootstrap-t (B-t) method

Assuming the quantity of interest is θ , one can compute the estimate $\hat{\theta}$ and its corresponding estimate of the standard error $s.e.(\hat{\theta})$ from the given data. The bootstrap estimates $\hat{\theta}_j^*$ and their estimated standard errors $s.e.(\hat{\theta}_j^*)$ are then computed for each bootstrap

sample, $j = 1, 2, \dots, B$. The bootstrap-t (studentized) statistics are derived from these values;

$$R_j^* = \frac{\hat{\theta}_j^* - \hat{\theta}}{s.e.(\hat{\theta}_j^*)}, \quad j = 1, 2, \dots, B.$$

The $(1-\alpha)100\%$ bootstrap-t two-sided confidence interval for θ is

$$CI_{B-t} = [\hat{\theta} - t_r^* \times s.e.(\hat{\theta}), \hat{\theta} + t_s^* \times s.e.(\hat{\theta})], \quad (8)$$

where $r = 1 - (\alpha/2)$ and $s = \alpha/2$, and t_q^* denotes the q quantile of the distribution of R_j^* .

4. SIMULATION STUDY AND RESULTS

The confidence interval for the parameter of a ZTPS distribution estimated via various bootstrap methods was considered in this study. Due to the absence of a theoretical comparison, a Monte Carlo simulation study was conducted using R version 4.2.2 (Ihaka and Gentleman, 1996). The random samples from the ZTPS distribution were generated using the accept-reject method (Robert and Casella, 2004). The simulation covered different sample sizes ($n = 10, 30, 50, 100$, and 500) to observe the effects of small and large variances. To observe the effect of small and large variances, the true parameter (θ) was set as $0.25, 0.5, 1, 1.5$, and 2 (the variance of the random variables decreases as the value of θ increases). For each simulation, 1,000 bootstrap samples of size n were generated from the original sample, and the simulation was repeated 1,000 times. To evaluate the performance of the bootstrap methods, the empirical coverage probabilities, lower and upper error rates, and average lengths were compared. A bootstrap method with a coverage probability close to or greater than the nominal confidence level $(1-\alpha)$ of 0.95 indicates that it contains the true value and provides a precise estimate of the confidence interval for the parameter. In this study, we can conclude that a coverage probability is greater than or equal to the nominal confidence level when the empirical coverage probability is greater than or equal to 0.939 by using the one-proportion z -test with a significance level of 0.95 . Moreover, the

shortest average length can be used to more accurately estimate the bootstrap confidence interval for the parameter.

The empirical coverage probability and average length from the simulation study are presented in Table 1. For $n=10, 30$ and 50 , the empirical coverage probability of the four approaches had a tendency to go below 0.95 and did not attain the expected nominal confidence level. However, in these scenarios, the BCa bootstrap method provided higher empirical coverage probabilities than the other methods. For $n \geq 100$, all four methods performed comparably well in terms of empirical coverage probability, and they provided empirical coverage probabilities greater than 0.95 in almost all of situations. In this case ($n \geq 100$), the BCa bootstrap method and B-t method produced shorter lengths on average than the PB and SB methods. Therefore, when the size of the sample was enlarged, the empirical coverage probabilities of the bootstrap methods exhibited a tendency to rise and converge towards 0.95 . Furthermore, the average lengths of the bootstrap methods exhibited a reduction when the value of θ declined, reflecting the relationship between the variance and θ . As expected, the average lengths of the four approaches dropped with an increase in the sample size. While the B-t method demonstrated the lowest average length for small sample sets, it exhibited an empirical coverage probability that was notably lower than the nominal confidence level. In brief, the BCa bootstrap method demonstrated persistent superior performance in terms of empirical coverage probability across many scenarios.

Table 2 shows the lower and upper error rates of the bootstrap confidence intervals. The lower error rates of the PB and BCa bootstrap methods exceeded the upper error rates across all scenarios. On the other hand, in every situation, the lower error rates of the SB method are lower than the upper error rates. Furthermore, there is no substantial difference observed between the lower and upper error rates in the BCa bootstrap method.

The ability of the ZTPS distribution to fit data was compared to the Poisson and ZTPL distributions (Shanker et al., 2015). In order to compare distributions, we have taken into account the log-likelihood (Log L), Akaike's information criterion (AIC) (Akaike, 1973), and Bayesian information criterion (BIC) (Schwarz, 1978) as provided in Table 12. The AIC and BIC values indicate that the ZTPS distribution fits the number of flower heads as per the number of fly eggs better than other distributions.

The bootstrap method resamples using replacement from observed data to estimate the sampling distribution of a statistic. However, there are certain limitations to using the bootstrap method, especially when dealing with small sample sizes. With a small sample size, we have limited information about the underlying population. This means that resampling from the small dataset may not adequately capture the variability present in the population, leading to biased or unreliable estimates. Additionally, the coverage probability of bootstrap confidence intervals might deviate from the desired level. Bootstrap confidence intervals may end up too narrow or too wide, leading to incorrect inferences about the parameter. On the other hand, the bootstrap method becomes more effective at approximating the sampling distribution of a statistic as the sample size increases. This is because with a large sample, the bootstrap resamples from the original data, capturing the underlying distribution of the population better.

5. CONCLUSIONS AND DISCUSSION

We present four bootstrap methods, PB, SB, BCa, and B-t, to estimate the confidence interval for the ZTPS distribution parameter. The coverage probabilities of all bootstrap confidence intervals were markedly lower than 0.95 when the sample size was 10, 30, or 50. However, when the sample size was sufficiently large ($n \geq 100$), the coverage probabilities and average lengths derived from four bootstrap methods did not differ significantly. Based on our findings, the BCa bootstrap method consistently outperformed the others in terms of the empirical coverage probability and average length for almost all situations, as confirmed by

The authors would like to express their gratitude for the reviewers' dedication and commitment in enhancing the overall quality of this manuscript.

REFERENCES

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & B. F. Csaki (Eds.), *Second International Symposium on Information Theory* (pp. 267–281). Akademiai Kiado: Budapest.
- Chernick, M. R., & LaBudde, R. A. (2011). *An Introduction to Bootstrap Methods with Applications to R*. New York: John Wiley & Sons.
- David, F. N., & Johnson, N. L. (1952). The truncated Poisson. *Biometrics*, 8(4), 275–285.
- Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge: Cambridge University Press.
- DiCiccio, T. J., & Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science*, 11(3), 189–212.
- Efron, B. (1982). The jackknife, the bootstrap, and other resampling plans. *CBMS-NSF Regional Conference Series in Applied Mathematics*, Monograph 38. SIAM.
- Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association*, 82(397), 171–185.
- Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. New York: Chapman and Hall.
- Finney, D. J., & Varley, G. C. (1955). An example of the truncated Poisson distribution. *Biometrics*, 11(3), 387–394.
- Flowers-Cano, R. S., Ortiz-Gómez, R., León-Jiménez, J. E., López Rivera, R., & Perera Cruz, L. A. (2018). Comparison of bootstrap confidence intervals using Monte Carlo simulations. *Water*, 10(2), 166. <https://doi.org/10.3390/w10020166>
- Ghitany, M. E., Al-Mutairi, D. K., & Nadarajah, S. (2008). Zero-truncated Poisson-Lindley distribution and its application. *Mathematics and Computers in Simulation*, 79(3), 279–287.
- Henningsen, A., & Toomet, O. (2011). A package for maximum likelihood estimation in R. *Computational Statistics*, 26(3), 443–458.
- Hussain, T. (2020). A zero truncated discrete distribution: Theory and applications to count data. *Pakistan Journal of Statistics and Operation Research*, 16(1), 167–190.
- Ihaka, R., & Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5(3), 299–314.
- Kissell, R., & Poserina, J. (2017). *Optimal sports math, statistics, and fantasy*. London: Academic Press.
- Lindley, D. V. (1958). Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society: Series B*, 20(1), 102–107.
- Meeker, W. Q., Hahn, G. J., & Escobar, L. A. (2017). *Statistical intervals: A guide for practitioners and researchers*. New Jersey: John Wiley and Sons.
- Manoharan, T., Arasan, J. Midi, H., & Adam, M. B. (2017). Bootstrap intervals in the presence of left-truncation, censoring and covariates with a parametric distribution. *Sains Malaysiana*, 46(12), 2529–2539.
- Robert, C. P., & Casella, G. (2004). *Monte Carlo Statistical Methods*. New York: Springer.
- Ross, S. (2006). *A First Course in Probability*. New Jersey: Pearson Education.
- Sankaran, M. (1970). The discrete Poisson-Lindley distribution. *Biometrics*, 26(1), 145–149.
- Sangnawakij, P. (2021). Confidence interval for the parameter of the zero-truncated Poisson distribution. *The Journal of Applied Science*, 20(2), 13–22.
- Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461–464.
- Shanker, R. (2015). Shanker distribution and its applications. *International Journal of Statistics and Applications*, 5(6), 338–348.
- Shanker, R. (2016). The discrete Poisson-Shanker distribution. *Jacobs Journal of Biostatistics*, 1(1), 1–7.
- Shanker, R. (2017). A zero-truncated Poisson-Shanker distribution and its applications. *International Journal of Statistics and Applications*, 7(3), 159–169.
- Shanker, R., Fesshaye, H., Sujatha, S., & Yemane, A. (2015). On zero-truncation of Poisson and Poisson-Lindley distributions and their applications. *Biometrics and Biostatistics International Journal*, 2(6), 168–181.
- Siegel, A. F., & Wanger, M. R. (2022). *Practical Business Statistics*. San Diego: Academic Press.
- Turhan, N. S. (2020). Karl Pearson's chi-square tests. *Educational Research Review*, 15(9), 575–580.
- Van den Boogaard, H. F. P., & Hall, M. J. (2004). The construction of confidence intervals for frequency analysis using resampling techniques: a supplementary note. *Hydrology and Earth System Sciences*, 8(6), 1174–1178.
- Wood, M. (2004). Statistical inference using bootstrap confidence intervals. *Significance*, 1(4), 180–182.

APPENDIX

```
library(maxLik)
theta.hat <- function(x) {
  options(warn=-1)
  logLikFun <- function(param) {
    th <- param
    n <- length(x)
    term1 <- n*log((th^2)/(th^3+th^2+2*th+1))
    term2 <- sum(x)*log(th+1)
    term3 <- sum(log(x+(th^2+th+1)))
    return(term1-term2+term3)
  }
  mle <- maxLik(logLik=logLikFun,start=0.01)
  return(c(mle$estimate,vcov(mle)))
}
```

Exact solutions of the fractional Landau–Ginzburg–Higgs equation and the $(3+1)$ -dimensional space–time fractional modified KdV–Zakharov–Kuznetsov equation using the simple equation method

Aungkanaporn Chankaew¹, Sirasrete Phoosree² and Jiraporn Sanjun^{3*}

¹Rajamangala University of Technology Rattanakosin, Thailand

²Education Program in Mathematics, Faculty of Education, Surattani Rajabhat University, Thailand

³Department of Mathematics, Faculty of Science and Technology, Surattani Rajabhat University, Thailand

ABSTRACT

In this article, we analyze the fractional Landau–Ginzburg–Higgs equation and the $(3+1)$ -dimensional space–time fractional modified KdV–Zakharov–Kuznetsov equation, two ion wave equations, using the simple equation approach and the Bernoulli equation. The traveling wave solutions are demonstrated by the exponential and hyperbolic functions, which can be interpreted as kink waves. Their graphical representations are three-dimensional graphs, and contour graphs are shown using appropriate parameter values. Additionally, the results demonstrated that the technique employed in this study is a powerful analytical tool for obtaining exact traveling wave solutions to nonlinear models encountered in a variety of scientific and engineering fields.

KEYWORDS: simple equation method, fractional partial differential equations, $(3+1)$ -dimensional space–time fractional modified KdV–Zakharov Kuznetsov equation, fractional Landau–Ginzburg–Higgs equation

*Corresponding Author: jiraporn.san@sru.ac.th

Received: 18/06/2023; Revised: 21/11/2023; Accepted: 23/11/2023

1. INTRODUCTION

Many phenomena of nature are associated with the ideas of physical science: shallow water waves (Islam et al., 2023), nonlinear optics (Jhangeer et al., 2021), plasma physics (Salas et al., 2022), fluid dynamics (Shah et al., 2022), quantum mechanics (Alabedalhadi, 2022), biology (Simpson, 1963), hydrodynamics (Zeng & Liu, 2022), and acoustics (Li et al., 2022). The present work concerns the representation of these phenomena by means of fractional differential equations (FDEs), which are used to simulate various physical phenomena and processes for solving linear and nonlinear problems (Çenesiz et al., 2017).

Many researchers have used diverse methods to get exact solutions of fractional differential equations, such as the Kudryashov method (Thadee et al., 2022), the modified Kudryshov method (Kumar et

al., 2018), the riccati-sub equation method (Khodadad et al., 2017), the exp-function method (Islam et al., 2019), the first integral method (Eslami et al., 2014), the (G'/G) - expansion method (Baishya & Rangarajan, 2018), the new extended direct algebraic method (Rezazadeh et al., 2020), the extended auxiliary equation method (Sabi'u et al., 2023), the auxiliary equation method (Rezazadeh et al., 2020), and the Sine-Cosine method (Sabi'u et al., 2019), etc.

In the present work, we take into consideration the fractional Landau–Ginzburg–Higgs equation (Guner et al., 2017),

$$D_t^{2\alpha} u - D_x^{2\alpha} u - m^2 u + n^2 u^3 = 0, \quad 0 < \alpha \leq 1, \quad (1.1)$$

where $u = u(x, t)$, m and n are parameters. Furthermore, we analyze the $(3+1)$ -dimensional

space-time fractional modified KdV-Zakharov-Kuznetsov equation (Sahoo & Ray, 2015),

$$D_t^\alpha u + \delta u^2 D_x^\alpha u + D_x^{3\alpha} u + D_x^\alpha D_y^{2\alpha} u + D_x^\alpha D_z^{2\alpha} u = 0, \quad 0 < \alpha \leq 1, \quad (1.2)$$

where δ is an arbitrary constant and $D_t^\alpha u$ denotes Jumarie's modified Riemann-Liouville derivatives of u , where $u = u(x, y, z, t)$.

The fractional variational iteration transform method was used to analyze the analytical solution of the fractional Landau-Ginzburg-Higgs equation (Guner et al., 2017). The exact solution of the (3+1)-dimensional space-time fractional modified KdV-Zakharov-Kuznetsov equation was established by the improved generalized tanh-coth method in 2019 (Torvattanabun et al., 2019), the improved fractional sub-equation method in 2015 (Sahoo & Ray, 2015), the fractional variable method in 2017 (Çenesiz et al., 2017), and the unified method in 2019 (Osman et al., 2019), for which the fractional Landau-Ginzburg-Higgs equation and the (3+1)-dimensional space-time fractional modified KdV-Zakharov-Kuznetsov equation have not yet been analyzed using the simple equation method.

In this article, we have used Jumarie's modified Riemann-Liouville derivative and the simple equation method with the Bernoulli equation to solve the fractional Landau-Ginzburg-Higgs equation and the (3+1)-dimensional space-time fractional modified KdV-Zakharov-Kuznetsov equation. We have displayed the analytical solutions and the wave effects in a three-dimensional graph.

2. DESCRIPTIONS OF MODIFIED RIEMANN-LIOUVILLE DERIVATIVE AND THE PROPOSED METHOD

Jumarie's modified Riemann-Liouville derivative and the properties of the modified Riemann-Liouville derivative (Sahoo & Ray, 2015) of order α are defined by the expression

$$f^{(\alpha)}(x) = \lim_{h \rightarrow 0} \left(\frac{\sum_{k=0}^{\infty} (-1)^k \binom{\alpha}{k} f(x + (\alpha - k)h)}{h^\alpha} \right), \quad 0 < \alpha \leq 1, \quad (2.1)$$

which can be written as

$$D_x^\alpha f(x) = \begin{cases} \frac{1}{\Gamma(-\alpha)} \int_0^x (x - \xi)^{(-\alpha-1)} [f(\xi) - f(0)] d\xi & \text{if } \alpha < 0, \\ \frac{1}{\Gamma(1-\alpha)} \frac{d}{dx} \int_0^x (x - \xi)^{(-\alpha)} [f(\xi) - f(0)] d\xi & \text{if } 0 < \alpha < 1, \\ [f^{(\alpha-n)}(x)]^{(n)} & \text{if } n \leq \alpha < n+1, n \geq 1, \end{cases} \quad (2.2)$$

some properties of the modified Riemann-Liouville derivative as follows:

$$D_x^\alpha x^\gamma = \frac{\Gamma(\gamma+1)x^{\gamma-\alpha}}{\Gamma(\gamma+1-\alpha)}, \quad \gamma > 0, \quad x > 0, \\ D_x^\alpha (f(x)g(x)) = g(x)D_x^\alpha f(x) + f(x)D_x^\alpha g(x), \quad (2.3) \\ D_x^\alpha f(g(x)) = f'_g[g(x)]D_x^\alpha g(x) = D_g^\alpha f[g(x)](g'(x))^\alpha.$$

The aforementioned characteristics are crucial to the simple equation method (Phoosree & Thadee, 2022).

3. DESCRIPTIONS OF SIMPLE EQUATION METHOD

In this part, we will outline a straightforward technique for getting the traveling wave solution to fractional PDEs, known as the simple equation method. Assume that x, y and t are independent variables in a nonlinear partial differential equation that is given by

$$G(u, D_x^\alpha u, D_y^\alpha u, D_t^\alpha u, D_x^{2\alpha} u, D_y^\alpha D_x^\alpha u, D_t^\alpha D_x^\alpha u, \dots) = 0, \quad 0 < \alpha \leq 1, \quad (3.1)$$

where $u(x, y, t)$ is an unknown function and G is a polynomial of $u(x, y, t)$ and its derivatives. The main

steps of the SE method (Sanjun & Chankaew, 2022) are as follows:

Step 1. Wave transformation

We suppose that $u(x, y, t) = u(\xi)$,

$$\xi = \frac{kx^\alpha}{\Gamma(\alpha+1)} + \frac{ly^\alpha}{\Gamma(\alpha+1)} - \frac{\omega t^\alpha}{\Gamma(\alpha+1)}, \quad (3.2)$$

where constants k and l are non-zero constants, and ω is the speed of the traveling wave. We are capable of transforming Eq. (3.1) into an ordinary differential equation (ODE) for $u = u(\xi)$ using the traveling wave transformation Eq. (3.2) as follows:

$$Q(u, \frac{du}{d\xi}, \frac{d^2u}{d\xi^2}, \frac{d^3u}{d\xi^3}, \dots) = 0,$$

where Q is a polynomial of $u(\xi)$ and its derivatives, in which the prime indicates the derivative with respect to ξ .

Step 2. Solution Assumption

Assume that Eq. (3.3) has the following formal solution:

$$u(\xi) = \sum_{i=0}^N a_i F^i(\xi), \quad (3.4)$$

where $a_i (i = 0, 1, 2, \dots, N)$ is a constant that will be determined later. The function $F(\xi)$ satisfies the simple equations (ordinary differential equations). The Bernoulli equation, a well-known nonlinear ordinary differential equation, will be used in this study. Their solutions can be expressed as simple functions. The Bernoulli equation,

$$F'(\xi) = cF(\xi) + dF^2(\xi). \quad (3.5)$$

Where c and d are non-zero constants.

Step 3. balancing the integer N

The balance number N can be determined by balancing the highest-order derivative and nonlinear terms in Eq (3.3).

Step 4. Solution attainment

We get the general solutions of Eq. (3.5) as follows:

Case I: if $c > 0$ and $d < 0$, we get

$$F(\xi) = \frac{ce^{c(\xi+\xi_0)}}{1 - de^{c(\xi+\xi_0)}}, \quad (3.6)$$

where ξ_0 is the constant of the integration.

Case II: if $c < 0$ and $d > 0$, we get

$$F(\xi) = -\frac{ce^{c(\xi+\xi_0)}}{1 + de^{c(\xi+\xi_0)}}, \quad (3.7)$$

where ξ_0 is the constant of the integration.

4. APPLICATIONS

Now, we aim to solve the fractional Landau–Ginzburg–Higgs equation (1.1) and the (3+1)-dimensional space-time fractional modified KdV–Zakharov–Kuznetsov equation (1.2) by applying the simple equation method described above.

4.1 Solutions of the fractional Landau–Ginzburg–Higgs equation

The fractional Landau–Ginzburg–Higgs equation is

$$D_t^\alpha (D_x^\alpha u) - D_x^\alpha (D_x^\alpha u) - m^2 u + n^2 u^3 = 0, \quad 0 < \alpha \leq 1, \quad (4.1.1)$$

where m and n are parameters. We will reduce it to an ODE using the traveling wave variable

$$\xi = \frac{kx^\alpha}{\Gamma(\alpha+1)} - \frac{\omega t^\alpha}{\Gamma(\alpha+1)}.$$

The substitution of the transformation into equation (4.1.1) leads to:

$$(\omega^2 - k^2)u'' - m^2 u + n^2 u^3 = 0. \quad (4.1.2)$$

Next, we balanced the highest-order derivative terms and the highest nonlinear terms of Eq.

(4.1.2). Then $N=1$. We have the solution to Eq. (4.1.2) as follows:

$$u(\xi) = \sum_{i=0}^1 a_i F^i(\xi) = a_0 + a_1 F, \quad (4.1.3)$$

where F satisfies Eq. (3.5). Therefore, the expressions for u'' and u^3 are expressed as:

$$\begin{aligned} u'' &= a_1 c^2 F + 3a_1 c d F^2 + 2a_1 d^2 F^3, \\ u^3 &= a_0^3 + 3a_0^2 a_1 F + 3a_0 a_1^2 F^2 + a_1^3 F^3. \end{aligned} \quad (4.1.4)$$

Substituting Eqs. (4.1.3) and (4.1.4) into Eq. (4.1.2), then equating the coefficient of F^i to zero, where $i=0,1,2,3$, yields

$$\begin{aligned} F^0(\xi): n^2 a_0^3 - m^2 a_0 &= 0, \\ F^1(\xi): (\omega^2 - k^2) a_1 c^2 - m^2 a_1 + 3n^2 a_0^2 a_1 &= 0, \\ F^2(\xi): 3(\omega^2 - k^2) a_1 c d + 3n^2 a_0 a_1^2 &= 0, \\ F^3(\xi): 2(\omega^2 - k^2) a_1 d^2 + n^2 a_1^3 &= 0. \end{aligned} \quad (4.1.5)$$

Solving this system of algebraic equations, we obtain

$$\begin{aligned} a_0 &= \pm \frac{m}{n}, \quad a_1 = \pm \frac{d}{n} \sqrt{-2(\omega^2 - k^2)} \\ \text{and } \omega &= \pm \sqrt{\frac{k^2 c^2 - 2m^2}{c^2}}, \end{aligned} \quad (4.1.6)$$

where $-2(\omega^2 - k^2) \geq 0$ and $\frac{k^2 c^2 - 2m^2}{c^2} > 0$.

The following two exact solutions are obtained by substituting Eq. (4.1.6) into Eq. (4.1.3). Then, we utilize the general solutions to the Bernoulli equations (3.6) and (3.7). We get four exact solutions of (4.1.1) written in terms to the exponential function.

Case I: if $c > 0$ and $d < 0$, we get

$$u_{1,2}(x, t) = \pm \frac{m}{n} \left(1 + 2d \left(\frac{e^{c(\xi + \xi_0)}}{1 - d e^{c(\xi + \xi_0)}} \right) \right), \quad (4.1.7)$$

where $\xi = \frac{kx^\alpha}{\Gamma(\alpha+1)} - \frac{\omega t^\alpha}{\Gamma(\alpha+1)}$ and ξ_0 is a constant of the integration.

Case II: if $c < 0$ and $d > 0$, we get

$$u_{3,4}(x, t) = \pm \frac{m}{n} \left(1 - 2d \left(\frac{e^{c(\xi + \xi_0)}}{1 + d e^{c(\xi + \xi_0)}} \right) \right), \quad (4.1.8)$$

where $\xi = \frac{kx^\alpha}{\Gamma(\alpha+1)} - \frac{\omega t^\alpha}{\Gamma(\alpha+1)}$ and ξ_0 is a constant of the integration.

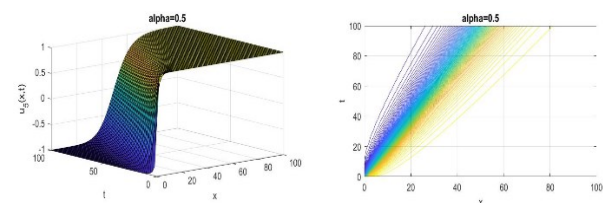
After that, we substitute $c = \pm 2$ and $d = \mp 1$ to get the following precise solutions in hyperbolic form:

$$u_{5,6}(x, t) = \pm \frac{m}{n} \tanh(\xi + \xi_0), \quad (4.1.9)$$

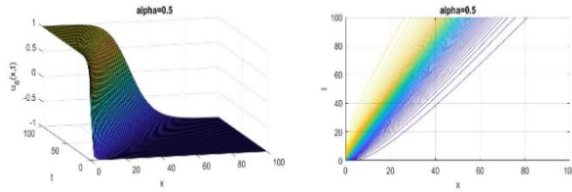
where $\xi = \frac{kx^\alpha}{\Gamma(\alpha+1)} - \frac{\omega t^\alpha}{\Gamma(\alpha+1)}$ and ξ_0 is a constant of the integration.

By using the simple equation method, we get the wave solution, which is expressed in terms of fractions of exponential functions, and then the wave solution is expressed in terms of hyperbolic functions when we set constants c and d .

In Figures 1 and 2, we present the three-dimensional and contour plots of some exact solutions, which are $u_5(x, t)$ and $u_6(x, t)$, expressed in Eq. (4.1.9). Employing $m=1, n=1, \alpha=0.5$ and $0 \leq x, t \leq 100$, we obtain the plots of the selected exact traveling wave solution, which represents the kink wave solutions.



Figures 1: kink wave solution of $u_5(x, t)$ in 3D and contour



Figures 2. kink wave solution of $u_6(x,t)$ in 3D and contour

4.2 Solutions of the (3+1) – dimensional space–time fractional modified KdV–Zakharov–Kuznetsov equation

The (3+1)–dimensional space–time fractional modified KdV–Zakharov–Kuznetsov equation is

$$D_t^\alpha v + \delta v^2 D_x^\alpha v + D_x^\alpha (D_x^\alpha (D_x^\alpha v)) + D_x^\alpha (D_y^\alpha (D_y^\alpha v)) + D_x^\alpha (D_z^\alpha (D_z^\alpha v)) = 0, \quad 0 < \alpha \leq 1, \quad (4.2.1)$$

where δ is a parameter. We will reduce it to an ODE using the traveling wave variable

$$\xi = \frac{kx^\alpha}{\Gamma(\alpha+1)} + \frac{ly^\alpha}{\Gamma(\alpha+1)} + \frac{mz^\alpha}{\Gamma(\alpha+1)} - \frac{\omega t^\alpha}{\Gamma(\alpha+1)}.$$

The substitution of the transformation into Eq.

(4.2.1) leads to:

$$(k^3 + l^2 + m^2)v''' - \omega v' + \delta k v^2 v' = 0. \quad (4.2.2)$$

Integrating Eq. (4.2.2) with respect to ξ taking the constant of integration to be zero, yields

$$k(k^2 + l^2 + m^2)v'' - \omega v + \delta k \frac{v^3}{3} = 0, \quad (4.2.3)$$

we balanced the highest-order derivative terms and the highest nonlinear terms of Eq. (4.2.3). Then, $N = 1$.

We have the solution form of Eq. (4.2.3) as follows:

$$v(\xi) = \sum_{i=0}^1 a_i F^i(\xi) = a_0 + a_1 F, \quad (4.2.4)$$

where F satisfies Eq. (3.5). Therefore, the following expressions for v'' and v^3 are:

$$\begin{aligned} v'' &= a_1 c^2 F + 3a_1 c d F^2 + 2a_1 d^2 F^3, \\ v^3 &= a_0^3 + 3a_0^2 a_1 F + 3a_0 a_1^2 F^2 + a_1^3 F^3. \end{aligned} \quad (4.2.5)$$

Substituting Eqs. (4.2.4) and (4.2.5) into Eq. (4.2.3) and then equating the coefficient of F^i to zero, where $i = 0, 1, 2, 3$, yields

$$\begin{aligned} F^0(\xi): -3\omega a_0 + \delta k a_0^3 &= 0, \\ F^1(\xi): -\omega a_1 + \delta k a_0^2 a_1 + k(k^2 + l^2 + m^2)a_1 c^2 &= 0, \\ F^2(\xi): \delta k a_0 a_1^2 + 3a_1 c d k(k^2 + l^2 + m^2) &= 0, \\ F^3(\xi): \delta k a_1^3 + 6a_1 d^2 k(k^2 + l^2 + m^2) &= 0. \end{aligned} \quad (4.2.6)$$

Solving this system of algebraic equations, we obtain

$$\begin{aligned} a_0 &= \pm \sqrt{\frac{3\omega}{\delta k}}, \quad a_1 = \pm \frac{2d}{c} \sqrt{\frac{3\omega}{\delta k}} \quad \text{and} \\ \omega &= -\frac{k(k^2 + l^2 + m^2)c^2}{2}, \end{aligned} \quad (4.2.7)$$

where $\frac{3\omega}{\delta k} > 0$.

The following two exact solutions are obtained by substituting Eq. (4.2.7) into Eq. (4.2.4). Then, we utilize the general solutions to the Bernoulli equations (3.6) and (3.7). We get four exact solutions of Eq. (4.2.1) from the exponential term.

Case I: if $c > 0$ and $d < 0$, we get

$$v_{1,2}(x, y, z, t) = \pm c \sqrt{\frac{-6\beta}{\delta}} \left(\frac{1}{2} + d \left(\frac{e^{c(\xi + \xi_0)}}{1 - d e^{c(\xi + \xi_0)}} \right) \right), \quad (4.2.8)$$

where $\xi = \frac{kx^\alpha}{\Gamma(\alpha+1)} + \frac{ly^\alpha}{\Gamma(\alpha+1)} + \frac{mz^\alpha}{\Gamma(\alpha+1)} - \frac{\omega t^\alpha}{\Gamma(\alpha+1)}$,

$\beta = k^2 + l^2 + m^2$ and ξ_0 is a constant of the integration.

Case II: if $c < 0$ and $d > 0$, we get

$$v_{3,4}(x, y, z, t) = \pm c \sqrt{\frac{-6\beta}{\delta}} \left(\frac{1}{2} - d \left(\frac{e^{c(\xi + \xi_0)}}{1 + de^{c(\xi + \xi_0)}} \right) \right), \quad (4.2.9)$$

$$\text{where } \xi = \frac{kx^\alpha}{\Gamma(\alpha+1)} + \frac{ly^\alpha}{\Gamma(\alpha+1)} + \frac{mz^\alpha}{\Gamma(\alpha+1)} - \frac{\omega t^\alpha}{\Gamma(\alpha+1)},$$

$\beta = k^2 + l^2 + m^2$ and ξ_0 is a constant of the integration.

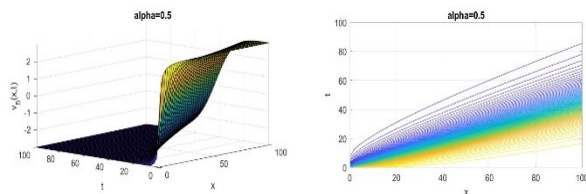
After that, we set $c = \pm 1$ and $d = \mp 1$ into Eqs. (4.2.8) and (4.2.9), and we get the following precise solutions in hyperbolic form:

$$v_{5,6}(x, y, z, t) = \pm \frac{1}{2} \sqrt{\frac{-6\beta}{\delta}} \tanh \left(\frac{\xi + \xi_0}{2} \right), \quad (4.2.10)$$

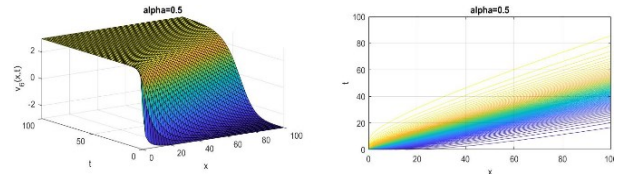
$$\text{where } \xi = \frac{kx^\alpha}{\Gamma(\alpha+1)} + \frac{ly^\alpha}{\Gamma(\alpha+1)} + \frac{mz^\alpha}{\Gamma(\alpha+1)} - \frac{\omega t^\alpha}{\Gamma(\alpha+1)},$$

$\beta = k^2 + l^2 + m^2$ and ξ_0 is a constant of the integration.

Next, we represent the shape of solutions to the (3+1)-dimensional space-time fractional modified KdV-Zakharov-Kuznetsov equation by setting some parameters $\delta = -1$, $k = 1$, $m = 1$, $l = 1$ and $\alpha = 0.5$ in the interval $0 \leq x, t \leq 100$, and as displayed in Figures 3 and 4, it produces a kink wave solution.



Figures 3. kink wave solution of $v_5(x, y, z, t)$ in 3D and contour



Figures 4. kink wave solution of $v_6(x, y, z, t)$ in 3D and contour

5. CONCLUSIONS

In this work, we have investigated exact traveling wave solutions for nonlinear fractional evolution equations, namely, the fractional Landau-Ginzburg-Higgs equation and the (3+1)-dimensional space-time fractional modified KdV-Zakharov-Kuznetsov equation, via the simple equation method, and the solutions are found in exponential and hyperbolic forms. This method allows us to solve nonlinear fractional evolution equations in mathematical physics.

The simple equation method has many advantages: it is straight-forward and concise. Furthermore, this study shows that the proposed method is quite efficient and practically well suited for finding exact solutions to the fractional Landau-Ginzburg-Higgs equation and the (3+1)-dimensional space-time fractional modified KdV-Zakharov-Kuznetsov equation. Therefore, the simple equation method is reliable for constructing exact traveling wave solutions to nonlinear models encountered in a variety of scientific and engineering fields.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude for the reviewers' dedication and commitment in enhancing the overall quality of the manuscript and the ideas that helped further develop this paper.

REFERENCES

- Alabedalhadi, M. (2022). Exact travelling wave solutions for nonlinear system of spatiotemporal fractional quantum mechanics equations. *Alexandria Engineering Journal*, 61(2), 1033–1044.
- Baishya, C., & Rangarajan, R. (2018). A new application of G'/G -expansion method for travelling wave solutions of fractional PDEs. *International Journal of Applied Engineering Research*, 13(11), 9936–9942.
- Çenesiz, Y., Tasbozan, O., & Kurt, A. (2017). Functional Variable Method for conformable fractional modified KdV-ZK equation and Maccari system. *Euro-Tbilisi Mathematical Journal*, 10(1), 117–125.
- Eslami, M., Fathi, V., B., Mirzazadeh, M., & Biswas, A. (2014). Application of first integral method to fractional partial differential equations. *Indian Journal of Physics*, 88, 177–184.
- Guner, O., Bekir, A., & Korkmaz, A. (2017). Tanh-type and sech-type solitons for some space-time fractional PDE models. *The European Physical Journal Plus*, 132, 1–12.
- Islam, M. T., Akbar, M. A., & Azad, M. A. K. (2019). Closed-form travelling wave solutions to the nonlinear space-time fractional coupled Burgers' equation. *Arab. Journal of Basic and Applied Sciences*, 26(1), 1–11.
- Islam, S. R., Arafat, S. Y., & Wang, H. (2023). Abundant closed-form wave solutions to the simplified modified Camassa-Holm equation. *Journal of Ocean Engineering and Science*, 8(3), 238–245.
- Jhangeer, A., Faridi, W. A., Asjad, M. I., & Akgül, A. (2021). Analytical study of soliton solutions for an improved perturbed Schrödinger equation with Kerr law non-linearity in non-linear optics by an expansion algorithm. *Partial Differential Equations in Applied Mathematics*, 4, 100102.
- Kumar, D., Seadawy, A. R., & Joardar, A. K. (2018). Modified Kudryashov method via new exact solutions for some conformable fractional differential equations arising in mathematical biology. *Chinese journal of physics*, 56(1), 75–85.
- Khodadad, F. S., Nazari, F., Eslami, M., & Rezazadeh, H. (2017). Soliton solutions of the conformable fractional Zakharov-Kuznetsov equation with dual-power law nonlinearity. *Optical and Quantum Electronics*, 49, 1–12.
- Li, Y., Meyer, J., Lokki, T., Cuenca, J., Atak, O., & Desmet, W. (2022). Benchmarking of finite-difference time-domain method and fast multipole boundary element method for room acoustics. *Applied Acoustics*, 191, 108662.
- Osman, M. S., Rezazadeh, H., & Eslami, M. (2019). Traveling wave solutions for (3+1) dimensional conformable fractional Zakharov-Kuznetsov equation with power law nonlinearity. *Nonlinear Engineering*, 8(1), 559–567.
- Phoosree, S., & Thadee, W. (2022). Wave effects of the fractional shallow water equation and the fractional optical fiber equation. *Frontiers in Applied Mathematics and Statistics*, 8, 900369.
- Rezazadeh, H., Sabi'u, J., Jena, R. M., & Chakraverty, S. (2020). New optical soliton solutions for Triki-Biswas model by new extended direct algebraic method. *Modern Physics Letters B*, 34(supp01), 2150023.
- Rezazadeh, H., Tariq, K. U., Sabi'u, J., & Bekir, A. (2020). Abundant traveling wave solutions to the resonant nonlinear Schrödinger's equation with variable coefficients. *Modern Physics Letters B*, 34(12), 2050118.
- Sabi'u, J., Jibril, A., & Gadu, A. M. (2019). New exact solution for the (3+1) conformable space-time fractional modified Korteweg-de-Vries equations via Sine-Cosine Method. *Journal of Taibah University for Science*, 13(1), 91–95.
- Sabi'u, J., Shaayesteh, M. T., Taheri, A., Rezazadeh, H., Inc, M., & Akgül, A. (2023). New exact solitary wave solutions of the generalized (3+1)-dimensional nonlinear wave equation in liquid with gas bubbles via extended auxiliary equation method. *Optical and Quantum Electronics*, 55(7), 586.
- Sahoo, S., & Ray, S. S. (2015). Improved fractional sub-equation method for (3+1)-dimensional generalized fractional KdV-Zakharov-Kuznetsov equations. *Computers & Mathematics with Applications*, 70(2), 158–166.
- Sanjun, J., & Chankaew, A. (2022). Wave solutions of the DMBBM equation and the cKG equation using the simple equation method. *Frontiers in Applied Mathematics and Statistics*, 8, 952668.
- Salas, A. H., El-Tantawy, S. A., & Youssef, A. A. A. R. (2020). New solutions for chirped optical solitons related to Kaup-Newell equation: Application to plasma physics. *Optik*, 218, 165203.
- Shah, Z., Dawar, A., Nasir, S., Islam, S., Deebani, W., & Shutaywi, M. (2022). Application of Arrhenius chemical process on unsteady mixed bio-convective flows of third-grade fluids having temperature-dependent in thermo-rheological properties. *Waves in Random and Complex Media*, 1–20.
- Simpson, G. G. (1963). Biology and the Nature of Science: Unification of the sciences can be most meaningfully sought through study of the phenomena of life. *Science*, 139(3550), 81–88.
- Thadee, W., Chankaew, A., & Phoosree, S. (2022). Effects of wave solutions on shallow-water equation, optical-fibre equation and electric-circuit equation. *Maejo International Journal of Science & Technology*, 16(3), 262–274.
- Torvattanabun, M., Simtrakankul, C., & Duangpithak, S. (2019). The (3+1)-dimensional Fractional Modified Kdv-Zakharov-Kuznetsov Equation By Using The Improved Generalized Tanh-Coth Method. *Journal of Mathematical Sciences: Advances and Applications*, 56, 53–65.
- Zeng, J., & Liu, H. (2022). An approximate explicit analytical solution for the frictionless swash hydrodynamics with an improved seaward boundary condition. *Coastal Engineering*, 174, 104127.

การเปรียบเทียบวิธีการประมาณค่าของช่วงความเชื่อมั่นเชิงเส้นกำกับ และช่วงความเชื่อมั่นที่แท้จริงสำหรับสัดส่วนของประชากร ที่มีการแจกแจงทวินามและการประยุกต์

Comparison of Methods on Asymptotic and Exact Confidence Intervals for a Binomial Proportion with Application

ชนกานต์ สังข์บุญชู*, ปิยพล ไพจิตร

Chanakan Sungboonchoo*, Piyaphon Paichit

ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร ประเทศไทย

Department of Statistics, Faculty of Science, Silpakorn University, Thailand

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าช่วงความเชื่อมั่นสำหรับสัดส่วนของประชากรที่มีการแจกแจงทวินามทั้ง 2 ประเภท คือ ช่วงความเชื่อมั่นที่แท้จริง และช่วงความเชื่อมั่นเชิงเส้นกำกับ โดยศึกษาเปรียบเทียบวิธีการประมาณค่าช่วงความเชื่อมั่น 5 วิธี ได้แก่ วิธีวัลด์ (Wald) วิธี Wilson score วิธี Agresti-Coull วิธี New และวิธี generalized Agresti-Coull จำลองข้อมูลโดยใช้วิธีมอนติคาร์โลให้มีการแจกแจงทวินามด้วยพารามิเตอร์ ขนาดตัวอย่าง (n) มีค่าเท่ากับ 10, 30, 50 และ 100 กำหนดสัดส่วน (p) มีค่าเท่ากับ 0.1, 0.3, 0.5, 0.7 และ 0.9 และกำหนดระดับความเชื่อมั่นเท่ากับ 95% โดยทำการทดลองซ้ำ 10,000 ครั้งในแต่ละสถานการณ์ ซึ่งเกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพของช่วงความเชื่อมั่น คือ ค่าความน่าจะเป็นคัมรวม และค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่น ผลการวิจัยสรุปได้ดังนี้ ช่วงความเชื่อมั่นวิธี Wilson score มีประสิทธิภาพดีที่สุดหลายสถานการณ์ที่ศึกษา ซึ่งช่วงความเชื่อมั่นนี้เหมาะกับสถานการณ์ที่ $n=10$ เมื่อ $p=0.5$ และ $n=30$ เมื่อ p มีค่าเข้าใกล้ 0 หรือ 1 และ $n=50$ สำหรับทุกค่าของ p ยกเว้น $p=0.5$ ช่วงความเชื่อมั่นวิธี Agresti-Coull มีประสิทธิภาพเป็นอันดับที่สอง ซึ่งเหมาะกับ $n=10$ เมื่อ $p=0.3$ และ 0.7 และ $n=30$ เมื่อ $p=0.3-0.7$ ส่วนช่วงความเชื่อมั่นวิธี New มีประสิทธิภาพเป็นอันดับที่สาม ซึ่งเหมาะกับ $n=10$ และ 100 เมื่อ p มีค่าเข้าใกล้ 0 หรือ 1 ช่วงความเชื่อมั่นวิธี generalized Agresti-Coull มีประสิทธิภาพเป็นอันดับที่สี่ ซึ่งเหมาะกับ $n=100$ เมื่อ $p=0.3$ และ 0.7 และช่วงความเชื่อมั่นวิธีวัลด์ (Wald) มีประสิทธิภาพที่ไม่ค่อยดี นอกจากนั้นช่วงความเชื่อมั่นทั้ง 5 วิธีสามารถนำไปประยุกต์ใช้กับข้อมูลจริงในสถานการณ์ต่าง ๆ

คำสำคัญ: ช่วงความเชื่อมั่น, พารามิเตอร์ค่าสัดส่วนทวินาม, ความน่าจะเป็นคัมรวม, การจำลองมอนติคาร์โล

ABSTRACT

The objective of this research was to compare efficiency of the methods for estimating two types of confidence interval for a binomial proportion, which are exact confidence interval and asymptotic confidence interval by studying comparison upon 5 methods for estimating confidence interval. These 5 methods are Wald method, Wilson Score method, Agresti-Coull method, New method, and generalized Agresti-Coull method. The data was simulated by using Monte Carlo method, utilizing binomial distribution with parameters n and p that determine sizes of the sample (n) equal to 10, 30, 50, and 100. Meanwhile, the values of proportion (p) were determined as 0.1, 0.3, 0.5, 0.7, and 0.9 with confidence level 95% by repeating experiment 10,000 times in each certain circumstance. Moreover, the criteria used to compare efficiency of confidence intervals was coverage probability as well as average length. According to research findings, it can be summarized as follows. Confidence interval of Wilson Score method showed the best efficiency in diverse circumstances in case study. Besides, this confidence interval was appropriate for the circumstance at $n=10$ when $p=0.5$ and $n=30$ when p was close to 0 or 1, and $n=50$ for every value of p , except $p=0.5$. Furthermore, confidence interval of Agresti-Coull method had the most second efficiency, which was appropriate for $n=10$ when $p=0.3$ and 0.7 , and $n=30$ when $p=0.3-0.7$, whereas confidence interval of New method had the third most efficiency, which was appropriate for $n=10$ and 100 when value of p was close to 0 or 1. At the same time, confidence interval of generalized Agresti-Coull method had the fourth most efficiency, which was appropriate for $n=100$ when $p=0.3$ and 0.7 .

However, confidence interval of Wald method possess quite low precision. In addition, confidence interval of these 5 methods can be applicable to the real data in any circumstances.

KEYWORDS: Confidence interval, Binomial proportion parameter, Coverage probability, Monte Carlo simulation

*Corresponding Author: sungboonchoo_c@silpakorn.edu

Received: 12/01/2023; Revised: 19/06/2023; Accepted: 14/07/2023

1. บทนำ

การอนุมานเชิงสถิติ (statistical inference) คือ กระบวนการหาข้อสรุปเกี่ยวกับลักษณะของประชากร โดยใช้ข้อมูลจากตัวอย่างที่ได้มาโดยวิธีการเชิงสุ่มจากประชากรนั้น ในการอนุมานเชิงสถิติแบ่งออกเป็น 2 ลักษณะ คือ การทดสอบสมมติฐานและการประมาณค่า ซึ่งการประมาณค่ามี 2 แบบ คือ การประมาณค่าแบบจุด (point estimation) เป็นการประมาณค่าพารามิเตอร์ด้วยค่าสถิติเพียงค่าเดียว ซึ่งตัวประมาณแบบจุดที่ได้มีโอกาสคลาดเคลื่อนไปจากค่าพารามิเตอร์ได้มาก และการประมาณค่าแบบช่วง (interval estimation) เป็นการประมาณค่าพารามิเตอร์ด้วยช่วง ๆ หนึ่งหรือค่า 2 ค่าที่เป็นขีดจำกัดล่าง (lower limit) และขีดจำกัดบน (upper limit) ของช่วงความเชื่อมั่น ค่า 2 ค่าดังกล่าวสร้างจากตัวประมาณแบบจุดและการแจกแจงภาวะน่าจะเป็นของตัวประมาณนั้น โดยคาดว่าพารามิเตอร์มีแนวโน้มจะอยู่ในช่วงนี้หรืออยู่ระหว่างค่า 2 ค่านี้อย่างมีความเชื่อมั่นระดับหนึ่ง ซึ่งจะเห็นได้ว่าการประมาณค่าแบบช่วงสามารถบอกขอบเขตของค่าพารามิเตอร์ที่สนใจได้แม่นยำกว่าการประมาณค่าแบบจุด

การประมาณค่าสัดส่วนของประชากรที่มีการแจกแจงทวินามนั้นมีการศึกษากันอย่างแพร่หลาย ซึ่งพารามิเตอร์สำหรับสัดส่วน (p) ของการแจกแจงทวินามนี้แทนความน่าจะเป็นของการเกิดสิ่งที่สนใจ โดยให้ตัวแปรสุ่ม X แทน จำนวนครั้งของการเกิดสิ่งที่สนใจจากการทดลองสุ่มที่กระทำซ้ำ ๆ กัน n ครั้ง โดยการทดลองแต่ละครั้งเป็นอิสระต่อกัน ซึ่งในการกระทำแต่ละครั้งจะมีผลลัพธ์ที่เกิดขึ้นเป็นไปได้ 2 ลักษณะ คือ ผลสำเร็จ (success) และผลไม่สำเร็จหรือผลล้มเหลว (failure) นั่นคือ $X \sim B(n, p)$ และมีฟังก์ชันมวลความน่าจะเป็น $f(x) = \binom{n}{x} p^x (1-p)^{n-x}, x=0,1,2,\dots,n$ การประมาณค่าสัดส่วนนั้นมีความสำคัญทั้งในด้านการแพทย์ เช่น Komoltri (2012) ได้ศึกษาการประมาณค่าสัดส่วนโดยผู้วิจัย

ต้องการทราบค่าความไว (sensitivity) และความจำเพาะ (specificity) ของ ultrasound, mammogram เทียบกับผลการตรวจทางพยาธิวิทยาของชิ้นเนื้อ (gold standard) ในการวินิจฉัยมะเร็งเต้านม ค่าต่าง ๆ เหล่านี้จะรายงานเป็นสัดส่วนของความไวหรือช่วงความเชื่อมั่นสำหรับสัดส่วนของค่าความไวในประชากร และในด้านสิ่งแวดล้อม เช่น Muenmee & Bootdee (2021) ได้ศึกษาความเข้มข้นของฝุ่น $PM_{2.5}$ ที่ปลดปล่อยมาจากพื้นที่เขตอุตสาหกรรมในอำเภอปลวกแดง จังหวัดระยอง และนำมาทำการประเมินความเสี่ยงต่อสุขภาพของมนุษย์ด้วยการคำนวณจากค่าสัดส่วนความเสี่ยง (Hazard Quotient; HQ) เป็นต้น

ข้อมูลทวิภาค (binary data) จะพบได้บ่อยในการวิเคราะห์ทางสถิติ ซึ่งจะเป็นข้อมูลของตัวแปรที่มีค่าที่เป็นไปได้เพียง 2 ค่า เช่น สำเร็จ-ไม่สำเร็จ ใช่-ไม่ใช่ เป็นต้น และช่วงความเชื่อมั่นที่นำมาใช้สำหรับการประมาณค่าสัดส่วนของข้อมูลทวิภาคหรือสัดส่วนของประชากรที่มีการแจกแจงทวินาม คือ ช่วงความเชื่อมั่นของวิธี Wald (Wald, 1943) ซึ่งเป็นวิธีดั้งเดิมและมีวิธีการคำนวณที่ง่ายจึงนิยมใช้กันอย่างแพร่หลาย จากการศึกษางานวิจัยส่วนใหญ่ที่เกี่ยวข้องกับการประมาณค่าสัดส่วนโดยใช้ช่วงความเชื่อมั่นของวิธี Wald จะพบปัญหาที่สำคัญ คือ Brown et al. (2001) พบว่าช่วงความเชื่อมั่นของวิธี Wald มีความกว้างของค่าความน่าจะเป็นคลุมรวม (coverage probability) สูงในกรณีที่ค่าพารามิเตอร์สัดส่วนมีค่าเข้าใกล้ 0 หรือ 1 Agresti & Coull, 1998; Brown et al. (2002) ได้กล่าวถึงช่วงความเชื่อมั่นของวิธี Wald มีประสิทธิภาพที่ไม่ดีนักเนื่องจากให้ค่าความน่าจะเป็นคลุมรวมต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด จึงมีงานวิจัยที่เสนอวิธีการปรับปรุงช่วงความเชื่อมั่นสำหรับการประมาณค่าสัดส่วนของประชากรที่มีการแจกแจงทวินามหลากหลายวิธี ซึ่งจะมีช่วงความเชื่อมั่นของวิธี Wilson score (Wilson, 1927) เป็นอีกทางเลือกหนึ่ง เพราะเป็นวิธีที่มีสูตรการคำนวณช่วงความเชื่อมั่นที่ง่ายและมีประสิทธิภาพที่ดีกว่าช่วงความเชื่อมั่นของวิธี Wald แต่จะมี

ประสิทธิภาพที่ไม่ดีนักสำหรับในกรณีที่ค่าพารามิเตอร์สัดส่วนมีค่าเข้าใกล้ 0 หรือ 1 โดยทั่วไปช่วงความเชื่อมั่นสำหรับสัดส่วนของประชากรที่มีการแจกแจงทวินามจะแบ่งออกเป็น 2 ประเภท ได้แก่ 1) ช่วงความเชื่อมั่นที่แท้จริง (exact confidence interval) ซึ่งโดยทั่วไปช่วงความเชื่อมั่นประเภทนี้จะไม่มีรูปแบบของสูตรช่วงความเชื่อมั่นที่ชัดเจน สำหรับการสร้างช่วงความเชื่อมั่นที่แท้จริงจะนิยมใช้วิธีแก้ปัญหาคณิตศาสตร์เชิงตัวเลข (numerical solution) เช่น ช่วงความเชื่อมั่นของวิธี Clopper-Pearson (Clopper & Pearson, 1934) และ 2) ช่วงความเชื่อมั่นเชิงเส้นกำกับ (asymptotic confidence interval) จะสร้างขึ้นโดยใช้ทฤษฎีเชิงเส้นกำกับ (asymptotic theory) ซึ่งลักษณะทั่วไปของช่วงความเชื่อมั่นประเภทนี้จะไม่มีรูปแบบของสูตรการคำนวณช่วงความเชื่อมั่นที่ชัดเจน เช่น ช่วงความเชื่อมั่นของวิธี Wald) วิธี Wilson score และวิธี Agresti-Coull เป็นต้น Cao et al. (2022) ได้ทำการปรับปรุงช่วงความเชื่อมั่นจากวิธี Wilson score เพื่อแก้ไขปัญหาในกรณีที่ค่าพารามิเตอร์สัดส่วนมีค่าเข้าใกล้ 0 หรือ 1 จึงนำเสนอช่วงความเชื่อมั่นสำหรับการประมาณค่าสัดส่วนของประชากรที่มีการแจกแจงทวินามแบบใหม่ คือ วิธี New จะมีลักษณะเป็นช่วงความเชื่อมั่นที่แท้จริง นั่นคือวิธี New จะเป็นวิธีการคำนวณที่ซับซ้อนไม่มีสูตรการคำนวณช่วงความเชื่อมั่นที่ชัดเจน จึงต้องเขียนโปรแกรมคอมพิวเตอร์ในการคำนวณหาช่วงความเชื่อมั่น ซึ่งปรับปรุงโดยใช้วิธี Cornish-Fisher expansion (Fisher & Cornish, 1960) ในการประมาณค่าควอนไทล์ของการแจกแจงทวินามและสร้างช่วงความเชื่อมั่นโดยใช้วิธีสกอร์ (score method) และทำการศึกษเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าช่วงความเชื่อมั่นสำหรับการประมาณค่าสัดส่วนของประชากรที่มีการแจกแจงทวินามทั้ง 7 วิธี ได้แก่ 1) วิธี New 2) วิธี Wilson score 3) วิธี Agresti-Coull 4) วิธี Jeffreys 5) วิธี Mid-P 6) วิธี modified midpoint Wilson score และ 7) วิธี randomized plausible จากผลการศึกษาพบว่า เมื่อค่าพารามิเตอร์สัดส่วนมีค่าเข้าใกล้ 0 หรือ 1 (<0.17 หรือ >0.83) ช่วงความเชื่อมั่นของวิธี New จะมีประสิทธิภาพดีที่สุด และเมื่อค่าพารามิเตอร์สัดส่วนมีค่าอยู่ในช่วง 0.17-0.83 ช่วงความเชื่อมั่นของวิธี Wilson score และวิธี modified midpoint Wilson score จะมีประสิทธิภาพดีที่สุด อย่างไรก็ตามสำหรับในกรณีค่าพารามิเตอร์สัดส่วนมีค่าอยู่ในช่วง

0.17-0.83 วิธี New จะมีประสิทธิภาพดี แต่แนะนำให้ใช้วิธี Wilson score เพราะมีสูตรการคำนวณที่ง่ายกว่าสูตรของวิธี modified midpoint Wilson score และ วิธี New Kim & Jang (2022) ได้ศึกษาประสิทธิภาพของช่วงความเชื่อมั่นของวิธี Wilson score และวิธี Agresti-Coull พบว่าช่วงความเชื่อมั่นทั้ง 2 วิธีนี้มีความเอนเอียงแบบเป็นระบบ (systematic bias) และมีความกวัดแกว่งของค่าความน่าจะเป็นกลุ่มรวม จึงได้ปรับปรุงช่วงความเชื่อมั่นจากวิธี Agresti-Coull โดยนำเสนอช่วงความเชื่อมั่นสำหรับการประมาณค่าสัดส่วนของประชากรที่มีการแจกแจงทวินาม คือ วิธี generalized Agresti-Coull โดยทำการปรับลดความเอนเอียงแบบเป็นระบบ ด้วยวิธี saddlepoint (saddlepoint approximation) ซึ่งช่วงความเชื่อมั่นที่นำเสนอนี้มีลักษณะเป็นช่วงความเชื่อมั่นเชิงเส้นกำกับ นั่นคือมีสูตรการคำนวณช่วงความเชื่อมั่นที่ชัดเจน และทำการเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าช่วงความเชื่อมั่นสำหรับการประมาณค่าสัดส่วนของประชากรที่มีการแจกแจงทวินามทั้ง 4 วิธี ได้แก่ 1) วิธี generalized Agresti-Coull 2) วิธี Wald 3) วิธี Wilson score และ 4) วิธี Agresti-Coull จากการศึกษาพบว่า ช่วงความเชื่อมั่นของวิธี Wilson score มีประสิทธิภาพดีที่สุดสำหรับค่าความน่าจะเป็นกลุ่มรวม ถึงแม้ว่าจะมีค่าความกว้างของช่วงความเชื่อมั่นที่กว้างมากในทางกลับกันช่วงความเชื่อมั่นของวิธี Agresti-Coull จะให้ค่าความน่าจะเป็นกลุ่มรวมที่ค่อนข้างคงที่กว่าของวิธี Wilson score แต่จะมีความกว้างของช่วงความเชื่อมั่นที่มากกว่าวิธี Wilson score ส่วนวิธี generalized Agresti-Coull จะให้ค่าความน่าจะเป็นกลุ่มรวมที่ค่อนข้างคงที่กว่าของวิธี Wilson score และมีความกว้างของช่วงความเชื่อมั่นที่น้อยกว่าวิธี Agresti-Coull สำหรับทุกขนาดตัวอย่างที่ทำการศึกษา

จากการศึกษาวิธีการปรับปรุงช่วงความเชื่อมั่นแบบต่างๆ ในงานวิจัยนี้จึงสนใจศึกษาเปรียบเทียบวิธีการประมาณค่าของช่วงความเชื่อมั่นทั้ง 2 ประเภท คือ ช่วงความเชื่อมั่นเชิงเส้นกำกับและช่วงความเชื่อมั่นที่แท้จริงสำหรับสัดส่วนของประชากรที่มีการแจกแจงทวินาม โดยทำการศึกษเปรียบเทียบ 5 วิธี ได้แก่ 1) วิธี Wald 2) วิธี Wilson score 3) วิธี Agresti-Coull 4) วิธี New และ 5) วิธี generalized Agresti-Coull โดยเกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพของช่วงความเชื่อมั่นวิธีต่างๆ

มี 2 เกณฑ์ คือ ค่าความน่าจะเป็นค้ำรวม (coverage probability) และค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่น (average length) ซึ่งวิธีที่มีประสิทธิภาพดีจะต้องให้ค่าความน่าจะเป็นค้ำรวมเป็นไปตามค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดและมีค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นน้อยที่สุด และการประยุกต์ใช้โดยการแนะนำช่วงความเชื่อมั่นสำหรับสัดส่วนของประชากรที่มีการแจกแจงทวินามให้เหมาะสมกับสถานการณ์ต่าง ๆ ที่ศึกษา

2. วิธีการทดลอง

การเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าช่วงความเชื่อมั่นสำหรับสัดส่วนของประชากรที่มีการแจกแจงทวินาม 5 วิธี ได้แก่ 1) วิธีไวล์ด (Wald) 2) วิธี Wilson score 3) วิธี Agresti-Coull 4) วิธี New และ 5) วิธี generalized Agresti-Coull มีขั้นตอนการวิจัยดังต่อไปนี้

1. กำหนดขนาดตัวอย่าง (n) เท่ากับ 10, 30, 50 และ 100 และกำหนดค่าพารามิเตอร์สำหรับสัดส่วนเท่ากับ 0.1, 0.3, 0.5, 0.7 และ 0.9
2. กำหนดระดับความเชื่อมั่น $(1-\alpha)100\%$ เท่ากับ 95%
3. ใช้โปรแกรม R เวอร์ชัน 4.2.2 ในการจำลองข้อมูลที่มีการแจกแจงทวินามที่มีค่าพารามิเตอร์ (n, p) โดยใช้การจำลองมอนติคาร์โล (Monte Carlo simulation) ซึ่งทำการทดลองซ้ำ 10,000 ครั้งในแต่ละสถานการณ์
4. คำนวณช่วงความเชื่อมั่นสำหรับสัดส่วนของประชากรที่มีการแจกแจงทวินามที่ศึกษาในงานวิจัยนี้ทั้งหมด 5 วิธี ดังนี้

4.1 วิธีไวล์ด (Wald)

Wald (1943) ได้เสนอวิธีการสร้างช่วงความเชื่อมั่นโดยใช้ทฤษฎีบทขีดจำกัดส่วนกลาง (central limit theorem) ภายใต้การแจกแจงปกติ โดยใช้ตัวประมาณค่าพารามิเตอร์สำหรับสัดส่วน คือ $\hat{p} = \frac{X}{n}$ ซึ่งเป็นตัวประมาณภาวน่าจะเป็นสูงสุด (maximum likelihood estimator $\hat{p}$) โดยที่ จะมีค่าเฉลี่ยเท่ากับ p และมีความแปรปรวนเท่ากับ $\frac{p(1-p)}{n}$ และจากทฤษฎีบทขีดจำกัดส่วนกลาง ถ้าขนาดตัวอย่าง (n) มีขนาดใหญ่ ($n \rightarrow \infty$) จะได้ว่า $Z = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \sim N(0,1)$ และสามารถประมาณค่าช่วงความเชื่อมั่นได้โดย

$$P\left[-Z_{\alpha/2} < \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} < Z_{\alpha/2}\right] = 1 - \alpha \quad \text{กำหนดให้}$$

ขีดจำกัดล่างและขีดจำกัดบนของช่วงความเชื่อมั่นไวล์ดเขียนแทนด้วยสัญลักษณ์ L_W และ U_W ตามลำดับ ดังนั้นช่วงความเชื่อมั่น $(1-\alpha)100\%$ สำหรับสัดส่วน (p) ของประชากรที่มีการแจกแจงทวินามของวิธีไวล์ด เป็นดังนี้

$$L_W \text{ คือ } \hat{p} - Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$U_W \text{ คือ } \hat{p} + Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

เมื่อ $\hat{p}$ คือ สัดส่วนของการเกิดสิ่งที่สนใจในตัวอย่าง ซึ่ง $\hat{p} = \frac{X}{n}$; X แทน จำนวนครั้งของการเกิดสิ่งที่สนใจในตัวอย่างสุ่มขนาด n และ $Z_{\alpha/2}$ คือ ค่าควอนไทล์ที่

$1-\alpha/2$ ของการแจกแจงปกติมาตรฐาน

4.2 วิธี Wilson score

Wilson (1927) ได้พัฒนาช่วงความเชื่อมั่นโดยใช้แนวคิดของวิธีสกอร์ (score method) โดยสร้างช่วงความเชื่อมั่นภายใต้การทดสอบสมมติฐาน $H_0: p = p_0$ เทียบกับ $H_1: p \neq p_0$ ที่ระดับนัยสำคัญ α และจะยอมรับสมมติฐานหลัก H_0 เมื่อ

$$\frac{|\hat{p} - p_0|}{\sqrt{\frac{p_0(1-p_0)}{n}}} < Z_{\alpha/2} \quad (1)$$

จะสามารถหาค่าพารามิเตอร์ p_0 โดยกำหนดให้ $p = p_0$ ได้จากการจัดสมการที่ (1) ให้อยู่ในรูปของสมการกำลังสอง (quadratic equation) แล้วแก้สมการหาค่าพารามิเตอร์ p_0 กำหนดให้ขีดจำกัดล่างและขีดจำกัดบนของช่วงความเชื่อมั่นวิธี Wilson score เขียนแทนด้วยสัญลักษณ์ L_{WS} และ U_{WS} ตามลำดับ ดังนั้นช่วงความเชื่อมั่น $(1-\alpha)100\%$ สำหรับสัดส่วน (p) ของประชากรที่มีการแจกแจงทวินามของวิธี Wilson score เป็นดังนี้

$$L_{WS} \text{ คือ } \frac{\left(\hat{p} + \frac{Z_{\alpha/2}^2}{2n}\right) - Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{Z_{\alpha/2}^2}{4n^2}}}{1 + \frac{Z_{\alpha/2}^2}{n}}$$

$$U_{WS} \text{ คือ } \frac{\left(\hat{p} + \frac{Z_{\alpha/2}^2}{2n}\right) + Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{Z_{\alpha/2}^2}{4n^2}}}{1 + \frac{Z_{\alpha/2}^2}{n}}$$

เมื่อ $\hat{p}$ คือ สัดส่วนของการเกิดสิ่งที่สนใจในตัวอย่าง
ซึ่ง $\hat{p} = \frac{X}{n}$; X แทน จำนวนครั้งของการเกิดสิ่งที่สนใจ
ในตัวอย่างสุ่มขนาด n และ $Z_{\alpha/2}$ คือ ค่าควอนไทล์ที่
 $1-\alpha/2$ ของการแจกแจงปกติมาตรฐาน

4.3 วิธี Agresti-Coull

Agresti & Coull (1998) ได้ทำการปรับปรุงช่วงความเชื่อมั่นจากวิธีวิลด์โดยใช้หลักการเพิ่มค่า สังเกต 4 ค่า นั่นคือ การเพิ่มผลลัพธ์ของค่าสังเกตที่สนใจและไม่สนใจอย่างละ 2 ค่า (adding two successes and two failures) ในจำนวนครั้งของการทดลอง (n) ทำให้จำนวนครั้งของผลลัพธ์ที่เป็นไปได้ทั้งหมดเพิ่มขึ้นอีก 4 ครั้ง และกำหนดให้ $Z_{\alpha/2} \approx 2$ ซึ่งช่วงความเชื่อมั่นของวิธี Agresti-Coull กับช่วงความเชื่อมั่นของวิธีวิลด์จะแตกต่างกันที่ค่าประมาณของพารามิเตอร์สำหรับสัดส่วน จะเห็นได้ว่า
 $\tilde{p} = \frac{X+2}{n+4}$ กำหนดให้ขีดจำกัดล่างและขีดจำกัดบนของช่วงความเชื่อมั่นวิธี Agresti-Coull เขียนแทนด้วยสัญลักษณ์ L_{AC} และ U_{AC} ตามลำดับ ดังนั้นช่วงความเชื่อมั่น $(1-\alpha)100\%$ สำหรับสัดส่วน (p) ของประชากรที่มีการแจกแจงทวินามของวิธี Agresti-Coull เป็นดังนี้

$$L_{AC} \text{ คือ } \tilde{p} - Z_{\alpha/2} \sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}}$$

$$U_{AC} \text{ คือ } \tilde{p} + Z_{\alpha/2} \sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}}$$

เมื่อ $\tilde{p}$ คือ สัดส่วนของการเกิดสิ่งที่สนใจในตัวอย่าง ซึ่ง
 $\tilde{p} = \frac{X+2}{n+4}$; X แทน จำนวนครั้งของการเกิดสิ่งที่สนใจ
ในตัวอย่างสุ่มขนาด n , $\tilde{n} = n+4$ และ $Z_{\alpha/2}$ คือ
ค่าควอนไทล์ที่ $1-\alpha/2$ ของการแจกแจงปกติมาตรฐาน

4.4 วิธี New

Cao et al. (2022) ได้พัฒนาช่วงความเชื่อมั่นจากวิธี Wilson score จึงนำเสนอช่วงความเชื่อมั่นสำหรับการประมาณค่าสัดส่วนของประชากรที่มีการแจกแจงทวินามคือ วิธี New ซึ่งปรับปรุงโดยใช้วิธี Cornish-Fisher expansion (Fisher & Cornish, 1960) ในการประมาณค่าควอนไทล์ของการแจกแจงทวินาม โดยจะใช้คิวมุลแลนต์ (cumulant) ที่ 3 (κ_3) และ 4 (κ_4) ของการแจกแจงทวินามในการประมาณค่าควอนไทล์ที่ q (q -quantile (λ_q)) นั่นคือ

$$\lambda_q \approx z_q + \frac{z_q^2 - 1}{6} \kappa_3 + \frac{z_q^3 - 3z_q}{24} \kappa_4 - \frac{2z_q^3 - 5z_q}{36} \kappa_3^2$$

$$\kappa_3 = (1-2\pi) / \sqrt{n\pi(1-\pi)}$$

$$\kappa_4 = (1-6\pi(1-\pi)) / \sqrt{n\pi(1-\pi)}$$

และสร้างช่วงความเชื่อมั่นโดยใช้วิธีสโคร์ (score method) กำหนดให้ขีดจำกัดล่างและขีดจำกัดบนของช่วงความเชื่อมั่นวิธี New เขียนแทนด้วยสัญลักษณ์ L_N และ U_N ตามลำดับ ดังนั้นช่วงความเชื่อมั่น $(1-\alpha)100\%$ สำหรับสัดส่วน (π) ของประชากรที่มีการแจกแจงทวินามของวิธี New เป็นดังนี้

$$L_N \text{ คือ } \frac{p-\pi}{\sqrt{\pi(1-\pi)/n}} = \lambda_{1-\alpha/2}$$

$$U_N \text{ คือ } \frac{p-\pi}{\sqrt{\pi(1-\pi)/n}} = \lambda_{\alpha/2}$$

ช่วงความเชื่อมั่นของวิธี New จะมีวิธีการคำนวณที่ซับซ้อนไม่มีสูตรการคำนวณช่วงความเชื่อมั่นที่ชัดเจนจึงสามารถใช้โค้ดโปรแกรม R จากงานวิจัยของ Cao et al. (2022) ในการคำนวณหาช่วงความเชื่อมั่น

4.5 วิธี generalized Agresti-Coull

Kim & Jang (2022) ได้นำเสนอช่วงความเชื่อมั่นวิธี generalized Agresti-Coull ซึ่งสร้างโดยใช้วิธีการประมาณ saddlepoint (saddlepoint approximation) และมีพารามิเตอร์ 3 ค่าซึ่งเป็นองค์ประกอบในการสร้างช่วงความเชื่อมั่นเพื่อปรับลดค่าความเอนเอียง (bias) ระหว่างค่าความน่าจะเป็นคัมรวม (coverage probability) และระดับความเชื่อมั่น (confidence level) ซึ่งพารามิเตอร์ทั้ง 3 ค่านี้ ได้แก่ a, c และ d การปรับค่าความเอนเอียงของพารามิเตอร์จะขึ้นอยู่กับขนาดตัวอย่าง (n) โดยที่พารามิเตอร์ a จะควบคุมตำแหน่งของค่าประมาณแบบจุดของช่วงความเชื่อมั่น ส่วนพารามิเตอร์ c และ d จะควบคุมค่าขีดจำกัดบนและขีดจำกัดล่างของช่วงความเชื่อมั่นของค่าประมาณแบบจุดแต่ละค่า กำหนดให้ขีดจำกัดล่างและขีดจำกัดบนของช่วงความเชื่อมั่นวิธี generalized Agresti-Coull เขียนแทนด้วยสัญลักษณ์ L_{gAC} และ U_{gAC} ตามลำดับ ดังนั้นช่วงความเชื่อมั่น $(1-\alpha)100\%$ สำหรับสัดส่วน (p) ของประชากรที่มีการแจกแจงทวินามของวิธี generalized Agresti-Coull เป็นดังนี้

$$L_{gAC} \text{ คือ } \tilde{p}_a - Z_{\alpha/2} \sqrt{\frac{\tilde{p}_c(1-\tilde{p}_c)}{n+d}}$$

$$U_{gAC} \text{ คือ } \tilde{p}_a + Z_{\alpha/2} \sqrt{\frac{\tilde{p}_c(1-\tilde{p}_c)}{n+d}}$$

เมื่อ $\tilde{p}_a$ และ $\tilde{p}_c$ คือ สัดส่วนของการเกิดสิ่งที่สนใจในตัวอย่าง ซึ่ง $\tilde{p}_a = \frac{X+a}{n+2a}$ และ $\tilde{p}_c = \frac{X+c}{n+2c}$; X แทนจำนวนครั้งของการเกิดสิ่งที่สนใจในตัวอย่างสุ่มขนาด n และ $Z_{\alpha/2}$ คือ ค่าควอนไทล์ที่ $1-\alpha/2$ ของการแจกแจงปกติมาตรฐาน

สำหรับช่วงความเชื่อมั่น 95% สำหรับสัดส่วน (p) ของประชากรที่มีการแจกแจงทวินาม ค่าพารามิเตอร์

$$a = 2.1 - \frac{1}{\sqrt{n}}, c = 1.7 - \frac{2}{\sqrt{n}}, d = 4 - \frac{1}{\sqrt{n}}$$

5. คำนวณค่าประมาณความน่าจะเป็นคัมรวม ($\hat{c}$) สำหรับช่วงความเชื่อมั่น 95% สำหรับสัดส่วนของประชากรที่มีการแจกแจงทวินามทั้ง 5 วิธี สามารถคำนวณได้ดังนี้

$$\text{ค่าประมาณความน่าจะเป็นคัมรวม} = \frac{\text{จำนวนครั้งทั้งหมดที่ช่วงความเชื่อมั่นครอบคลุมค่าพารามิเตอร์ } p}{M}$$

เมื่อ M คือ จำนวนรอบในการทำซ้ำ ในที่นี้ $M=10,000$ รอบ

6. เปรียบเทียบค่าประมาณความน่าจะเป็นคัมรวม ($\hat{c}$) ที่ได้สำหรับช่วงความเชื่อมั่นทั้ง 5 วิธี ว่ามีค่าเท่ากับค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด (c_0) อย่างมีนัยสำคัญหรือไม่ ทำได้โดยการทดสอบสมมติฐานทางสถิติดังนี้

$H_0: c \geq c_0$ (ค่าความน่าจะเป็นคัมรวมมีค่าไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด)

$H_1: c < c_0$ (ค่าความน่าจะเป็นคัมรวมมีค่าต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด)

$$\text{ตัวสถิติทดสอบ คือ } Z = \frac{\hat{c} - c_0}{\sqrt{\frac{c_0(1-c_0)}{M}}}$$

เมื่อ M คือ จำนวนรอบในการทำซ้ำเท่ากับ 10,000 รอบ

$$\text{จะปฏิเสธสมมติฐาน } H_0 \text{ เมื่อ } \frac{\hat{c} - c_0}{\sqrt{\frac{c_0(1-c_0)}{M}}} < -Z_\alpha$$

$$\text{หรือ } \hat{c} < c_0 - Z_\alpha \sqrt{\frac{c_0(1-c_0)}{M}}$$

เมื่อกำหนดระดับนัยสำคัญของการทดสอบ (α) เท่ากับ 0.05 และค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด (c_0) คือ 0.95 จะได้ว่า

$$\hat{c} < 0.95 - 1.645 \sqrt{\frac{0.95(1-0.95)}{10,000}} \text{ หรือ } \hat{c} < 0.946$$

ดังนั้น วิธีการประมาณค่าช่วงความเชื่อมั่นที่ให้ค่าประมาณความน่าจะเป็นคัมรวม ($\hat{c}$) ไม่ต่ำกว่า

ค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด ($c_0 = 0.95$) เมื่อค่าประมาณความน่าจะเป็นคัมรวมมีค่าไม่ต่ำกว่า 0.946 ($\hat{c} \geq 0.946$)

ถ้าช่วงความเชื่อมั่นใดให้ค่าประมาณความน่าจะเป็นคัมรวมไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด แสดงว่าช่วงความเชื่อมั่นนั้นมีประสิทธิภาพ และจะพิจารณาค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่น แต่ถ้าช่วงความเชื่อมั่นใดให้ค่าประมาณความน่าจะเป็นคัมรวมต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด แสดงว่าช่วงความเชื่อมั่นนั้นไม่มีประสิทธิภาพ และจะไม่นำมาพิจารณาค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่น

7. คำนวณค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่น (average length; AL) โดยคำนวณค่าขีดจำกัดล่างและขีดจำกัดบนของช่วงความเชื่อมั่นทั้ง 5 วิธี แล้วนำไปคำนวณโดยใช้สูตร

$$AL = \frac{\sum_{i=1}^{10,000} (U_i - L_i)}{10,000}$$

เมื่อ U_i และ L_i คือ ขีดจำกัดบนและขีดจำกัดล่างของช่วงความเชื่อมั่น ในการทำซ้ำรอบที่ i

โดยที่ $i = 1, 2, \dots, 10,000$

ถ้าช่วงความเชื่อมั่นใดให้ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นน้อยที่สุด แสดงว่าช่วงความเชื่อมั่นนั้นมีประสิทธิภาพมากที่สุด

8. เปรียบเทียบค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นทั้ง 5 วิธี

9. สรุปผลการศึกษาของช่วงความเชื่อมั่นทั้ง 5 วิธีในแต่ละสถานการณ์

3. ผลและวิจารณ์ผลการวิจัย

ผลการศึกษางานวิจัยนี้ได้ทำการเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าช่วงความเชื่อมั่นสำหรับสัดส่วนของประชากรที่มีการแจกแจงทวินาม 5 วิธี ได้แก่ 1) วิธีวัลด์ (Wald) 2) วิธี Wilson score 3) วิธี Agresti-Coull 4) วิธี New และ 5) วิธี generalized Agresti-Coull ซึ่งเกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพมี 2 เกณฑ์ คือ 1) ค่าความน่าจะเป็นคัมรวม และ 2) ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่น ซึ่งวิธีที่มีประสิทธิภาพดีจะต้องให้ค่าความน่าจะเป็นคัมรวมไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด นั่นคือวิธีการประมาณค่าช่วงความเชื่อมั่นที่ให้ค่าประมาณความน่าจะเป็นคัมรวมไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด

0.95 เมื่อค่าประมาณความน่าจะเป็นคัมรวมมีค่าไม่ต่ำกว่า 0.946 และมีค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นน้อยที่สุด โดยการใช้การจำลองมอนติคาร์โลในสถานการณ์ต่าง ๆ ที่ศึกษา ดังแสดงในตารางที่ 1 และ 2 ตามลำดับ

ตารางที่ 1 แสดงค่าประมาณความน่าจะเป็นคัมรวมของช่วงความเชื่อมั่น 95% สำหรับสัดส่วนของประชากรที่มีการแจกแจงทวินามทั้ง 5 วิธี ผลการศึกษาพบว่าวิธี Wald (Wald) ให้ค่าประมาณความน่าจะเป็นคัมรวมต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดในหลายสถานการณ์ ยกเว้นในกรณีที่ $n=30$ เมื่อ $p=0.3-0.7$ และ $n=100$ เมื่อ $p=0.3$ และ 0.7 และมีแนวโน้มที่จะให้ค่าประมาณความน่าจะเป็นคัมรวมที่มากขึ้นเมื่อขนาดตัวอย่างใหญ่ขึ้น วิธี Wilson score จะให้ค่าประมาณความน่าจะเป็นคัมรวมไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด ในกรณีที่ $n=10$ เมื่อ $p=0.5$ $n=30$ เมื่อ $p=0.1, 0.5$ และ 0.9 และ $n=50$ สำหรับทุกค่าของ p ยกเว้น $p=0.5$ ช่วงความเชื่อมั่นวิธี Wald (Wald) และวิธี Wilson score จะให้ค่าประมาณความน่าจะเป็นคัมรวมที่มีความกว้างสูงในกรณีที่ค่าพารามิเตอร์สัดส่วนมีค่าเข้าใกล้ 0 หรือ 1 ส่วนวิธี Agresti-Coull ให้ค่าประมาณความน่าจะเป็นคัมรวมไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดในหลายสถานการณ์ ยกเว้นในกรณีที่ $n=10$ เมื่อ $p=0.1$ และ 0.9 และ $n=50$ และ 100 เมื่อ $p=0.5$ สำหรับวิธี New จะให้ค่าประมาณความน่าจะเป็นคัมรวมไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดในหลายสถานการณ์ที่ศึกษา ยกเว้นในกรณีที่ $n=10$ เมื่อ $p=0.3$ และ 0.7 $n=30$ เมื่อ $p=0.1$ และ 0.9 $n=50$ เมื่อ $p=0.1, 0.5$ และ 0.9 และ $n=100$ เมื่อ $p=0.5$ และวิธี generalized Agresti-Coull จะให้ค่าประมาณความน่าจะเป็นคัมรวมไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดในหลายสถานการณ์ ยกเว้นในกรณีที่ $n=10$ สำหรับทุกค่าของ p ยกเว้น $p=0.5$ $n=30$ เมื่อ $p=0.3$ และ 0.7 และ $n=50$ และ 100 เมื่อ $p=0.5$ ซึ่งผลการศึกษาสอดคล้องกับกราฟในรูปที่ 1 แสดงค่าประมาณความน่าจะเป็นคัมรวมของช่วงความเชื่อมั่น 95% สำหรับสัดส่วนของประชากรที่มีการแจกแจงทวินามทั้ง 5 วิธี โดยที่ W แทนวิธี Wald, WS แทนวิธี Wilson score, AC แทนวิธี Agresti-Coull, N แทนวิธี New และ gAC แทนวิธี generalized Agresti-Coull จะพบว่าเมื่อขนาดตัวอย่างใหญ่ขึ้นจะทำให้ช่วงความเชื่อมั่นทั้ง 5 วิธี มีแนวโน้มให้ค่าประมาณความน่าจะเป็นคัมรวมเพิ่มขึ้น

ตารางที่ 2 แสดงค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่น 95% สำหรับสัดส่วนของประชากรที่มีการแจกแจงทวินามทั้ง 5 วิธี ผลการศึกษาพบว่าช่วงความเชื่อมั่นทั้ง 5 วิธีให้ค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่นลดลงเมื่อตัวอย่างมีขนาดใหญ่ขึ้นในทุกระดับของค่า p วิธี Wilson score ให้ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นน้อยที่สุดในกรณีที่ $n=10$ เมื่อ $p=0.5$ $n=30$ เมื่อ $p=0.1$ และ 0.9 และ $n=50$ สำหรับทุกค่าของ p ยกเว้น $p=0.5$ วิธี Agresti-Coull จะให้ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นน้อยที่สุดในกรณีที่ $n=10$ เมื่อ $p=0.3$ และ 0.7 และ $n=30$ เมื่อ $p=0.3-0.7$ ส่วนวิธี New ให้ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นน้อยที่สุดในกรณีที่ $n=10$ และ 100 เมื่อ $p=0.1$ และ 0.9 และวิธี generalized Agresti-Coull จะให้ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นน้อยที่สุดในกรณีที่ $n=100$ เมื่อ $p=0.3$ และ 0.7 ซึ่งผลการศึกษาสอดคล้องกับกราฟในรูปที่ 2 แสดงค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่น 95% สำหรับสัดส่วนของประชากรที่มีการแจกแจงทวินามทั้ง 5 วิธี โดยที่ W แทนวิธี Wald, WS แทนวิธี Wilson score, AC แทนวิธี Agresti-Coull, N แทนวิธี New และ gAC แทนวิธี generalized Agresti-Coull

4. การประยุกต์ใช้กับข้อมูลจริง

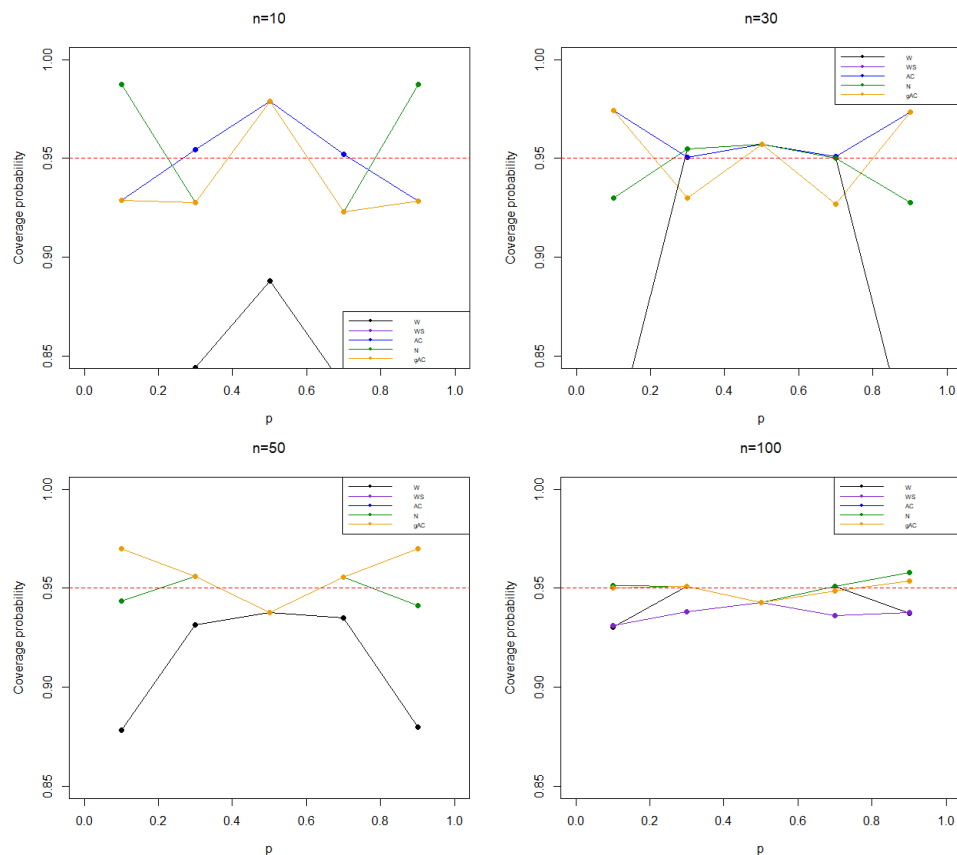
การศึกษาเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าช่วงความเชื่อมั่นสำหรับสัดส่วนของประชากรที่มีการแจกแจงทวินามทั้ง 5 วิธี สามารถนำมาประยุกต์ใช้กับข้อมูลในสถานการณ์จริง ซึ่งข้อมูลที่น่านำมาประยุกต์ใช้ในงานวิจัยนี้มาจากงานวิจัยของ Mizumoto et al. (2020) ได้ศึกษาข้อมูลเกี่ยวกับผู้โดยสารในเรือสำราญ Diamond Princess ที่ล่องเรืออยู่ในเมืองโยโกฮามะ ประเทศญี่ปุ่น ซึ่งมีผู้โดยสารในเรือได้ตรวจพบผู้ติดเชื้อไวรัสโคโรนา (COVID-19) จำนวน 634 คน ซึ่งพบว่าเป็นชาวญี่ปุ่นจำนวน 270 คน เป็นชาวอเมริกันจำนวน 88 คน และที่เหลือเป็นบุคคลที่มีสัญชาติต่าง ๆ จาก 28 ประเทศ

ในงานวิจัยนี้ได้นำข้อมูลมาทำการศึกษา 2 กรณี ได้แก่ กรณี 1) ศึกษาช่วงความเชื่อมั่น 95% สำหรับสัดส่วนของผู้โดยสารที่ติดเชื้อไวรัสโคโรนาในเรือสำราญที่เป็นชาวอเมริกัน และกรณี 2) ศึกษาช่วงความเชื่อมั่น

ตารางที่ 1 ค่าประมาณความน่าจะเป็นคัมรวมของช่วงความเชื่อมั่น 95% สำหรับสัดส่วนของประชากรที่มีการแจกแจงทวินามทั้ง 5 วิธี

ขนาดตัวอย่าง (n)	พารามิเตอร์ p	วิธี				
		Wald	Wilson score	Agresti-Coull	New	Generalized Agresti-Coull
10	0.1	0.6540	0.9288	0.9288	0.9876*	0.9288
	0.3	0.8444	0.9278	0.9544*	0.9278	0.9278
	0.5	0.8881	0.9788*	0.9788*	0.9788*	0.9788*
	0.7	0.8350	0.9230	0.9521*	0.9230	0.9230
	0.9	0.6469	0.9284	0.9284	0.9874*	0.9284
30	0.1	0.8066	0.9744*	0.9744*	0.9302	0.9744*
	0.3	0.9548*	0.9300	0.9507*	0.9548*	0.9300
	0.5	0.9573*	0.9573*	0.9573*	0.9573*	0.9573*
	0.7	0.9502*	0.9270	0.9508*	0.9502*	0.9270
	0.9	0.8023	0.9733*	0.9733*	0.9275	0.9733*
50	0.1	0.8783	0.9701*	0.9701*	0.9436	0.9701*
	0.3	0.9315	0.9560*	0.9560*	0.9560*	0.9560*
	0.5	0.9377	0.9377	0.9377	0.9377	0.9377
	0.7	0.9349	0.9555*	0.9555*	0.9555*	0.9555*
	0.9	0.8799	0.9701*	0.9701*	0.9411	0.9701*
100	0.1	0.9305	0.9313	0.9500*	0.9513*	0.9500*
	0.3	0.9509*	0.9383	0.9509*	0.9509*	0.9509*
	0.5	0.9428	0.9428	0.9428	0.9428	0.9428
	0.7	0.9510*	0.9361	0.9488*	0.9510*	0.9488*
	0.9	0.9373	0.9378	0.9535*	0.9581*	0.9535*

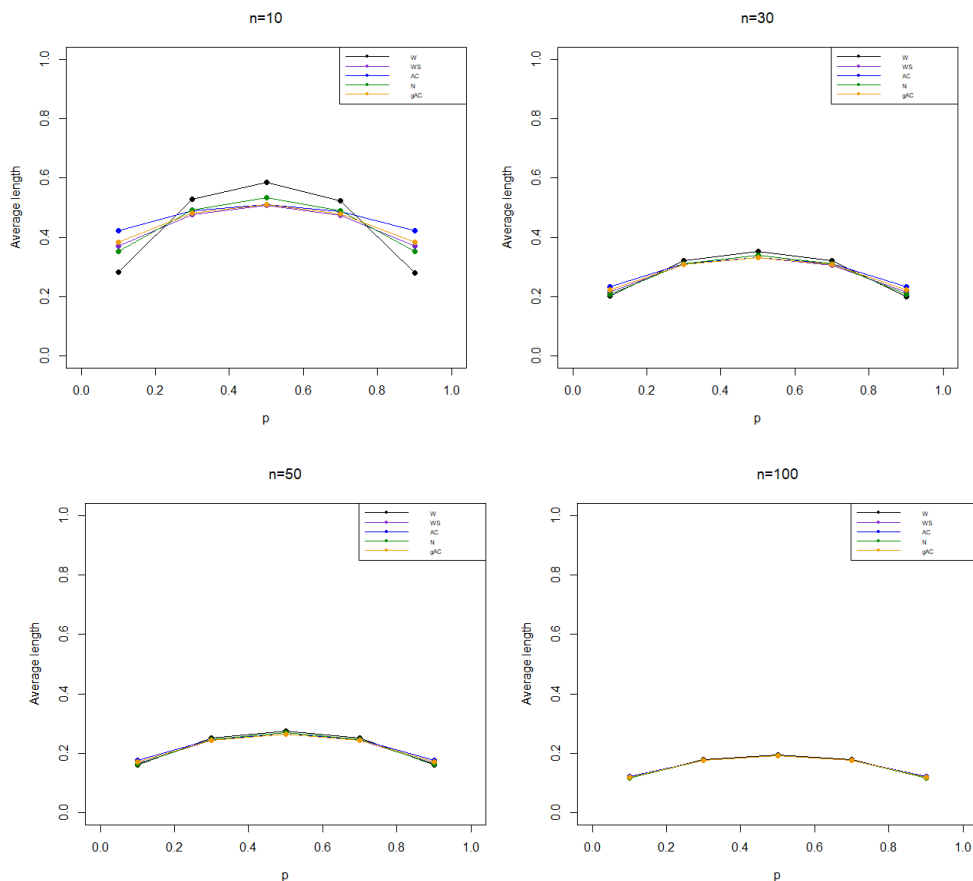
หมายเหตุ: * หมายถึง วิธีการประมาณช่วงความเชื่อมั่นให้ค่าความน่าจะเป็นคัมรวมไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด

รูปที่ 1 ค่าประมาณความน่าจะเป็นคัมรวมของช่วงความเชื่อมั่น 95% สำหรับสัดส่วนของประชากรที่มีการแจกแจงทวินามทั้ง 5 วิธี กรณีขนาดตัวอย่าง (n) แตกต่างกัน

ตารางที่ 2 ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่น 95% สำหรับสัดส่วนของประชากรที่มีการแจกแจงทวินามทั้ง 5 วิธี

ขนาดตัวอย่าง (n)	พารามิเตอร์ p	วิธี				
		Wald	Wilson score	Agresti-Coull	New	Generalized Agresti-Coull
10	0.1	0.2818	0.3696	0.4217	0.3534**	0.3839
	0.3	0.5283	0.4762	0.4896**	0.4932	0.4824
	0.5	0.5856	0.5066**	0.5099	0.5330	0.5108
	0.7	0.5235	0.4737	0.4880**	0.4900	0.4801
	0.9	0.2797	0.3691	0.4214	0.3527**	0.3835
30	0.1	0.2011	0.2148**	0.2335	0.2080	0.2229
	0.3	0.3216	0.3070	0.3102**	0.3121	0.3087
	0.5	0.3517	0.3318	0.3317**	0.3393	0.3322
	0.7	0.3212	0.3067	0.3100**	0.3118	0.3084
	0.9	0.2000	0.2141**	0.2330	0.2071	0.2223
50	0.1	0.1607	0.1662**	0.1761	0.1630	0.1710
	0.3	0.2509	0.2437**	0.2453	0.2463	0.2446
	0.5	0.2744	0.2646	0.2644	0.2682	0.2647
	0.7	0.2509	0.2437**	0.2453	0.2462	0.2446
	0.9	0.1612	0.1666**	0.1765	0.1634	0.1714
100	0.1	0.1157	0.1175	0.1214	0.1166**	0.1196
	0.3	0.1786	0.1759	0.1765	0.1768	0.1763**
	0.5	0.1950	0.1914	0.1913	0.1925	0.1914
	0.7	0.1785	0.1759	0.1764	0.1767	0.1762**
	0.9	0.1160	0.1178	0.1217	0.1169**	0.1199

หมายเหตุ: ** หมายถึง วิธีการประมาณช่วงความเชื่อมั่นให้ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นน้อยที่สุดโดยพิจารณาจากช่วงความเชื่อมั่นที่ให้ค่าความน่าจะเป็นคัมรวมไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดซึ่งแสดงในตารางที่ 1



รูปที่ 2 ค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่น 95% สำหรับสัดส่วนของประชากรที่มีการแจกแจงทวินามทั้ง 5 วิธี กรณีขนาดตัวอย่างแตกต่างกัน

ตารางที่ 3 ช่วงความเชื่อมั่น 95% สำหรับสัดส่วนของประชากรที่มีการแจกแจงทวินาม และค่าความกว้างของช่วงความเชื่อมั่นทั้ง 5 วิธี เมื่อนำมาประยุกต์ใช้กับข้อมูลจริง

วิธีการประมาณค่าช่วงความเชื่อมั่น	กรณีศึกษา			
	กรณี 1 ($p = 88/634 = 0.14$)		กรณี 2 ($p = 270/634 = 0.43$)	
	ช่วงความเชื่อมั่น 95%	ค่าความกว้างของช่วงความเชื่อมั่น	ช่วงความเชื่อมั่น 95%	ค่าความกว้างของช่วงความเชื่อมั่น
วิธี Wald (Wald)	(0.1119, 0.1657)	0.0538	(0.3874, 0.4644)	0.0770
วิธี Wilson score	(0.1141, 0.1679)	0.0538	(0.3879, 0.4647)	0.0768
วิธี Agresti-Coull	(0.1141, 0.1681)	0.0540	(0.3880, 0.4647)	0.0767
วิธี New	(0.1135, 0.1674)	0.0539	(0.3878, 0.4646)	0.0767
วิธี generalized Agresti-Coull	(0.1142, 0.1681)	0.0540	(0.3880, 0.4647)	0.0767

ตารางที่ 4 วิธีการประมาณค่าช่วงความเชื่อมั่น 95% สำหรับสัดส่วนของประชากรที่มีการแจกแจงทวินามที่เหมาะสมกับสถานการณ์ต่าง ๆ ที่ศึกษา

ขนาดตัวอย่าง (n)	พารามิเตอร์ p				
	0.1	0.3	0.5	0.7	0.9
10	N	AC	WS	AC	N
30	WS	AC	AC	AC	WS
50	WS	WS	-	WS	WS
100	N	gAC	-	gAC	N

หมายเหตุ: WS แทน วิธี Wilson score; AC แทน วิธี Agresti-Coull; N แทน วิธี New และ gAC แทน วิธี generalized Agresti-Coull

- แทน ไม่มีวิธีการประมาณค่าช่วงความเชื่อมั่นใดที่เหมาะสม

ความเชื่อมั่น 95% และค่าความกว้างของช่วงความเชื่อมั่นทั้ง 5 วิธี สำหรับสัดส่วนของผู้โดยสารที่ติดเชื้อไวรัสโคโรนาในเรือสำราญที่เป็นชาวอเมริกันและชาวญี่ปุ่น ดังแสดงในตารางที่ 3

ผลจากตารางที่ 3 จะพบว่าวิธีการประมาณค่าช่วงความเชื่อมั่นทั้ง 5 วิธี ในกรณีที่ศึกษาทั้ง 2 กรณี คือกรณีที่ตัวอย่างมีขนาดใหญ่ ($n = 634$) และค่า p มีค่าเข้าใกล้ 0 ($p = 0.14$) และค่า $p = 0.43$ จะให้ค่าประมาณช่วงความเชื่อมั่น 95% และค่าความกว้างของช่วงความเชื่อมั่นสำหรับสัดส่วนของประชากรที่มีการแจกแจงทวินามที่ใกล้เคียงกันในทุกกรณี เนื่องจากตัวอย่างที่นำมาศึกษามีขนาดใหญ่ ซึ่งจะพบว่ามีผลสอดคล้องกับผลการศึกษาเชิงจำลองในกรณีที่ตัวอย่างมีขนาดใหญ่ ช่วงความเชื่อมั่นทั้ง 5 วิธีจะมีประสิทธิภาพมากขึ้นในการคำนวณหาค่าประมาณช่วงความเชื่อมั่น 95% และค่าความกว้างของช่วงความเชื่อมั่นจึงมีค่าที่ใกล้เคียงกันในทุกกรณี สำหรับการศึกษกรณี 1 ช่วงความเชื่อมั่นวิธี Wald และวิธี Wilson score จะมีค่าประมาณความกว้างของช่วงความเชื่อมั่นน้อยที่สุดจึงมีประสิทธิภาพมากที่สุด รองลงมาคือช่วงความเชื่อมั่นวิธี New วิธี Agresti-Coull และวิธี generalized Agresti-Coull ส่วนกรณีที่ศึกษาที่ 2 ช่วงความ

เชื่อมั่นวิธี Agresti-Coull วิธี New และวิธี generalized Agresti-Coull จะให้ค่าประมาณความกว้างของช่วงความเชื่อมั่นน้อยที่สุดจึงมีประสิทธิภาพมากที่สุด รองลงมาคือช่วงความเชื่อมั่น วิธี Wilson score และวิธี Wald

สรุปผลการวิจัย

ผลจากการศึกษาเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าช่วงความเชื่อมั่น 95% สำหรับสัดส่วนของประชากรที่มีการแจกแจงทวินาม 5 วิธี ได้แก่ 1) วิธี Wald 2) วิธี Wilson score 3) วิธี Agresti-Coull 4) วิธี New และ 5) วิธี generalized Agresti-Coull โดยพิจารณาประสิทธิภาพของช่วงความเชื่อมั่นที่ให้ค่าความน่าจะเป็นคุ่มรวมไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดไว้คือ 0.95 และให้ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นน้อยที่สุด จะพบว่าช่วงความเชื่อมั่นที่เหมาะสมกับสถานการณ์ต่าง ๆ ที่ศึกษาโดยจำแนกตามขนาดตัวอย่าง (n) และ ค่าพารามิเตอร์ p แสดงดังตารางที่ 4

ตารางที่ 4 แสดงวิธีการประมาณค่าช่วงความเชื่อมั่น 95% สำหรับสัดส่วนของประชากรที่มีการแจกแจงทวินาม

ที่เหมาะสมกับสถานการณ์ต่าง ๆ ที่ศึกษาโดยจำแนกตามขนาดตัวอย่าง (n) และค่าพารามิเตอร์ p จะพบว่าช่วงความเชื่อมั่นวิธี Wilson score มีประสิทธิภาพดีที่สุดใหลายสถานการณ์ที่ศึกษา ซึ่งช่วงความเชื่อมั่นนี้เหมาะกับสถานการณ์ที่ตัวอย่างมีขนาดเล็ก $n=10$ เมื่อ $p=0.5$ และตัวอย่างขนาดกลาง $n=30$ เมื่อ p มีค่าเข้าใกล้ 0 หรือ 1 ($p=0.1, 0.9$) และ $n=50$ สำหรับทุกค่าของ p ยกเว้น $p=0.5$ ช่วงความเชื่อมั่นวิธี Agresti-Coull มีประสิทธิภาพเป็นอันดับที่สอง ซึ่งเหมาะกับตัวอย่างขนาดเล็ก $n=10$ เมื่อ $p=0.3$ และ 0.7 และตัวอย่างขนาดกลาง $n=30$ เมื่อ $p=0.3-0.7$ ส่วนช่วงความเชื่อมั่นวิธี New มีประสิทธิภาพเป็นอันดับที่สาม ซึ่งเหมาะกับตัวอย่างขนาดเล็ก $n=10$ และตัวอย่างขนาดใหญ่ $n=100$ เมื่อ p มีค่าเข้าใกล้ 0 หรือ 1 ($p=0.1, 0.9$) ช่วงความเชื่อมั่นวิธี generalized Agresti-Coull มีประสิทธิภาพเป็นอันดับที่สี่ ซึ่งเหมาะกับตัวอย่างขนาดใหญ่ $n=100$ เมื่อ $p=0.3$ และ 0.7 และช่วงความเชื่อมั่นวิธี Wald มีประสิทธิภาพที่ไม่ดีนักเนื่องจากให้ค่าความน่าจะเป็นคุ่มรวมต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดไว้ในหลายสถานการณ์ที่ศึกษาแต่มีแนวโน้มให้ประสิทธิภาพดีเมื่อตัวอย่างมีขนาดใหญ่ นอกจากนี้การประยุกต์ช่วงความเชื่อมั่นทั้ง 5 วิธีนำไปใช้กับข้อมูลจริงในสถานการณ์ต่าง ๆ จะให้ผลการศึกษาประสิทธิภาพของช่วงความเชื่อมั่นแต่ละวิธีมีความสอดคล้องกับผลการศึกษาเชิงจำลอง

เอกสารอ้างอิง

- Agresti, A., & Coull, B. A. (1998). Approximate is better than "Exact" for interval estimation of binomial proportions. *The American Statistician*, 52(2), 118-126. <https://doi.org/10.1080/00031305.1998.10480550>
- Brown, L. D., Cai, T., & Dasgupta, A. (2001). Interval estimation for a binomial proportion. *Statistical Science*, 16(2), 101-133. <https://doi.org/10.1214/ss/1009213286>
- Brown, L. D., Cai, T., & Dasgupta, A. (2002). Confidence intervals for a binomial proportion and asymptotic expansions. *The Annals of Statistics*, 30(1), 160-201. <https://doi.org/10.1214/aos/1015362189>
- Cao, Y., Xu, Y., Tan, M. T., Chen, P., & Duan, C. (2022). A simple and improved score confidence interval for a single proportion. *Communications in Statistics-Theory and Methods*, 51(8), 2659-2675. <https://doi.org/10.1080/03610926.2020.1779747>
- Clopper, C. J., & Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26(4), 404-413.
- Fisher, A. R., & Cornish, E. A. (1960). The percentile points of distributions having known cumulants. *American Society for Quality*, 2(2), 209-225.
- Kim, J., & Jang, W. (2022). A generalized Agresti-Coull type confidence interval for a binomial proportion. *Journal of the Korean Statistical*

- Society*, 51, 356-377. <https://doi.org/10.1007/s42952-021-00143-3>
- Komoltri, J. (2012). Sample size calculation. *Journal of Mental Health of Thailand*, 20(3), 192-198. (in Thai)
- Mizumoto, K., Kagaya, K., Zarebski, A., & Chowell, G. (2020). Estimating the Asymptomatic Proportion of Coronavirus Disease 2019 (COVID-19) Case on Board the Diamond Princess Cruise Ship, Yokohama, Japan, 2020. *Euro Surveill*, 25(10), 1-5. <https://doi.org/10.2807/1560-7917.ES.2020.25.10.2000180>
- Muenmee, S., & Bootdee, S. (2021). Health risk assessment of exposure PM2.5 from industrial area in Pluak Daeng District, Rayong province. *Naresuan Phayao Journal*, 14(3), 95-110. (in Thai)
- Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is Large. *American Mathematical Society*, 54(3), 426-482.
- Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22(158), 201-212. <https://doi.org/10.2307/2276774>

ช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยประชากรของการแจกแจงเรย์ลีย์: กรณีศึกษา ความเร็วคลื่นเฉือนของดินในประเทศไทย

Confidence Interval for the Population Mean of Rayleigh Distribution with Application to Shear wave Velocity of Soils in Thailand

ธนกฤต เตียวพงษ์พันธุ์ และ ภทรวรรณ แสงนวกิจ*

Thanakit Teawpongpan and Patarawan Sangnawakij*

สาขาวิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ ประเทศไทย

Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, Thailand

บทคัดย่อ

การแจกแจงเรย์ลีย์เป็นการแจกแจงความน่าจะเป็นแบบต่อเนื่องชนิดหนึ่งที่ถูกนำไปใช้ในหลากหลายด้าน โดยเฉพาะทางด้านวิศวกรรม และทางการแพทย์ การแจกแจงนี้ถูกพบบ่อยครั้งในปรากฏการณ์ทางธรรมชาติ เนื่องจากการแจกแจงเรย์ลีย์เกี่ยวข้องกับข้อมูลเชิงปริมาณในการศึกษาครั้งนี้ผู้วิจัยจึงได้นำเสนอวิธีประมาณค่าแบบช่วงสำหรับค่าเฉลี่ยประชากร การสร้างช่วงความเชื่อมั่นในครั้งนี้อาศัยวิธีของวาลด์ วิธีสร้างช่วงความเชื่อมั่นสำหรับตัวอย่างขนาดใหญ่ และวิธีช่วงความเชื่อมั่นน้อยทั่วไป ผู้วิจัยได้ทำการศึกษาประสิทธิภาพของช่วงความเชื่อมั่นโดยอาศัยการจำลองและใช้เกณฑ์การวัดประสิทธิภาพจากค่าความน่าจะเป็นคัมรวมและความยาวเฉลี่ยของช่วงความเชื่อมั่น ผลที่ได้จากการจำลองพบว่าช่วงความเชื่อมั่นน้อยทั่วไปมีประสิทธิภาพที่ดีในทุกสถานการณ์ และดีกว่าช่วงความเชื่อมั่นที่นำมาเปรียบเทียบกับการศึกษาครั้งนี้ผู้วิจัยได้นำวิธีการทางทฤษฎีมาประยุกต์ใช้กับข้อมูลจริงที่เกี่ยวข้องกับความเร็วคลื่นเฉือนของดินในจังหวัดลำพูน ประเทศไทย ทั้งนี้เพื่อเป็นการยืนยันผลที่ได้จากการจำลอง

คำสำคัญ: การประมาณค่าแบบช่วง, การแจกแจงเรย์ลีย์, ความน่าจะเป็นคัมรวม, ช่วงความเชื่อมั่นน้อยทั่วไป

ABSTRACT

Rayleigh distribution is one of the continuous probability distributions applied in several areas, especially in engineering and medicine. This probability model is also widely occurred in natural phenomena. Since Rayleigh distribution is related to quantitative data, we are interested in studying interval estimation for the population mean. The methods based on Wald-type confidence interval, large-sample confidence interval, and generalized confidence intervals are used in construction. Then, the performance of the proposed confidence intervals is conducted using simulations in terms of coverage probability and expected length. The results illustrate that the novel generalized confidence interval performs better than the compared confidence intervals in all cases in the study. Finally, a real data set on the shear wave velocity of soils in Lamphun province of Thailand is used in computation, to confirm the results from simulation.

KEYWORDS: Interval Estimation, Rayleigh Distribution, Coverage Probability, Generalized Confidence Interval

*Corresponding Author: patarawan@mathstat.sci.tu.ac.th

Received: 01/02/2023; Revised: 19/06/2023; Accepted: 14/07/2023

1. บทนำ

การประมาณค่าพารามิเตอร์ (Parameter estimation) เป็นวิธีการทางสถิติที่ใช้ในการศึกษาค่าจริงของประชากร โดยอาศัยข้อมูลจากตัวอย่างซึ่งเป็นส่วนหนึ่งของประชากร ที่สนใจศึกษา การประมาณค่าพารามิเตอร์แบ่งออกเป็น 2 แบบ ได้แก่ การประมาณค่าแบบจุด และการประมาณ

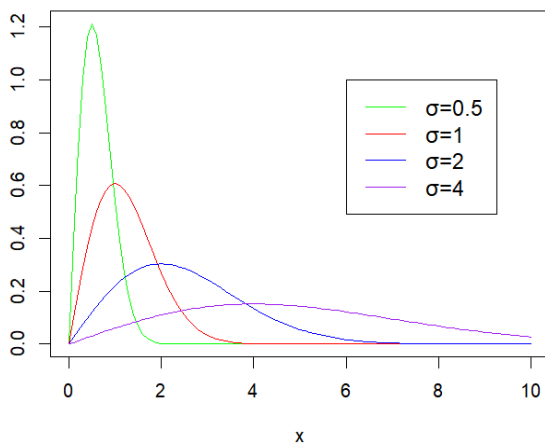
ค่าแบบช่วง โดยการประมาณค่าแบบจุดจะเป็นการประมาณค่าพารามิเตอร์จากค่าสถิติเพียงค่าเดียว ซึ่งต่างจากการประมาณค่าแบบช่วงที่ให้ช่วงค่าสถิติสองค่า อย่างไรก็ตามเนื่องจากการประมาณค่าแบบช่วงสามารถอธิบายระดับความเชื่อมั่น (Confidence level) ที่ค่าสถิติจะประมาณค่าพารามิเตอร์ได้ถูกต้อง (Casella & Berger,

2002) ดังนั้นจึงเป็นวิธีที่สามารถให้ผลสรุปได้มากกว่าการประมาณค่าแบบจุด

การแจกแจงเรย์ลีย์ (Rayleigh distribution) เป็นการแจกแจงชนิดต่อเนื่องที่เกี่ยวข้องกับตัวแปรสุ่มที่มีค่ามากกว่าศูนย์ การแจกแจงนี้ถูกพบบ่อยครั้งในข้อมูลที่เกิดขึ้นจากปรากฏการณ์ทางธรรมชาติ เช่น ความเร็วลม ปริมาณฝุ่นละอองขนาดเล็กในอากาศ และความเร็วคลื่นเฉือนของดินซึ่งเป็นสัญญาณของการเกิดแผ่นดินไหว (Pishgar-Komleh et al., 2015) นอกจากนี้ยังพบได้ในข้อมูลที่เกี่ยวข้องกับสัญญาณภาพถ่ายทางดาวเทียม และสัญญาณภาพถ่ายอัลตราซาวด์ (Ultrasound) (Sarti et al., 2005; Ahn et al., 2012; Nar et al., 2018) สำหรับทฤษฎีความน่าจะเป็นเกี่ยวกับการแจกแจงเรย์ลีย์มีดังนี้ กำหนดให้ X เป็นตัวแปรสุ่มที่มีการแจกแจงเรย์ลีย์ด้วยพารามิเตอร์กำหนดรูปร่าง (Shape parameter) σ ในที่นี้แทนด้วย $Ra(\sigma)$ แล้ว ฟังก์ชันความหนาแน่นความน่าจะเป็น (Probability density function) ของ X คือ

$$f_X(x; \sigma) = \frac{x}{\sigma^2} \exp\left\{-\frac{x^2}{2\sigma^2}\right\} \quad (1)$$

โดยที่ $x > 0$ และ $\sigma > 0$ ค่าคาดหวังและความแปรปรวนของตัวแปรสุ่ม X คือ $E(X) = \psi = \sigma\sqrt{\pi}/2$ และ $Var(X) = (4-\pi)\sigma^2/2$ ตามลำดับ โดยที่ $\pi \approx 3.142$ (Ateeq et al., 2019) การแจกแจงเรย์ลีย์มีลักษณะเบ้ขวา โดยความเบ้ของข้อมูลจะขึ้นอยู่กับค่าพารามิเตอร์ σ ดังแสดงในรูปที่ 1 จากรูปพบว่าเมื่อ σ มีค่าน้อยข้อมูลจะมีลักษณะเบ้มาก แต่เมื่อ σ มีค่ามากขึ้นความเบ้ของข้อมูลจะลดลง



รูปที่ 1 กราฟความน่าจะเป็นของการแจกแจงเรย์ลีย์ เมื่อกำหนดค่า σ ต่างกัน

จากที่ได้กล่าวไปข้างต้นว่าการแจกแจงเรย์ลีย์ถูกนำไปประยุกต์ใช้ในหลายด้าน ตัวอย่างงานวิจัยที่เกี่ยวข้องมีดังนี้ การสร้างแบบจำลองความเร็วคลื่นเฉือนของดินแบบบูรณาการสำหรับแหล่งก๊าซไกรนิงเกน (Groningen) ในประเทศเนเธอร์แลนด์ เพื่อใช้จำลองความเสี่ยงในการเกิดแผ่นดินไหว (Kruiver et al., 2017) การประมาณค่าพลังงานและความเร็วลม ณ เกาะเซนต์มาร์ติน ประเทศบังกลาเทศ (Islam et al., 2017) และการพยากรณ์ความเร็วลมเพื่อทำนายสภาพอากาศในมณฑลเจียงซู ประเทศจีน (Xu et al., 2020) ในทางทฤษฎีมีนักวิจัยได้นำเสนอวิธีประมาณค่าพารามิเตอร์ในการแจกแจงเรย์ลีย์ดังนี้ Elfattah et al. (2006) ได้ศึกษาผลกระทบของรูปแบบการสุ่มตัวอย่างแบบเซ็นเซอร์ (Censored sampling) ที่มีผลต่อการประมาณค่าพารามิเตอร์ σ โดยอาศัยวิธีภาวะน่าจะเป็นสูงสุด Dey & Dey (2009) ได้ศึกษาวิธีประมาณค่าแบบจุดสำหรับ σ โดยอาศัยวิธีเบย์ส (Bayesian's method) ที่มีการแจกแจงก่อนเป็นฟังก์ชันสูญเสีย (Loss function) และฟังก์ชันความน่าเชื่อถือได้ (Reliability function) Ahmed et al. (2013) ได้ทำการเปรียบเทียบประสิทธิภาพของการประมาณค่าแบบจุดสำหรับ σ ด้วยวิธีประมาณค่าแบบเบย์ที่มีการแจกแจงก่อนเป็นฟังก์ชันสูญเสีย และวิธีภาวะน่าจะเป็นสูงสุด Seo et al. (2016) ได้สร้างช่วงความเชื่อมั่นที่แน่นอน (Exact confidence interval) ในการแจกแจงเรย์ลีย์สำหรับพารามิเตอร์ σ โดยอาศัยปริมาณหมุน (Pivotal quantity) ที่มีการแจกแจงไคกำลังสอง (Chi-square distribution) และ Liebenberg & Allison (2023) ได้ศึกษาวิธีการสร้างสถิติทดสอบภาวะสารูปดี (Goodness of fit test) สำหรับการทดสอบการแจกแจงเรย์ลีย์

จากการทบทวนวรรณกรรมที่เกี่ยวข้องจะพบว่ามีผู้เสนอตัวประมาณค่าพารามิเตอร์แบบจุดและแบบช่วงสำหรับ σ ในการแจกแจงเรย์ลีย์อย่างต่อเนื่อง แต่การประมาณค่าแบบช่วงสำหรับค่าเฉลี่ยประชากร (ψ) ในการแจกแจงเรย์ลีย์นั้น ยังไม่พบว่ามีผู้ที่ทำการศึกษามากนัก โดยเฉพาะอย่างยิ่งการสร้างช่วงความเชื่อมั่นที่อาศัยวิธีของวาลด์จากการประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุด วิธีสร้างช่วงความเชื่อมั่นสำหรับตัวอย่างขนาดใหญ่ที่อาศัยตัวประมาณโมเมนต์ รวมทั้งวิธีการสร้างช่วงความเชื่อมั่นนัยทั่วไปนั้น ยังไม่มีผู้ทำการศึกษา ดังนั้นการนำเสนอช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยประชากรในการแจกแจงเรย์ลีย์ และการศึกษาประสิทธิภาพของช่วงความเชื่อมั่นที่ได้จากการ

จำลองจึงเป็นวัตถุประสงค์หลักของงานวิจัยฉบับนี้ นอกจากการศึกษาทางด้านทฤษฎีแล้วผู้วิจัยได้นำข้อมูลจริงซึ่งเป็นค่าความเร็วคลื่นเฉือนของดินในจังหวัดลำพูนมาใช้ในการคำนวณซึ่งรายละเอียดจะกล่าวถึงในบทประยุกต์ต่อไป

2. วิธีการทางสถิติ

กำหนดให้ $X = (X_1, X_2, \dots, X_n)$ เป็นตัวอย่างสุ่ม (Random sample) ขนาด n จากประชากรที่มีการแจกแจงเรย์ลีย์ด้วยพารามิเตอร์ σ ซึ่งไม่ทราบค่า และมีฟังก์ชันความหนาแน่นความน่าจะเป็นตามสมการ (1) ค่าคาดหวังของ X คือ $\psi = \sigma\sqrt{\pi/2}$ โดยที่ $\sigma > 0$ และ $\pi \approx 3.142$ ในการศึกษาครั้งนี้ ψ คือ พารามิเตอร์ที่สนใจศึกษาและต้องการหาค่าประมาณ

การประมาณค่าแบบจุดของ σ และ ψ โดยอาศัยวิธีโมเมนต์จะพิจารณาโมเมนต์ที่ k รอบจุดกำเนิดของประชากร ($E(X^k)$) และโมเมนต์ที่ k รอบจุดกำเนิดของตัวอย่าง ($\sum_{i=1}^n X_i^k / n$) โดยที่ $k = 1, 2, \dots$ และเนื่องจาก $\sum_{i=1}^n X_i^k / n$ เป็นตัวประมาณไม่เอนเอียงของ $E(X^k)$ ดังนั้น

$$E(X^k) = \frac{1}{n} \sum_{i=1}^n X_i^k$$

เมื่อกำหนดให้ $k = 1$ จะได้ว่า

$$E(X) = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{และ} \quad \sigma\sqrt{\pi/2} = \bar{X}$$

กล่าวคือ $\hat{\sigma}_{mm} = \bar{X}\sqrt{2/\pi}$ เป็นตัวประมาณโมเมนต์สำหรับ σ และจะได้ว่าตัวประมาณโมเมนต์ของ ψ คือ $\hat{\psi}_{mm} = \hat{\sigma}_{mm}\sqrt{\pi/2}$ สำหรับวิธีภาวะน่าจะเป็นสูงสุดเป็นวิธีที่ใช้ในการประมาณค่าพารามิเตอร์ที่ทำให้ล็อกของฟังก์ชันภาวะน่าจะเป็น (Log-likelihood function) ของพารามิเตอร์ เมื่อทราบ $(X_1, X_2, \dots, X_n) = (x_1, x_2, \dots, x_n)$ มีค่าสูงสุด และเนื่องจาก X_i มีการแจกแจง $Ra(\sigma)$ โดยที่ $i = 1, 2, \dots, n$ ดังนั้นล็อกของฟังก์ชันภาวะน่าจะเป็นของ σ คือ

$$\ln L(\sigma; x_i) = \ln \prod_{i=1}^n x_i - 2n \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n x_i \quad (2)$$

เมื่อหาอนุพันธ์อันดับหนึ่งของ (2) เทียบกับ σ และกำหนดให้เท่ากับศูนย์แล้ว จะได้ตัวประมาณภาวะน่าจะเป็นสูงสุดของ σ คือ $\hat{\sigma}_{MLE} = \sqrt{\sum_{i=1}^n X_i^2 / 2n}$ อาศัยสมบัติไม่แปรเปลี่ยน (Invariance property) ดังนั้นตัว

ประมาณภาวะน่าจะเป็นสูงสุดของ ψ คือ $\hat{\psi}_{MLE} = \hat{\sigma}_{MLE}\sqrt{\pi/2}$

ตัวประมาณค่าเฉลี่ยประชากรที่ได้จากวิธีโมเมนต์ และวิธีภาวะน่าจะเป็นสูงสุดข้างต้นจะถูกนำไปใช้ในการสร้างช่วงความเชื่อมั่นในหัวข้อต่อไป

1. ช่วงความเชื่อมั่นจากวิธีของวาลด์

วิธีของวาลด์เป็นวิธีที่ใช้ในการสร้างช่วงความเชื่อมั่นที่อาศัยสมบัติของตัวประมาณภาวะน่าจะเป็นสูงสุด ถ้าสมมติให้ $\hat{\psi}_{MLE}$ เป็นตัวประมาณภาวะน่าจะเป็นสูงสุดของ $\psi = \sigma\sqrt{\pi/2}$ เมื่อตัวอย่างมีขนาดใหญ่ ($n \rightarrow \infty$) แล้ว $\hat{\psi}_{MLE}$ จะมีการแจกแจงปกติโดยประมาณ (Approximate normal distribution) ด้วยค่าเฉลี่ย ψ และความแปรปรวน $\text{Var}(\hat{\psi})$ ดังนั้นในกรณีของตัวประมาณ $\hat{\psi}_{MLE}$ ที่ได้นำเสนอไปข้างต้นจะหาความแปรปรวนได้จาก

$$\text{Var}(\hat{\psi}_{MLE}) = \text{Var}\left(\sqrt{\frac{\pi}{2}}\hat{\sigma}_{MLE}\right) = \frac{\pi}{2} \text{Var}(\hat{\sigma}_{MLE})$$

โดยที่ $\text{Var}(\hat{\sigma}_{MLE}) = \text{Var}\left(\sqrt{\sum_{i=1}^n X_i^2} / 2n\right)$ เมื่ออาศัยวิธีเดลตา (Delta method) จากนิยาม

$$\text{Var}(g(x)) \cong \text{Var}(X)[g'(x)]^2_{X=E(X)}$$

(Casella & Berger, 2002) ในการประมาณความแปรปรวนของ $g(x) = \sqrt{\sum_{i=1}^n X_i^2}$ จะได้ว่า

$$\text{Var}\left(\sqrt{\sum_{i=1}^n X_i^2}\right) \cong \frac{(4-\pi)n\pi\sigma^2}{8} \quad \text{และ} \quad \text{Var}(\hat{\sigma}_{MLE}) \cong \frac{(4-\pi)\pi\sigma^2}{16}$$

อาศัยการพิสูจน์ข้างต้นทำให้ได้ความแปรปรวนโดยประมาณของ $\hat{\psi}_{MLE}$ ดังนี้

$$\text{Var}(\hat{\psi}_{MLE}) \cong \left(\frac{(4-\pi)\pi^2}{32}\right)\sigma^2 \quad (3)$$

ดังนั้นช่วงความเชื่อมั่น $(1-\alpha) 100\%$ สำหรับ ψ จากวิธีของวาลด์ คือ

$$CI_{NMLE} = (l_{NMLE}, u_{NMLE}) \\ = \left(\hat{\psi}_{MLE} - z_{1-\alpha/2}\sqrt{\text{var}(\hat{\psi}_{MLE})}, \hat{\psi}_{MLE} + z_{1-\alpha/2}\sqrt{\text{var}(\hat{\psi}_{MLE})}\right)$$

เมื่อ $\text{var}(\hat{\psi}_{MLE})$ คือ ตัวประมาณค่าความแปรปรวนของ $\hat{\psi}_{MLE}$ ซึ่ง $\text{var}(\hat{\psi}_{MLE})$ หาได้จากการแทนค่า σ^2 ในสมการ (3) ด้วย $\hat{\sigma}_{MLE}^2 = \sum_{i=1}^n X_i^2 / 2n$ และ $z_{1-\alpha/2}$ คือเปอร์เซ็นต์ไทล์ที่ $(1-\alpha/2) \times 100$ ของการแจกแจงปกติมาตรฐาน (Standard normal distribution)

2. ช่วงความเชื่อมั่นสำหรับตัวอย่างขนาดใหญ่ที่อาศัย $\hat{\psi}_{mm}$

การสร้างช่วงความเชื่อมั่นในหัวข้อนี้อาศัยทฤษฎีขีดจำกัดส่วนกลาง (Central limit theorem) เนื่องจาก $\hat{\psi}_{mm} = \hat{\sigma}_{mm} \sqrt{\frac{\pi}{2}} = \bar{X}$ ดังนั้น ปริมาณหมุน (Pivotal quantity) ที่ได้จาก $\frac{\hat{\psi}_{mm} - \psi}{\sqrt{\text{Var}(\hat{\psi}_{mm})}}$ จึงมีการแจกแจงปกติมาตรฐาน เมื่อผู้วิจัยทำการหาความแปรปรวนของ $\hat{\psi}_{mm}$ จึงได้ว่า

$$\text{Var}(\hat{\psi}_{mm}) = \text{Var}\left(\sqrt{\frac{\pi}{2}} \hat{\sigma}_{mm}\right) = \frac{\pi}{2} \text{Var}(\hat{\sigma}_{mm})$$

โดยที่ $\text{Var}(\hat{\sigma}_{mm}) = (4-\pi)\sigma^2 / n\pi$ เมื่อแทนค่าลงในสมการข้างต้นจะได้

$$\text{Var}(\hat{\psi}_{mm}) = \left(\frac{4-\pi}{2n}\right)\sigma^2 \quad (4)$$

ดังนั้นช่วงความเชื่อมั่น $(1-\alpha) 100\%$ สำหรับ ψ จากวิธีที่อาศัยตัวประมาณโมเมนต์ $\hat{\psi}_{mm}$ คือ

$$CI_{Nmm} = (l_{Nmm}, u_{Nmm}) \\ = (\hat{\psi}_{mm} - z_{1-\alpha/2} \sqrt{\text{var}(\hat{\psi}_{mm})}, \hat{\psi}_{mm} + z_{1-\alpha/2} \sqrt{\text{var}(\hat{\psi}_{mm})})$$

เมื่อ $\text{var}(\hat{\psi}_{mm})$ คือ ตัวประมาณค่าความแปรปรวนของ $\hat{\psi}_{mm}$ ซึ่ง $\text{var}(\hat{\psi}_{mm})$ หาได้จากการแทนค่า σ^2 ในสมการ (4) ด้วย $\hat{\sigma}_{MLE}^2 = 2\bar{X}^2 / \pi$

3. ช่วงความเชื่อมั่นสำหรับตัวอย่างขนาดใหญ่ที่อาศัย $\hat{\sigma}_{MLE}^2$

การสร้างช่วงความเชื่อมั่นในหัวข้อนี้อาศัยสมบัติของตัวประมาณภาวะนั้นจะเป็นสูงสุดสำหรับ σ^2 นั่นคือ $\hat{\sigma}_{MLE}^2 = \sum_{i=1}^n X_i^2 / 2n$ เนื่องจากทราบว่า $E(\hat{\sigma}_{MLE}^2) = \sigma^2$ และตัวแปรสุ่ม $Y = \sum_{i=1}^n X_i^2$ มีการแจกแจงแกมมา (Gamma distribution) ด้วยพารามิเตอร์ n และ $2\sigma^2$

หรือ $\text{Gamma}(n, 2\sigma^2)$ เมื่อ $\sigma^2 > 0$ และฟังก์ชันความหนาแน่นความน่าจะเป็นของ Y คือ

$$f_Y(y; \sigma) = \frac{1}{\Gamma(n)(2\sigma^2)^n} y^{n-1} e^{-y/(2\sigma^2)}$$

โดยที่ $y > 0$ และ n คือ ขนาดตัวอย่าง สำหรับ $E(Y) = 2n\sigma^2$ และ $\text{Var}(Y) = 4n\sigma^4$ ดังนั้น ความแปรปรวนของ $\hat{\sigma}_{MLE}^2$ จึงหาได้จาก

$$\text{Var}(\hat{\sigma}_{MLE}^2) = \text{Var}\left(\frac{1}{2n} \sum_{i=1}^n X_i^2\right) = \frac{\sigma^4}{n} \quad (5)$$

ในกรณีที่ตัวอย่างมีขนาดใหญ่ $\hat{\sigma}_{MLE}^2$ จะมีการแจกแจงปกติโดยประมาณด้วยค่าเฉลี่ย σ^2 และความแปรปรวน σ^4/n ดังนั้นช่วงความเชื่อมั่น $(1-\alpha) 100\%$ จากวิธีตัวอย่างขนาดใหญ่สำหรับ ψ ที่อาศัยตัวประมาณ $\hat{\sigma}_{MLE}^2$ คือ

$$CI_{MLE} = (l_{MLE}, u_{MLE}) = \left(\sqrt{\frac{\pi}{2}} \sqrt{L_{ML}}, \sqrt{\frac{\pi}{2}} \sqrt{U_{ML}}\right)$$

$$\text{เมื่อ } L_{ML} = \hat{\sigma}_{MLE}^2 - z_{1-\alpha/2} \sqrt{\text{var}(\hat{\sigma}_{MLE}^2)},$$

$U_{ML} = \hat{\sigma}_{MLE}^2 + z_{1-\alpha/2} \sqrt{\text{var}(\hat{\sigma}_{MLE}^2)}$ และ $\text{var}(\hat{\sigma}_{MLE}^2)$ คือ ตัวประมาณค่าความแปรปรวนของ $\hat{\sigma}_{MLE}^2$ ซึ่งหาได้จากการแทนค่า σ ในสมการ (5) ด้วย σ_{MLE}

4. ช่วงความเชื่อมั่นนัยทั่วไป

ช่วงความเชื่อมั่นนัยทั่วไปเป็นการสร้างช่วงความเชื่อมั่นสำหรับพารามิเตอร์ด้วยการหาเปอร์เซ็นต์ไทล์ของปริมาณหมุนนัยทั่วไป (Generalized pivotal quantity) ตามแนวคิดของ Weerahandi (1993) โดยปริมาณหมุนดังกล่าวมีสมบัติดังต่อไปนี้

1) ถ้าให้ $X=x$ แล้ว ฟังก์ชัน $R(X;x,\theta)$ ปราศจากพารามิเตอร์ที่ไม่ทราบค่า

2) ค่าสังเกตของ $R(X;x,\theta)$ หรือ $R(x;x,\theta)$ มีค่าเท่ากับ θ และไม่ขึ้นกับพารามิเตอร์รบกวน (Nuisance parameter) เมื่อ θ คือ พารามิเตอร์ทั่วไปในการแจกแจงความน่าจะเป็น

สำหรับการศึกษาค้นคว้าครั้งนี้ผู้วิจัยได้นำเสนอปริมาณหมุนนัยทั่วไปใหม่เพื่อใช้ในการสร้างช่วงความเชื่อมั่นสำหรับ ψ โดยมีข้อตกลงเบื้องต้นดังนี้ กำหนดให้ $X=(X_1, X_2, \dots, X_n)$ เป็นตัวอย่างสุ่มขนาด n จากการแจกแจง $Ra(\sigma)$ เมื่อ $\sigma > 0$ และเนื่องจาก $Y = \sum_{i=1}^n X_i^2$ มีการแจกแจง $\text{Gamma}(n, 2\sigma^2)$ ดังนั้นฟังก์ชันก่อกำเนิด

โมเมนต์ (Moment generating function) ของ $W = Y / \sigma^2$ จึงหาได้จาก

$$M_W(t) = E(e^{tW}) = M_Y\left(\frac{t}{\sigma^2}\right) = \frac{1}{(1-2t)^n}$$

ในที่นี้ W มีการแจกแจง $\text{Gamma}(n, 1/2)$ ซึ่งพบว่าการแจกแจงไม่ขึ้นอยู่กับการพารามิเตอร์และไม่มีพารามิเตอร์รบกวนเข้ามาเกี่ยวข้อง ดังนั้นจึงสามารถหาปริมาณพหุนัยทั่วไปสำหรับ σ^2 ได้จาก

$$R(X; x, \sigma^2) = \frac{Y}{W} \quad (6)$$

จาก (6) พบว่า 1) $R(X; x, \sigma^2)$ ไม่ขึ้นอยู่กับการพารามิเตอร์ และ 2) เมื่อแทนค่า $X = x$ แล้ว $R(x; x, \sigma^2) = \sigma^2$ ซึ่งเป็นไปตามเงื่อนไขของ Weerahandi (1993) ดังนั้นช่วงความเชื่อมั่นนัยทั่วไป $(1-\alpha) 100\%$ ของ σ^2 คือ

$$(R_{\alpha/2}(X; x, \sigma^2), R_{1-\alpha/2}(X; x, \sigma^2))$$

และจะสามารถหาช่วงความเชื่อมั่นนัยทั่วไป $(1-\alpha) 100\%$ สำหรับ ψ ได้จาก

$$CI_{GCI} = (l_{GCI}, u_{GCI}) \\ = \left(\sqrt{\pi/2} \sqrt{R_{\alpha/2}(X; x, \sigma^2)}, \sqrt{\pi/2} \sqrt{R_{1-\alpha/2}(X; x, \sigma^2)} \right)$$

โดยที่ $R_{\alpha/2}(X; x, \sigma^2)$ และ $R_{1-\alpha/2}(X; x, \sigma^2)$ คือ เปอร์เซ็นไทล์ที่ $(\alpha/2) \times 100$ และ $(1-\alpha/2) \times 100$ ของ $R(X; x, \sigma^2)$ ตามลำดับ

4. การจำลองและผลลัพธ์

การศึกษาประสิทธิภาพของช่วงความเชื่อมั่น CI_{NMLE} , CI_{Nmm} , CI_{MLE} และ CI_{GCI} ในครั้งนี้อาศัยการจำลองข้อมูลภายใต้สถานการณ์ต่าง ๆ โดยใช้โปรแกรม R (R Core Team, 2023) ผู้วิจัยกำหนดขนาดตัวอย่าง (n) เท่ากับ 10, 20, 30, 50, 100 และ 200 เพื่อแทนตัวอย่างขนาดเล็กไปจนถึงขนาดใหญ่ จำลองค่าของตัวแปรสุ่ม X จากการแจกแจงเรย์ลีย์ที่มีพารามิเตอร์ σ เท่ากับ 0.5, 1, 2 และ 3 จากรูปที่ 1 จะเห็นว่าลักษณะของข้อมูลจะแตกต่างกัน ทั้งนี้ในการสร้างเลขสุ่มผู้วิจัยได้อาศัยฟังก์ชัน *rrayleigh* จากโปรแกรม R (Wolodzko, 2022) และเมื่อคำนวณค่าเฉลี่ยประชากร (ψ) จึงมีค่าเท่ากับ 0.63, 1.25, 2.51 และ 3.76 จากนั้นจึงหาช่วงความเชื่อมั่นของ ψ จาก 4 วิธีโดยกำหนดระดับความเชื่อมั่น $(1-\alpha) 100\%$ เท่ากับ 0.95

การจำลองในแต่ละสถานการณ์ข้างต้น ผู้วิจัยกำหนดจำนวนการทำซ้ำ 10,000 รอบ แล้วคำนวณค่าความน่าจะเป็นคัมรวม (CP) และค่าความยาวช่วงโดยเฉลี่ย (EL) จาก

$$CP = \frac{1}{10,000} \sum_{i=1}^{10,000} (L_i \leq \psi \leq U_i) \quad \text{และ} \\ EL = \frac{1}{10,000} \sum_{i=1}^{10,000} (U_i - L_i)$$

โดยที่ $\sum_{i=1}^{10,000} (L_i \leq \psi \leq U_i)$ คือ ผลรวมของจำนวนครั้งที่ค่าขอบเขตล่าง L_i และขอบเขตบน U_i ในการทำซ้ำรอบที่ i ครอบคลุมค่าพารามิเตอร์ ψ และ $U_i - L_i$ คือความยาวของช่วงความเชื่อมั่นในการทำซ้ำรอบที่ i สำหรับช่วงความเชื่อมั่นที่มีประสิทธิภาพดีนั้นจะต้องมีค่าความน่าจะเป็นคัมรวมใกล้เคียงกับระดับความเชื่อมั่นที่กำหนด คือ $1-\alpha = 0.95$ และมีค่าความยาวช่วงโดยเฉลี่ยน้อยในแต่ละสถานการณ์ ทั้งนี้ผู้วิจัยได้พิจารณาขอบเขตที่ค่าความน่าจะเป็นคัมรวมจากการจำลองไม่แตกต่างจาก 0.95 อย่างมีนัยสำคัญทางสถิติจาก

$$(1-\alpha) \pm z_{1-\alpha} \sqrt{\frac{\alpha(1-\alpha)}{10,000}} = 0.95 \pm 1.645 \sqrt{\frac{0.05(0.95)}{10,000}} = \\ (0.9464, 0.9536)$$

ช่วงความเชื่อมั่นที่สอดคล้องกับเกณฑ์ดังกล่าวจะแสดงให้เห็นว่าเป็นวิธีที่สามารถประมาณค่าพารามิเตอร์ได้ถูกต้องและแม่นยำ

ตารางที่ 1 แสดงค่าความน่าจะเป็นคัมรวมและค่าความยาวเฉลี่ยของช่วงความเชื่อมั่น 95% สำหรับ ψ จากการจำลอง ซึ่งผลสรุปที่ได้มีดังนี้

1) จากการศึกษาพบว่าค่าความน่าจะเป็นคัมรวมของ CI_{NMLE} ไม่เป็นไปตามระดับความเชื่อมั่นที่กำหนด กล่าวคือมีค่าสูงกว่า 0.95 อย่างมากในทุกกรณีที่ทำการศึกษา ส่วน CI_{MLE} และ CI_{Nmm} มีค่าความน่าจะเป็นคัมรวมใกล้เคียง 0.95 เฉพาะในกรณีที่ $n \geq 50$ และ $n \geq 30$ ตามลำดับ เท่านั้น

2) CI_{GCI} ให้ค่าความน่าจะเป็นคัมรวมใกล้เคียง 0.95 ในทุกกรณีที่ทำการศึกษา โดยค่าความน่าจะเป็นคัมรวมที่ได้นั้นมีความคงเส้นคงวา (Consistency) กล่าวคือเมื่อ n , σ หรือ ψ มีค่าเปลี่ยนไป ความน่าจะเป็นคัมรวมของ CI_{GCI} ยังคงมีค่าใกล้เคียง 0.95 เสมอ

ตารางที่ 1 ค่าความน่าจะเป็นคัมรวมและค่าความยาวช่วงโดยเฉลี่ยของช่วงความเชื่อมั่น 95% สำหรับค่าเฉลี่ยประชากรในการแจกแจงเรย์ลีย์ ที่ได้จากการจำลอง

n	พารามิเตอร์		ความน่าจะเป็นคัมรวม			ความยาวช่วงโดยเฉลี่ย				
	σ	ψ	NMLE	MLE	Nmm	GCI	NMLE	MLE	Nmm	GCI
10	0.5	0.63	0.9985	0.9024	0.9347	0.9511	0.9964	0.4062	0.4064	0.4200
	1	1.25	0.9989	0.9018	0.9325	0.9463	1.9895	0.8111	0.8116	0.8385
	2	2.51	0.9987	0.9060	0.9356	0.9486	3.9952	1.6288	1.6289	1.6837
	3	3.76	0.9979	0.9040	0.9342	0.9515	5.9662	2.4324	2.4324	2.5142
20	0.5	0.63	1.0000	0.9287	0.9439	0.9513	1.0015	0.2799	0.2869	0.2849
	1	1.25	1.0000	0.9217	0.9388	0.9497	2.0036	0.5600	0.5738	0.5702
	2	2.51	1.0000	0.9253	0.9387	0.9467	4.0133	1.1217	1.1501	1.1426
	3	3.76	1.0000	0.9258	0.9384	0.9470	6.0158	1.6813	1.7232	1.7123
30	0.5	0.63	1.0000	0.9350	0.9440	0.9525	1.0060	0.2275	0.2347	0.2303
	1	1.25	1.0000	0.9309	0.9439	0.9464	2.0066	0.4537	0.4684	0.4593
	2	2.51	1.0000	0.9279	0.9414	0.9463	4.0146	0.9078	0.9369	0.9189
	3	3.76	1.0000	0.9371	0.9447	0.9491	6.0267	1.3628	1.4068	1.3793
50	0.5	0.63	1.0000	0.9418	0.9450	0.9489	1.0068	0.1751	0.1817	0.1764
	1	1.25	1.0000	0.9437	0.9509	0.9544	2.0127	0.3501	0.3633	0.3527
	2	2.51	1.0000	0.9406	0.9465	0.9495	4.0263	0.7004	0.7268	0.7054
	3	3.76	1.0000	0.9392	0.9445	0.9483	6.0337	1.0495	1.0891	1.0572
100	0.5	0.63	1.0000	0.9472	0.9470	0.9510	1.0078	0.1233	0.1285	0.1237
	1	1.25	1.0000	0.9450	0.9486	0.9492	2.0146	0.2466	0.2568	0.2473
	2	2.51	1.0000	0.9476	0.9529	0.9515	4.0291	0.4931	0.5136	0.4946
	3	3.76	1.0000	0.9420	0.9469	0.9487	6.0393	0.7391	0.7697	0.7413
200	0.5	0.63	1.0000	0.9486	0.9501	0.9496	1.0077	0.0870	0.0908	0.0871
	1	1.25	1.0000	0.9480	0.9507	0.9494	2.0165	0.1741	0.1817	0.1743
	2	2.51	1.0000	0.9480	0.9494	0.9478	4.0329	0.3481	0.3633	0.3485
	3	3.76	1.0000	0.9503	0.9489	0.9523	6.0432	0.5217	0.5444	0.5223

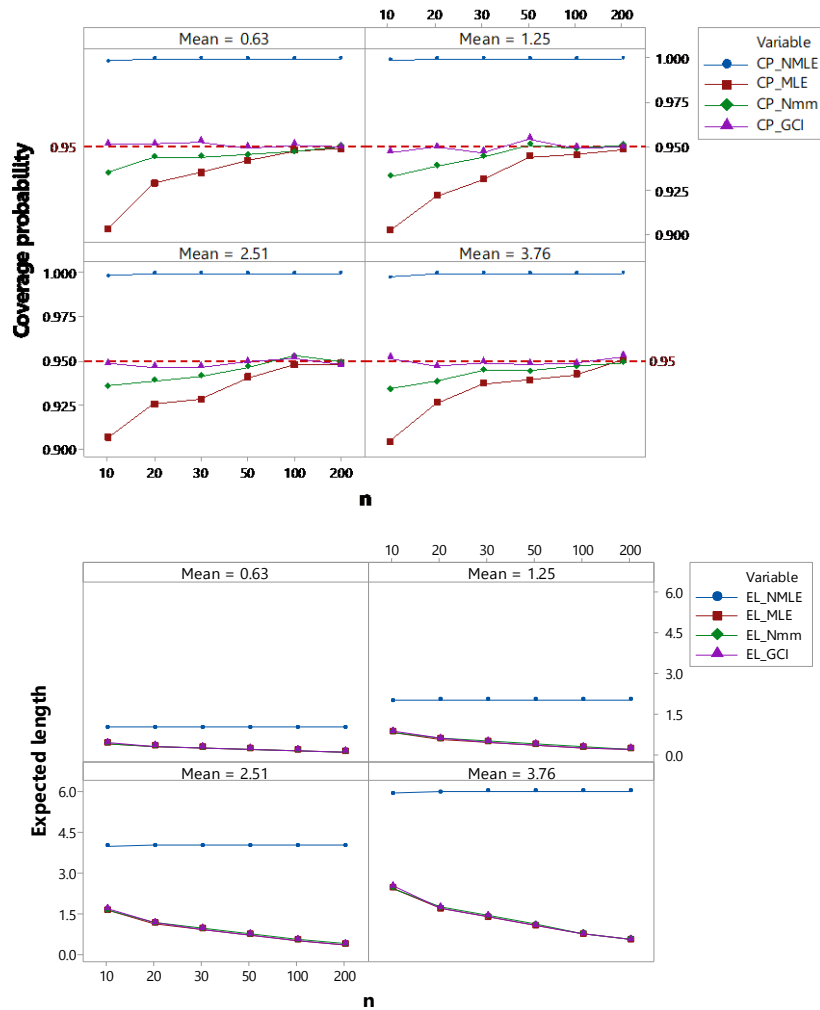
3) เมื่อพิจารณาประสิทธิภาพของตัวประมาณในเรื่องความยาวเฉลี่ยของช่วงความเชื่อมั่นพบว่า CI_{GCI} , CI_{Nmm} และ CI_{MLE} มีค่าความยาวช่วงโดยเฉลี่ยไม่แตกต่างกันมากนัก ในขณะที่ CI_{NMLE} ที่มีค่าความยาวช่วงโดยเฉลี่ยมากที่สุดเมื่อเทียบกับช่วงความเชื่อมั่นจากวิธีการอื่น ๆ อย่างไรก็ตามจะเห็นได้ว่าความยาวช่วงโดยเฉลี่ยที่ได้จากแต่ละวิธีจะมีค่าน้อยลง เมื่อ n มีค่าเพิ่มขึ้น และจะแปรผันตามค่าของ ψ

จากผลการจำลองจึงสรุปได้ว่า CI_{GCI} เป็นช่วงความเชื่อมั่นที่มีประสิทธิภาพสูงที่สุดในการศึกษาครั้งนี้ เนื่องจากสามารถประมาณค่าเฉลี่ยประชากรในการแจกแจงเรย์ลีย์ได้ด้วยค่าความน่าจะเป็นที่กำหนด และมีความยาวช่วงที่เหมาะสมในทุกสถานการณ์ดังแสดงในรูปที่ 2 ดังนั้นผู้วิจัยจึงเสนอให้ใช้ CI_{GCI} ในทางปฏิบัติสำหรับกรณีที่ตัวอย่างมีขนาดมากกว่า 30 หน่วย CI_{MLE} สามารถใช้เป็นวิธีทางเลือกในการประมาณค่าเฉลี่ยประชากรของ

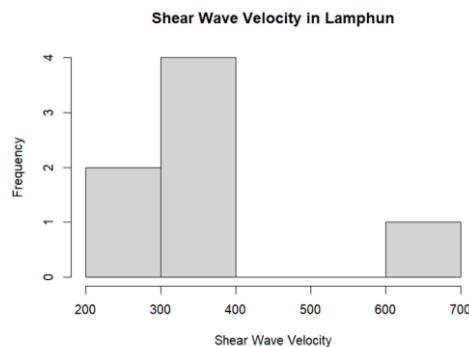
การแจกแจงเรย์ลีย์ได้ เนื่องจากเป็นวิธีที่ให้ค่าความน่าจะเป็นคัมรวมที่ยอมรับได้และมีความยาวช่วงสั้น นอกจากนั้น CI_{MLE} เป็นวิธีที่ไม่ต้องอาศัยโปรแกรมคอมพิวเตอร์ช่วยในการคำนวณ เนื่องจากมีสูตรปิดในการหาช่วงความเชื่อมั่น เชื่อมั่น 95% สำหรับสัดส่วนของผู้โดยสารที่ติดเชื้อไวรัส

การประยุกต์ใช้ข้อมูลจริง

ความเร็วคลื่นเฉือนของดิน (Shear-wave velocity of soils) คือ สมบัติของดินที่สามารถขยายแรงแผ่นดินไหวให้มีความรุนแรงมากขึ้นได้ ในงานวิศวกรรมธรณีเทคนิค จึงนิยมสำรวจและวัดค่าดังกล่าวมาทำการศึกษาแหล่งกำเนิดคลื่นแผ่นดินไหว ในปัจจุบันความเร็วคลื่นเฉือนของดินสามารถหาได้จากการลงภาคสนาม



รูปที่ 2 กราฟแสดงค่าความน่าจะเป็นคุ่มรวม (บน) และค่าความยาวโดยเฉลี่ย (ล่าง) ของช่วงความเชื่อมั่น 95% สำหรับค่าเฉลี่ยประชากรในการแจกแจงเรย์ลีย์ ที่ได้จากการจำลอง



รูปที่ 3 ฮิสโตแกรมของข้อมูลความเร็วคลื่นเฉือนของดิน (หน่วย: เมตร/วินาที) ในจังหวัดลำพูน

(วัดในหลุมเจาะสำรวจ) และสำรวจคลื่นไหวสะเทือน พื้นผิวแบบหลายช่องรับสัญญาณ (Multi-channel Analysis of Surface Wave: MASW) วิธีหลังนี้เป็นวิธีใหม่ที่มีราคาไม่สูงมากนัก และสามารถทำได้รวดเร็วกว่าวิธีแรก (ฐานบ ธิติมากร และคณะ, 2551) อย่างไรก็ตามถึงแม้ประเทศไทยจะไม่ได้รับผลกระทบโดยตรงจากแผ่นดินไหวในระดับรุนแรงมากนัก แต่ก็มีก่เกิดแผ่นดินไหวที่ประชาชนสามารถรับรู้ถึงแรงสั่นสะเทือนได้

จากข้อมูลสถิติพบว่าตั้งแต่ พ.ศ. 2550 ถึง พ.ศ. 2559 อัตราการเกิดแผ่นดินไหวในประเทศไทยมีมากขึ้น (โชติ เนื่องนันท์ และคณะ, 2561) การเกิดแผ่นดินไหวเป็นปรากฏการณ์ทางธรรมชาติที่ไม่สามารถคาดเดาการเกิดล่วงหน้าได้ แต่ด้วยวิธีการที่ทันสมัยจึงทำให้นักวิจัยทราบแนวโน้มในการเกิดเหตุการณ์ดังกล่าวได้เช่นกัน และหนึ่งในนั้นคือการวิเคราะห์คลื่นพื้นผิวจากค่าความเร็วคลื่นเฉือน

ตารางที่ 2 การตรวจสอบความเหมาะสมของตัวแบบการแจกแจงความน่าจะเป็นของข้อมูลตัวอย่าง

เกณฑ์/การแจกแจง	การแจกแจงเรย์ลีย์	การแจกแจงแบบเลขชี้กำลัง	การแจกแจงโคกำลังสอง
AIC	90.6010	98.5239	154.0379
BIC	90.5469	98.4699	153.9838

ตารางที่ 3 ช่วงความเชื่อมั่น 95% สำหรับค่าเฉลี่ยของความเร็วคลื่นเฉือนของดิน (เมตร/วินาที) ในจังหวัดลำพูน

วิธี	ขีดจำกัดล่าง	ขีดจำกัดบน	ค่าความยาวช่วง
CI_{NMLE}	65.7395	607.3215	541.5819
CI_{MLE}	171.3344	444.0158	272.6814
CI_{Nmm}	222.4714	503.6477	281.1763
CI_{GCI}	244.0206	525.4918	281.4712

การศึกษาในครั้งนี้ผู้วิจัยได้นำค่าความเร็วคลื่นเฉือนของดินในจังหวัดลำพูน ซึ่งเป็นจังหวัดที่มีพื้นที่ใกล้เคียงมีรอยเลื่อนที่มีพลังหลายรอยเลื่อน และมีประวัติของการเกิดแผ่นดินไหวในพื้นที่มาทำการศึกษา ข้อมูลที่ใช้ในครั้งนี้ถูกเก็บรวบรวมโดย ธนบดี แร่นาค (2557) จากการตรวจวัดด้วยเครื่อง MASW แบบ 24 ช่องสัญญาณ ค่าความเร็วคลื่นเฉือนที่วัดได้จาก 7 สถานีที่ระดับความลึก 30 เมตร (เมตร/วินาที) แสดงในรูปที่ 3 เมื่อคำนวณค่าสถิติเบื้องต้นแล้วพบว่า ค่าเฉลี่ยตัวอย่าง $\bar{X}$ เท่ากับ 363.0596 ค่ามัธยฐานเท่ากับ 344.9510 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 120.2079 เมตร/วินาที อย่างไรก็ตามเพื่อให้การประมาณค่าสอดคล้องกับการแจกแจงที่แท้จริงของข้อมูล ผู้วิจัยจะทำการตรวจสอบความเหมาะสมของตัวแบบความน่าจะเป็น (Model selection) สำหรับข้อมูลชุดนี้โดยอาศัยเกณฑ์ข้อสนเทศของอะกะอิเกะ (Akaike information criterion: AIC) และเกณฑ์ข้อสนเทศของเบส์ (Bayesian information criterion: BIC) ก่อน จากการทบทวนวรรณกรรมพบว่าความเร็วคลื่นเฉือนของดินอาจมีการแจกแจงเรย์ลีย์ การแจกแจงแบบเลขชี้กำลัง และการแจกแจงโคกำลังสองได้ (Pasari & Dikshit, 2015) ดังนั้นเมื่อพิจารณาค่า AIC และ BIC ของการแจกแจงต่าง ๆ ในตารางที่ 2 แล้วพบว่า การแจกแจงเรย์ลีย์มีค่า AIC และ BIC ต่ำที่สุด ดังนั้นจึงเป็นการแจกแจงที่เหมาะสมที่สุดสำหรับข้อมูลชุดนี้ จากนั้นผู้วิจัยได้ทำการทดสอบความมีนัยสำคัญของการแจกแจงโดยใช้สถิติทดสอบคอลโมโกรอฟ-สมิรโนฟ (Kolmogorov-Smirnov test) ภายใต้สมมติฐาน คือ

H_0 : ความเร็วคลื่นเฉือนของดินมีการแจกแจงเรย์ลีย์

H_1 : ความเร็วคลื่นเฉือนของดินไม่มีการแจกแจงเรย์ลีย์

ที่ระดับนัยสำคัญ 0.05 พบว่าความเร็วคลื่นเฉือนของดินที่ทำการศึกษาในครั้งนี้มีการแจกแจงเรย์ลีย์ (ค่าสถิติเท่ากับ

0.3983 และค่า p-value เท่ากับ 0.1630) ข้อมูลชุดนี้จึงสามารถนำไปใช้ในการประมาณค่าเฉลี่ยประชากรจาก CI_{NMLE} , CI_{Nmm} , CI_{MLE} และ CI_{GCI} ได้ผลที่ได้จากตารางที่ 3 พบว่า CI_{MLE} , CI_{Nmm} และ CI_{GCI} ให้ค่าความยาวช่วงความเชื่อมั่นไม่แตกต่างกันมากนัก ในขณะที่ CI_{NMLE} มีความยาวช่วงมากที่สุด ซึ่งสอดคล้องกับผลที่ได้จากการจำลอง แต่อย่างไรก็ตามผลจากการจำลองแสดงให้เห็นชัดเจนว่าเมื่อตัวอย่างมีขนาดเล็ก CI_{GCI} ให้ประสิทธิภาพดีที่สุด ดังนั้นจึงสรุปได้ว่าค่าเฉลี่ยของความเร็วคลื่นเฉือนของดินในจังหวัดลำพูนมีค่าอยู่ระหว่าง 244.0206 ถึง 525.4918 เมตร/วินาที และเมื่อพิจารณาตามเกณฑ์ National Earthquake Hazards Reduction Program (NEHRP) แล้วพบว่าชั้นดินในบริเวณและระยะเวลาที่นำข้อมูลมาทำการศึกษาจัดอยู่ในประเภทดินแข็ง (Stiff soil) ค่อนข้างแน่นมากและหินผุ (Dense soil and soft rock) ซึ่งอาจเกิดแผ่นดินไหวได้มาก

สรุปและอภิปรายผลการศึกษา

ในการศึกษาครั้งนี้ผู้วิจัยได้นำเสนอช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยประชากรในการแจกแจงเรย์ลีย์ 4 วิธี คือ วิธีของวาลด์ วิธีช่วงความเชื่อมั่นสำหรับตัวอย่างขนาดใหญ่ที่อาศัยตัวประมาณโมเมนต์ วิธีช่วงความเชื่อมั่นสำหรับตัวอย่างขนาดใหญ่ที่อาศัย $\hat{\sigma}_{MLE}^2$ และวิธีช่วงความเชื่อมั่นนัยทั่วไป จากการจำลองภายใต้สถานการณ์ต่าง ๆ พบว่าช่วงความเชื่อมั่นนัยทั่วไปมีประสิทธิภาพดีที่สุด กล่าวคือเป็นวิธีที่ให้ค่าความน่าจะเป็นคุ้มครองใกล้เคียงกับระดับความเชื่อมั่นที่กำหนด ทั้งนี้เป็นผลเนื่องมาจากวิธีนี้ได้ถูกพิจารณาภายใต้การแจกแจงที่แท้จริงของปริมาณหมุน (ไม่ได้อาศัยการแจกแจงปกติโดยประมาณอย่างวิธีการอื่น) ในทางปฏิบัติจึงสามารถนำวิธีช่วงความเชื่อมั่นนัยทั่วไป (CI_{GCI}) ไปใช้ได้กับทุกขนาดตัวอย่าง ในขณะที่ช่วงความเชื่อมั่นสำหรับตัวอย่างขนาดใหญ่นั้นได้

ถูกสร้างขึ้นมาโดยอาศัยทฤษฎีที่จำกัดส่วนกลางจึงทำให้มีประสิทธิภาพดีเมื่อตัวอย่างมีขนาดใหญ่พอ ซึ่งการศึกษานี้แนะนำให้ใช้เมื่อตัวอย่างมีขนาดมากกว่า 50 หน่วย ช่วงความเชื่อมั่นนี้ทั่วไปนอกจากจะมีประสิทธิภาพที่ดีกว่าช่วงความเชื่อมั่นที่นำมาเปรียบเทียบในด้านความน่าจะเป็นคุ่มรวมแล้ว ยังมีความยาวของช่วงสั้นซึ่งแสดงให้เห็นว่าสามารถประมาณค่าเฉลี่ยได้ใกล้เคียงกับค่าจริง และวิธีนี้สามารถใช้โปรแกรมคอมพิวเตอร์ช่วยในการหาขีดจำกัดบนและขีดจำกัดล่างได้สะดวก การศึกษาในครั้งนี้ได้นำวิธีที่นำเสนอในทางทฤษฎีมาประยุกต์ใช้กับข้อมูลจริง ซึ่งผลที่ได้พบว่ามีความสอดคล้องกับผลที่ได้จากการจำลอง

กิตติกรรมประกาศ

ผู้วิจัยขอขอบคุณบรรณาธิการและกรรมการที่ได้ให้ข้อเสนอแนะสำหรับบทความวิจัยฉบับนี้

เอกสารอ้างอิง

- Ahmed, A. A., Ahmad, S. P., & Reshi, J. A. (2013). Bayesian analysis of Rayleigh distribution. *International Journal of Scientific and Research Publications*, 3(10), 1-9.
- Ahn, C. Y., Jung, Y. M., Kwon, O. I., & Seo, J. K. (2012). Fast segmentation of ultrasound images using robust Rayleigh distribution decomposition. *Pattern Recognition*, 45(9), 3490-3500.
- Ateeq, K. Qasim, T. B., & Alvi, A. R. (2019). An extension of Rayleigh distribution and applications. *Cogent Mathematics & Statistics*, 6(1622191), 1-16.
- Casella, G., & Berger, R. L. (2002). *Statistical Inference*. California, USA: Duxbury Press.
- Dey, S. (2009). Comparison of Bayes estimators of the parameter and reliability function for Rayleigh distribution under different loss functions. *Malaysian Journal of Mathematical Sciences*, 3(2), 249-266.
- Dey, S., & Dey, T. (2014). Statistical inference for the Rayleigh distribution under progressively type-II censoring with binomial removal. *Applied Mathematical Modelling*, 38(3), 974-982.
- Abd Elfattah, A. M., Hassan, A. S., & Ziedan, D. M. (2006). Efficiency of maximum likelihood estimators under different censored sampling schemes for Rayleigh distribution. *Interstat*, 1, 1-16.
- Islam, D. K., Chaichana, T., Dussadee, N., & Intaniwet, A. (2017). Statistical distribution and energy estimation of the wind speed at Saint Martin's Island, Bangladesh. *International Journal of Renewable Energy*, 15(1), 77-88.
- Kruiver, P. P., Dedem, E. V., Romijn, R., Lange, G. D., Korff, M., Stafleu, J., Gunnink, J. L., Marek, A. R., Bommer, J. J., Elk, J. A., & Doornhof, D. (2017). An integrated shear-wave velocity model for the Groningen gas field, The Netherlands. *Bulletin of Earthquake Engineering*, 15(9), 3555-3580.
- Liebenberg, S. C., & Allison, S. J. (2023). A review of goodness-of-fit tests for the Rayleigh distribution. *Austrian Journal of Statistics*, 1-22.
- Nar, F., Okman, O. E., Özgür, A., & Çetin, M. (2018). Fast target detection in radar images using Rayleigh mixtures and summed area tables. *Digital Signal Processing*, 77(2018), 86-101.
- Pasari, S., & Dikshit, O. (2015). Earthquake interevent time distribution in Kachchh, Northwestern India. *Earth, Planets and Space*, 67(1), 1-17.
- Pishgar-Komleh, S. H., Keyhani, A., & Sefeedpari, P. (2015). Wind speed and power density analysis based on Weibull and Rayleigh distributions (a case study: Firouzkooch county of Iran). *Renewable and Sustainable Energy Reviews*, 42, 313-322.
- R Core Team. (2023). R: A Language and Environment for Statistical Computing. Vienna, Austria.
- Sarti, A., Corsi, C., Mazzini, E., & Lamberti, C. (2005). Maximum likelihood segmentation of ultrasound images with Rayleigh distribution. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 52(6), 947-960.
- Seo, J. I., Jeon, J. W., & Kang, S. B. (2016). Exact interval inference for the two-parameter Rayleigh distribution based on the upper record values. *Journal of Probability and Statistics*, 2016, 1-5.
- Weerahandi, S. (1993). Generalized confidence intervals. *Journal of the American Statistical Association*, 88(423), 899-905.
- Wolodko, T. (2020). *Additional Univariate and Multivariate Distributions*. <https://cran.r-project.org/web/packages/extraDistr/extraDistr.pdf>
- Xu, W., Ning, L., & Luo, Y. (2020). Wind speed forecast based on post-processing of numerical weather predictions using a gradient boosting decision tree algorithm. *Atmosphere*, 11(7), 1-38.
- Nungun, C., Phungkhom, N., Srikongruk, S., Mameechai, D., Thongperm, P., & Noolek, W. (2561). Trends in earthquakes in Thailand and Burma. *Journal of Thaksin University*, 21(3), 168-173. (in Thai)
- Thitimakorn, T., Luang-Sunthornkul, N., Junthamas, P., & Hansiri, A. (2551). Determining the shear wave velocity of the Bangkok soil layer From surface wave analysis. *Substance Engineering, Research and Development Edition*, 19(4), 30-35. (in Thai)
- Raenak, T. (2557). Determination of soil shear wave velocity using an MASW surface wave receiver belt. Mueang District area Lamphun Province [Unpublished Bachelor thesis]. Chulalongkorn University. (in Thai)

การเปรียบเทียบตัวแบบการพยากรณ์ราคาหุ้นโดยใช้ตัวแบบอาร์มาและจีวีเอ็ม

A Comparison of Forecasting Models of Stock Price using

ARIMA and GVM Models

บุญกอง ทะกลโยธิน

Boonkong Dhakonlayodhin

ภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์ประยุกต์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ ประเทศไทย

Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Thailand

บทคัดย่อ

การศึกษานี้มีวัตถุประสงค์เพื่อสร้างตัวแบบพยากรณ์ราคาหุ้น โดยศึกษาตัวแบบอาร์มาและตัวแบบเกรย์ ภายใต้ตัวแบบ GM(1,1) และตัวแบบเกรย์เวอร์เฮลท์ (Grey Verhulst Model: GVM) สำหรับข้อมูลราคาหุ้น BBL รายวัน ตั้งแต่เดือนกรกฎาคม 2565 ถึงเดือนมิถุนายน 2566 จำนวน 240 วัน ผู้วิจัยได้แบ่งข้อมูลออกเป็น 2 ชุด ชุดที่ 1 ตั้งแต่เดือนกรกฎาคม 2565 ถึงเดือนพฤษภาคม 2566 จำนวน 222 วัน สำหรับการสร้างตัวแบบพยากรณ์ด้วยวิธีการทางสถิติทั้งหมด 3 วิธี ได้แก่ ตัวแบบอาร์มา และตัวแบบเกรย์ภายใต้ตัวแบบ GM(1,1) และตัวแบบเกรย์เวอร์เฮลท์ ข้อมูลชุดที่ 2 เดือนมิถุนายน 2566 จำนวน 18 วัน นำมาใช้ในการเปรียบเทียบความแม่นยำของตัวแบบพยากรณ์ด้วยเกณฑ์ร้อยละค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย และเกณฑ์รากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ที่ต่ำที่สุด ผลการศึกษาพบว่า ตัวแบบที่มีความแม่นยำมากที่สุด คือ ตัวแบบ GM(1,1) มีความผิดพลาดในการพยากรณ์ร้อยละ 0.724 (MAPE = 0.724) หรือมีความผิดพลาดจากการพยากรณ์ 1.463 (RMSE = 1.463) ซึ่งทั้ง 2 เกณฑ์ที่ใช้ในการเปรียบเทียบความถูกต้องของตัวแบบพยากรณ์ให้ผลเป็นไปในทิศทางเดียวกัน จึงทำให้ความน่าเชื่อถือได้มากยิ่งขึ้นว่าตัวแบบ GM(1,1) เป็นวิธีที่มีความเหมาะสมกับอนุกรมเวลาชุดนี้มากที่สุด

คำสำคัญ: ตัวแบบอาร์มา, ตัวแบบเกรย์, ตัวแบบเกรย์เวอร์เฮลท์, ร้อยละของค่าความคลาดเคลื่อนสัมบูรณ์ค่าเฉลี่ย, รากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ย

ABSTRACT

The objective of this study is to construct an appropriate forecasting model for the BBL Stock Price via the use of Box-Jenkins method and Grey models. Daily average data, which were collected during July 2022 to June 2023 with 240 days in total, were divided into 2 datasets. The first dataset, which consisted of 222 days from July 2022 to May 2023, was used for constructing the forecasting model via the use of 3 statistical methods, which are Box-Jenkins, GM(1,1), and Grey Verhulst Model (GVM) methods. The second dataset, which consisted of 18 days in June 2023, was used for comparing the accuracy of the forecasting model. The criteria for comparing the accuracy were the mean absolute percentage error (MAPE) and root mean square error (RMSE). Research findings indicated that for all three forecasting methods, the most accurate method was the GM(1,1) method with 0.724 percent forecasting error (MAPE = 0.724) or forecasting error of 1.463 (RMSE = 1.463). The comparison of the accuracy of the forecasting model for both criteria provided the same result. Therefore, it made more reliable that the GM (1,1) method was the most suitable method for this time series.

KEYWORDS: ARIMA model; Grey Model; GVM Model; Mean Absolute Percentage Error; Root Mean Square Error

*Corresponding Author: boonkong.d@sci.kmutnb.ac.th

Received: 14/07/2023; Revised: 06/08/2023; Accepted: 08/08/2023

1. บทนำ

ในยุคปัจจุบันนี้การลงทุนในตลาดหลักทรัพย์ของประเทศไทยเป็นสิ่งที่น่าสนใจเนื่องจากการลงทุนในตลาดหลักทรัพย์สามารถให้ผลตอบแทนที่สูงกว่าการฝากเงินไว้กับธนาคารแต่เนื่องจากราคาหลักทรัพย์มีความผันผวนไม่แน่นอนอันเกิดมาจากสาเหตุหลายปัจจัยและนำไปสู่ความเสี่ยงในการลงทุนโดยเฉพาะอย่างยิ่งเมื่อราคาคงต่ำลงมากกว่าราคาเริ่มแรกที่นักลงทุนเข้าซื้อแม้หุ้นจะเป็นหลักทรัพย์ที่ผู้ลงทุนหลายคนลังเลที่จะลงทุน เพราะไม่แน่ใจในความเสี่ยงที่อาจจะเกิดขึ้น แต่คงปฏิเสธไม่ได้ว่าหุ้นยังคงมีเสน่ห์สำหรับผู้ลงทุนอีกหลายคนที่ไม่ว่าจะเสี่ยงแค่ไหนก็ต้องมีหุ้นอยู่ในพอร์ตการลงทุนเสมอขึ้นเป็นเพราะการลงทุนในหุ้นมีโอกาสได้รับผลตอบแทนที่น่าสนใจ ทั้งในรูปของเงินปันผล (Dividend) และกำไรจากการขายหุ้น (Capital Gain) อย่างไรก็ตามการลดความเสี่ยงจากการสูญเสียเงินลงทุนที่หายไปในตลาดหลักทรัพย์สามารถหลีกเลี่ยงได้หากนำผลการวิเคราะห์ในเชิงเทคนิคเข้ามาช่วยในการตัดสินใจลงทุนเข้าซื้อขายหลักทรัพย์โดยการเลือกเครื่องมือทางสถิติที่เหมาะสมกับทฤษฎีทางการเงินเพื่อเสริมสร้างหรือหาจังหวะในการเข้าซื้อขายหุ้นในตลาดหลักทรัพย์ของประเทศไทยอันนำไปสู่การลดความเสี่ยงของการขาดทุนที่จะเกิดขึ้นแก่ผู้ลงทุน การพยากรณ์เป็นการคาดคะเนของเหตุการณ์ในอนาคตโดยอาศัยรูปแบบของการเกิดเหตุการณ์หรือการพยากรณ์ที่เก็บข้อมูลจากอดีตรวมถึงความรู้ความสามารถของผู้พยากรณ์ หากทราบเหตุการณ์ต่าง ๆ ในอนาคตมีความเป็นไปได้ที่จะเพิ่มความเชื่อมั่นให้กับการวางแผนที่มีความถูกต้องและผิดพลาดน้อยที่สุดซึ่งมีความสอดคล้องกับการคาดการณ์ราคาหุ้น สำหรับหุ้นที่น่าสนใจ คือ BBL หรือ ธนาคารกรุงเทพ จำกัด (มหาชน) ซึ่งดำเนินธุรกิจธนาคารพาณิชย์ที่ให้บริการทางการเงินแก่ธุรกิจทั้งขนาดใหญ่ และ SME โดยมีสาขาในทุกพื้นที่ทั่วประเทศ และมีเครือข่ายในต่างประเทศครอบคลุมเขตเศรษฐกิจสำคัญของโลก

การพยากรณ์ คือ การคาดคะเนเหตุการณ์ในอนาคตโดยการศึกษาจากแนวโน้ม และรูปแบบของการเกิดในอดีต ปัจจุบันการพยากรณ์นั้นมีบทบาทสำคัญในการวางแผนและการตัดสินใจในการดำเนินงานในด้านต่างๆ เป็นอย่างมาก ซึ่งหากเราสามารถทราบได้ว่าเหตุการณ์ในอนาคตนั้น มีความเป็นไปได้ในทิศทางใด ก็จะทำให้การวางแผนดำเนินงานมีความถูกต้อง และความแม่นยำ แต่เหตุการณ์ต่าง ๆ ในอนาคตที่จะเกิดขึ้นนั้นอยู่ภายใต้

ความไม่แน่นอน ดังนั้นการพยากรณ์ที่ให้ความแม่นยำสูงจึงมีความสำคัญอย่างยิ่งต่อการวางแผนและการตัดสินใจดังกล่าว ในปัจจุบันได้มีผู้นำเสนอวิธีการพยากรณ์ไว้หลากหลายวิธี แต่การจะเลือกใช้วิธีการพยากรณ์แบบใดนั้น ขึ้นอยู่กับลักษณะของข้อมูล และพิจารณาของข้อมูลพยากรณ์ด้วย โดยพิจารณาจากวิธีการพยากรณ์ที่ให้ความคลาดเคลื่อนที่ต่ำสุด หรือให้ความถูกต้อง แม่นยำสูงที่สุด สำหรับการพยากรณ์อนุกรมเวลา (Time Series forecasting) มีอยู่ด้วยกันหลายวิธี วิธีที่นิยมใช้กันมาก คือ วิธีของ Box – Jenkins ด้วยตัวแบบอาร์มา (Autoregressive Integrated Moving Average Model: ARIMA) ซึ่งเป็นวิธีหาตัวแบบพยากรณ์โดยอาศัยความสัมพันธ์ของข้อมูลในอดีตเพื่อหาตัวแบบที่แสดงพฤติกรรมของข้อมูล และใช้เป็นแนวทางในการพยากรณ์ค่าในอนาคต ในบางครั้งเราจะพบว่าข้อมูลอนุกรมเวลาอาจมีความผันผวน มีความไม่แน่นอน ทำให้เกิดปัญหาในการตัดสินใจ อันเกิดจากธรรมชาติของข้อมูลที่เกิดขึ้นจริงในการดำเนินชีวิต โดยแบ่งข้อมูลออกเป็น 3 ระบบ คือ ระบบขาว (White System) ระบบดำ (Black System) และระบบเทาหรือระบบเกรย์ (Grey System) [1] โดยที่ระบบขาวเป็นกรณีที่ข้อมูลที่สมบูรณ์ชัดเจน ระบบดำเป็นกรณีที่ระบบข้อมูลที่ไม่มีข้อมูลชัดเจน (Black Box) ส่วนข้อมูลระบบเทาที่มีความคลุมเครือ มีข้อมูลบางส่วนที่ชัดเจนและบางส่วนไม่ชัดเจน อาจส่งผลต่อความแม่นยำของค่าพยากรณ์ได้ วิธีเกรย์สามารถใช้ในการทำนายข้อมูลที่มีจำนวนข้อมูลน้อย การทำนายด้วยวิธีเกรย์สามารถนำมาใช้ในหลายแนวทาง เช่น การทำนายภาวะทางเศรษฐกิจ เชิงการผลิต การตลาด การวางแผนทางธุรกิจ การทำนายสถานการณ์ทางสังคม การทำนายทรัพยากรธรรมชาติ และอื่นๆ [2, 3, 4, 5] วิธีเกรย์มีความสามารถในการทำนาย แม้ว่าจะมีข้อมูลที่ไม่เพียงพอหรือข้อมูลที่ไม่สมบูรณ์ [6,7] โดยการใช้ข้อมูลที่มีความเปลี่ยนแปลงน้อย (Grey Data) ในการประมาณการ วิธีนี้สามารถใช้กับข้อมูลที่มีลักษณะแบบเวลาสั้น หรือการเก็บข้อมูลที่ยากที่จะเป็นไปตามกระบวนการทางคณิตศาสตร์อื่น ๆ การใช้วิธีเกรย์ต้องใช้ขั้นตอนที่ถูกต้องในการกำหนดและประมวลผลข้อมูล เช่น การแก้ปัญหาเกรย์เบื้องต้น การสร้างโมเดลทางคณิตศาสตร์ เชิงสถิติ การประมาณค่าพารามิเตอร์ของโมเดล เป็นต้น ทั้งนี้การใช้วิธีเกรย์ควรพิจารณาความเหมาะสม และข้อจำกัดของวิธีในแต่ละกรณีการประยุกต์ใช้ ดังนั้นจึงมีการพัฒนาตัวแบบ GVM (Grey Verhulst Model) เพื่อใช้ในการพยากรณ์ข้อมูลอนุกรมเวลาที่ข้อมูลมีลักษณะแบบระบบเทา โดยใช้

ทฤษฎีระบบเกรย์ (Grey System Theory) เพื่อแก้ปัญหาดังกล่าวนั่นเอง

จากที่กล่าวมาข้างต้นผู้วิจัยจึงมีความสนใจในการศึกษาตัวแบบเกรย์ ภายใต้ตัวแบบ GM(1,1) และตัวแบบเกรย์เวอร์ไฮท์ (Grey Verhulst Model) และเปรียบเทียบประสิทธิภาพกับตัวแบบอาร์มาเพื่อหาตัวแบบที่เหมาะสมในการพยากรณ์ราคาหุ้นที่ให้ความถูกต้องและผิดพลาดน้อยที่สุด

2. การดำเนินการวิจัย

2.1 วิธีการพยากรณ์

2.1.1 ตัวแบบอาร์มา (Autoregressive Integrated Moving Average Model: ARIMA)

การพยากรณ์อนุกรมเวลาด้วยวิธีของบ็อกซ์-เจนกินส์ (Box-Jenkins) เป็นวิธีการพยากรณ์ที่มีความแม่นยำของการพยากรณ์ค่อนข้างสูงมากเมื่อเปรียบเทียบกับวิธีการพยากรณ์แบบอื่น ๆ การกำหนดตัวแบบให้กับอนุกรมเวลาตามวิธีของบ็อกซ์-เจนกินส์ จะกำหนดตัวแบบอยู่ในกลุ่มของตัวแบบ ARMA(p, q) (Autoregressive and Moving Average order p and q) หรือ ARIMA(p, d, q) (Autoregressive Integrated Moving Average of order p, d and q) ตัวแบบ ARMA(p, q) จะเป็นตัวแบบที่ใช้กับข้อมูลอนุกรมเวลาที่มีลักษณะเป็นอนุกรมเวลาที่เสถียร (Stationary) สำหรับตัวแบบ ARIMA(p, d, q) จะเป็นตัวแบบที่ใช้กับข้อมูลอนุกรมเวลาที่ไม่เป็นเสถียร (Nonstationary) แสดงได้ดังนี้

ตัวแบบอาร์มา

$$(1 - B)^d (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) Y_t = \delta + (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \varepsilon_t$$

เมื่อ Y_t คือ ข้อมูลอนุกรมเวลา ณ เวลา t

δ คือ ค่าคงที่

ϕ_i คือ พารามิเตอร์ของการถดถอยในตัว

(Autoregressive parameter)

θ_i คือ พารามิเตอร์ของค่าเฉลี่ยเคลื่อนที่

(Moving Average parameter)

d คือ จำนวนครั้งที่หาผลต่าง

ε_t คือ ค่าความคลาดเคลื่อน ณ เวลา t ที่มีการแจกแจงปกติซึ่งมีค่าเฉลี่ยเท่ากับ 0 ค่า

ความแปรปรวนเท่ากับ σ^2

B คือ backward-shift operator

โดยที่ $B^p X_t = X_{t-p}$ และ $B^q \varepsilon_t = \varepsilon_{t-q}$

2.1.2 ตัวแบบเกรย์ (Grey Model)

วิธีเกรย์ใช้กฎการอนุมานเกรย์ (Grey System Theory) เพื่อแยกออกมาเป็นสองส่วนหลักๆ คือ ข้อมูลที่มีความเปลี่ยนแปลงได้โดยตรง (Primary Data) และข้อมูลที่มีความเปลี่ยนแปลงตามกฎเกรย์ (Grey Data) โดยข้อมูลที่มีความเปลี่ยนแปลงได้โดยตรงจะถูกนำมาใช้ในการสร้างโมเดลทางคณิตศาสตร์ โดยเริ่มต้นจากการกำหนดข้อมูลราคาหุ้นโดยนำมาจัดเรียงตามสมการที่ (1)

$$Y = (y(1), y(2), \dots, y(k), y(k+1), \dots, y(n)) \quad (1)$$

โดยที่ Y คือ ราคาหุ้น BBL

$y(k)$ คือ ราคาหุ้นรายวันในลำดับที่ k

หลังจากนั้นนำค่าของ $y(k)$ แล้วจนถึงค่าที่ $y(k+1)$ ไม่เกินค่าที่ $y(n)$ เมื่อ $k \leq n-1$ นำ $y(k)$ จัดเรียงรูปแบบของ Y โดยที่ภายในลำดับ k ของข้อมูลเริ่มต้นที่จัดเก็บแล้ว นำมาเรียงเป็น

$$Y = (y(1), y(2), \dots, y(k), y(k+1), \dots, y(n))$$

การนำข้อมูลเริ่มต้นมาจัดเป็นรูปแบบจำลองของ

$$Y^{(0)} = (y^{(0)}(1), y^{(0)}(2), \dots, y^{(0)}(n))$$

โดยใช้วิธีการคำนวณแบบตัวสะสม (Accumulation Generation) จะให้ได้เป็น

$$Y^{(1)} = (y^{(1)}(1), y^{(1)}(2), \dots, y^{(1)}(n)) \text{ แล้วนำจัดรูปแบบได้}$$

$$y^{(1)}(k) + ay^{(1)}(k) = b \quad (2)$$

รูปแบบของสมการที่ (2) เรียกว่ารูปแบบเริ่มต้นของการจำลองแบบ GM(1,1) ซึ่งสัญลักษณ์ที่ใช้เป็น GM(1,1) เมื่อเริ่มคำนวณโดยกำหนด

$$Z^{(1)} = (y^{(1)}(1), y^{(1)}(2), \dots, y^{(1)}(n))$$

$Z^{(1)}$ เป็นลำดับคำนวณที่เกิดจาก $Y^{(1)}$ โดยวิธีเลือกปรับข้อมูลเฉลี่ยที่อยู่ติดกันแล้วแสดงวิธีคำนวณดังนี้

$$z^{(1)}(k) = \frac{1}{2} (y^{(1)}(k) + y^{(1)}(k-1)), k = 2, 3, \dots, n$$

ค่าที่ตามลำดับ k ใน $z^{(1)}(k)$ นำไปแทนค่าสมการที่ (3)

$$y^{(1)}(k) + az^{(1)}(k) = b \quad (3)$$

1) ตัวแบบ GM(1,1)

ตัวแบบ GM(1,1) เป็นรูปแบบหนึ่งที่ใช้กันแพร่หลายมากที่สุด และเป็นรูปแบบการพยากรณ์อนุกรมเวลาสมการเชิงอนุพันธ์ของ GM(1,1) มีค่าสัมประสิทธิ์ที่แปรผันตามระยะเวลา สำหรับขั้นตอนการสร้างแบบจำลอง GM(1,1) มีดังนี้ กำหนดข้อมูล $Y^{(0)}, Y^{(1)}$ และ $Z^{(1)}$ ตามเงื่อนไขของรูปแบบพื้นฐานของตัวแบบ GM(1,1) โดยที่ค่าของ $Y^{(0)}$ จะเป็นบวกเพื่อใช้ในการหาค่าของ $\hat{\alpha} = (a, b)^{(T)}$ เป็นไปตามลำดับของพารามิเตอร์ [8]

$$W = \begin{bmatrix} y^{(0)}(2) \\ y^{(0)}(3) \\ \vdots \\ y^{(0)}(n) \end{bmatrix}, B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix}$$

จากนั้นหาค่า $\hat{\alpha}$ จากสมการ (4)

$$\hat{\alpha} = (a, b)^{(T)} = (B^{(r)}B)^{(-1)} B^{(r)}W \quad (4)$$

โดยที่ $\hat{\alpha}$ คือ ค่าประมาณกำลังสองน้อยที่สุด (Least squares estimate) ของสมการ (3) ของตัวแบบ GM(1,1)

ผลการคำนวณจะได้เป็น $\hat{\alpha} = \begin{bmatrix} a \\ b \end{bmatrix}$ ซึ่งนำไปแทนในสมการ $y^{(1)}(k) + az^{(1)} = b$ จากนั้นหาอนุพันธ์ของรูปแบบฟังก์ชันตอบสนองเวลา (Time response function) ของตัวแบบสมการเชิงอนุพันธ์สามัญ (Ordinary Differential Equation: ODE) ดังสมการที่ (5)

$$\frac{dy^{(1)}}{dt} + ay = b \quad (5)$$

$$y^{(1)}(t) = \left(y^{(1)}(1) - \frac{b}{a} \right) e^{-at} + \frac{b}{a} \quad (6)$$

จากสมการ (6) มีอัตราการเปลี่ยนเวลา (1) ตามลำดับของตัวแบบจำลอง GM(1,1) ในสมการที่ (3) จะสามารถคำนวณค่าพยากรณ์ได้ดังสมการที่ (7)

$$\hat{y}^{(1)}(k+1) = \left(y^{(1)}(1) - \frac{b}{a} \right) e^{-ak} + \frac{b}{a}, k = 2, 3, \dots, n \quad (7)$$

2) ตัวแบบเกรย์เวอร์เอ้าท์ (Grey Verhulst Model)

ตัวแบบเกรย์เวอร์เอ้าท์ หรือตัวแบบ GVM ใช้โมเดลทางคณิตศาสตร์ที่เรียกว่าสมการการเจริญเติบโตแบบ logistic growth ซึ่งได้รับความนิยมและนำไปใช้ในการศึกษาและการอธิบายกระบวนการเจริญเติบโตของประชากรและสังคมในหลายรูปแบบ ตัวแบบเกรย์เวอร์เอ้าท์สามารถ

กำหนดเป็นลำดับการประมาณกำลังสองน้อยที่สุดของสมการความแตกต่างตัวแบบ GM(1,1) มีค่าจำกัดความดังนี้

$$\frac{dy^{(1)}}{dy} + ay^{(1)} = b \left(y^{(1)} \right)^2 \quad (8)$$

สมการผลต่างตัวแบบเกรย์ของสมการ (8) คือ

$$y^{(0)}(k) + az^{(1)}(k) = b \left(z^{(1)}(k) \right)^2 \quad (9)$$

จากสมการ (9) จะหาค่าประมาณกำลังสองน้อยที่สุดของ

สมการคือ $\hat{\alpha} = \begin{bmatrix} a \\ b \end{bmatrix}$ ซึ่งสามารถหาได้จาก

$$W = \begin{bmatrix} y^{(0)}(2) \\ y^{(0)}(3) \\ \vdots \\ y^{(0)}(n) \end{bmatrix}, B = \begin{bmatrix} -z^{(1)}(2) & \left(z^{(1)}(2) \right)^2 \\ -z^{(1)}(3) & \left(z^{(1)}(3) \right)^2 \\ \vdots & \vdots \\ -z^{(1)}(n) & \left(z^{(1)}(n) \right)^2 \end{bmatrix}$$

การแก้สมการให้แทนค่า $\hat{\alpha}$ ในสมการตัวแบบเกรย์เวอร์เอ้าท์ [4] คือ

$$y^{(1)}(t) = \frac{ay^{(1)}(0)}{e^{at} \left[a - by^{(1)}(0)(1 - e^{-at}) \right]}$$

และแทนค่าลงในฟังก์ชันตอบสนองเวลาของตัวแบบสมการ (10) จะได้ตัวแบบการพยากรณ์คือ

$$\hat{y}^{(1)}(k+1) = \frac{ay^{(1)}(0)}{by^{(1)}(0) + (a - by^{(1)}(0))e^{ak}}, k = 2, 3, \dots, n \quad (10)$$

2.2 เกณฑ์การตัดสินใจ

การนำวิธีการพยากรณ์มาเปรียบเทียบกับวิธีใดที่เหมาะสมกับข้อมูลที่นำมาศึกษา โดยพิจารณาจากค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Percentage Error: MAPE) และพิจารณาจากค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Root Mean Square Error: RMSE)

1) เกณฑ์การตัดสินใจจากค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE)

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100 \quad (11)$$

โดยที่ y_t คือ ค่าจริงของข้อมูลอนุกรมเวลา ณ เวลาที่ t

$\hat{y}_t$ คือ ค่าพยากรณ์ข้อมูลอนุกรมเวลา ณ เวลา ที่ t
 n คือ จำนวนข้อมูลหรือค่าสังเกต

2) เกณฑ์ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error: RMSE)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_t)^2} \quad (12)$$

โดยที่ y_t คือ ค่าจริงของข้อมูลอนุกรมเวลา ณ เวลา ที่ t
 $\hat{y}_t$ คือ ค่าพยากรณ์ของข้อมูลอนุกรมเวลา ณ เวลา ที่ t
 n คือ จำนวนข้อมูลหรือค่าสังเกต

เกณฑ์ในการเลือกตัวแบบพิจารณาจากค่า MAPE และ RMSE ของการพยากรณ์ในแต่ละวิธีมีค่าต่ำที่สุด ซึ่งแสดงให้เห็นว่าวิธีการพยากรณ์ดังกล่าวเหมาะสมกับข้อมูลอนุกรมเวลาชุดที่ใช้สร้างตัวแบบนั้น

2.3 ขั้นตอนการวิเคราะห์ข้อมูล

2.3.1 ตัวแบบอาร์มา

(1) พิจารณากราฟแสดงความเคลื่อนไหวของข้อมูล พบว่าราคาหุ้น BBL มีแนวโน้มทำให้ข้อมูลดังกล่าวไม่มีความเป็นสแตชันนารี จำเป็นต้องทำการแปลงข้อมูลให้เป็นสแตชันนารีเพื่อกำจัดแนวโน้ม

(2) กำหนดตัวแบบ $ARIMA(p,d,q)$ เมื่อวิเคราะห์กราฟ ACF และ PACF พิจารณาตัวแบบพยากรณ์ที่มีค่าพารามิเตอร์ Significant ต่ำที่สุด

(3) ตรวจสอบข้อตกลงเบื้องต้นในค่าความคลาดเคลื่อน ได้แก่ ความคลาดเคลื่อนกระจายคงที่และเป็นอิสระต่อกัน มีการแจกแจงปกติ มีค่าเฉลี่ยเท่ากับศูนย์และมีความแปรปรวนคงที่ σ^2

(4) นำค่าพยากรณ์ที่ได้จากการสร้างตัวแบบพยากรณ์ไปเปรียบเทียบกับข้อมูลชุดที่ 2 ที่ได้สำรองไว้เพื่อตรวจสอบความเหมาะสมของตัวแบบและคำนวณหาค่า MAPE และ RMSE

2.3.2 ตัวแบบเกรย์

(1) กำหนดข้อมูลเริ่มต้นตามลำดับของจำนวนข้อมูล

(2) หาค่าลำดับของค่ากลางและหาอนุพันธ์ของลำดับข้อมูล และหาสมการอนุพันธ์ของแบบจำลอง

(3) ประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุด

(4) นำค่าพยากรณ์ที่ได้จากการสร้างตัวแบบพยากรณ์ไปเปรียบเทียบกับข้อมูลชุดที่ 2 ที่ได้สำรองไว้และคำนวณหาค่า MAPE และ RMSE

3. ผลการวิเคราะห์ข้อมูล

จากการพิจารณาลักษณะการเคลื่อนไหวของอนุกรมเวลาราคาหุ้น BBL ตั้งแต่เดือนกรกฎาคม 2565 ถึงเดือนพฤษภาคม 2566 จำนวน 222 วัน พบว่าราคาหุ้น BBL มีอิทธิพลของแนวโน้ม (รูปที่ 1) และเมื่อพิจารณาจากการทดสอบด้วยการทดสอบแบบวิ่ง (Run test) เพื่อตรวจสอบอิทธิพลของแนวโน้ม พบว่าราคาหุ้น BBL มีค่ามัธยฐานในแต่ละวันแตกต่างกันที่ระดับนัยสำคัญ 0.05 ($p\text{-value} = 0.000$) จึงสามารถสรุปได้ว่าราคาหุ้น BBL มีอิทธิพลของแนวโน้ม

จากผลการวิเคราะห์ที่พบว่าอนุกรมเวลาราคาหุ้น BBL มีอิทธิพลแนวโน้ม จึงแปลงข้อมูลด้วยการหาผลต่างลำดับที่ 1 ($d = 1$) เพื่อกำจัดแนวโน้มและ นำมาสร้างกราฟฟังก์ชันสหสัมพันธ์ในตัวเอง (Autocorrelation function: ACF) และกราฟฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วน (Partial autocorrelation function: PACF) เพื่อสร้าง ตัวแบบพยากรณ์โดยวิธีบ็อกซ์-เจนกินส์ พบว่า อนุกรมเวลาหลังจากการแปลงข้อมูลมีลักษณะคงที่ (รูปที่ 2)

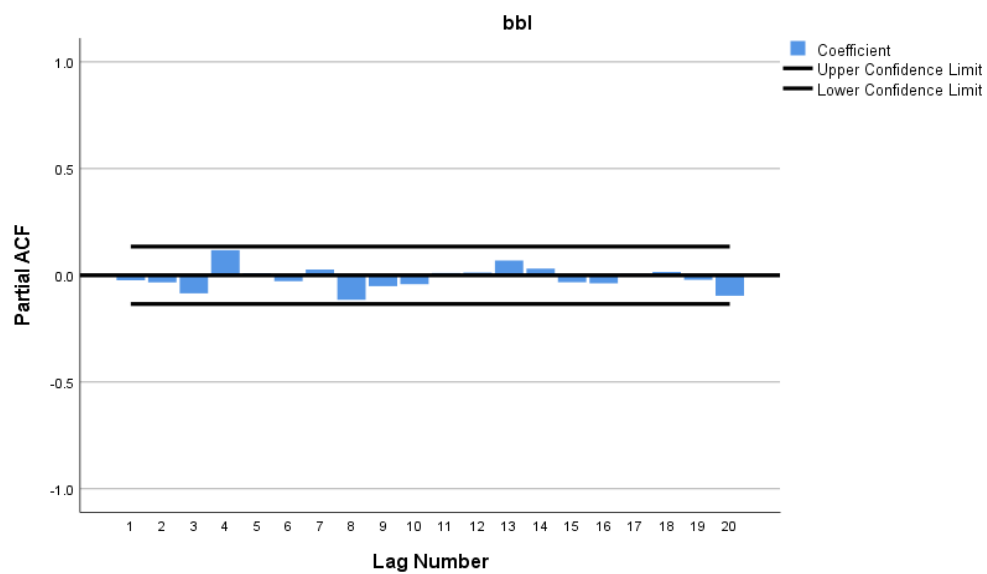
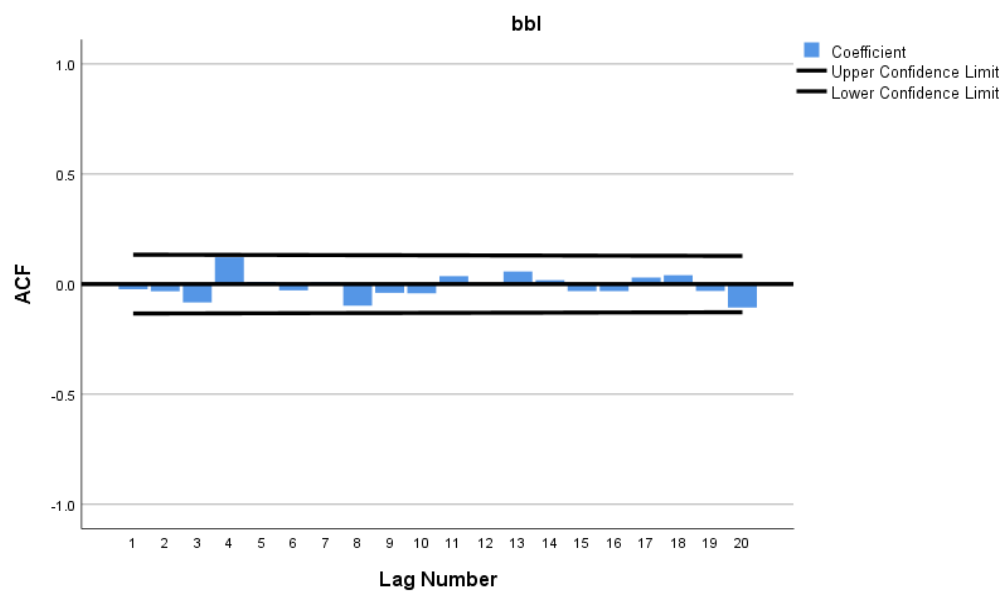
3.1 ตัวแบบอาร์มา

ข้อมูลที่น่ามาวิเคราะห์เป็นข้อมูลรายวัน จากการดำเนินการตามขั้นตอนการสร้างตัวแบบการพยากรณ์ โดยตัวแบบอาร์มาและทำการประมวลผลจากโปรแกรม SPSS โดยได้ตัวแบบการพยากรณ์ที่เหมาะสมคือ $ARIMA(0, 1, 0)$ โดยค่าประมาณของ ดังนั้นเมื่อนำ ไปแทนในสมการที่ 1 จะได้ตัวแบบการพยากรณ์ราคาหุ้น BBL โดยตัวแบบอาร์มาดังนี้

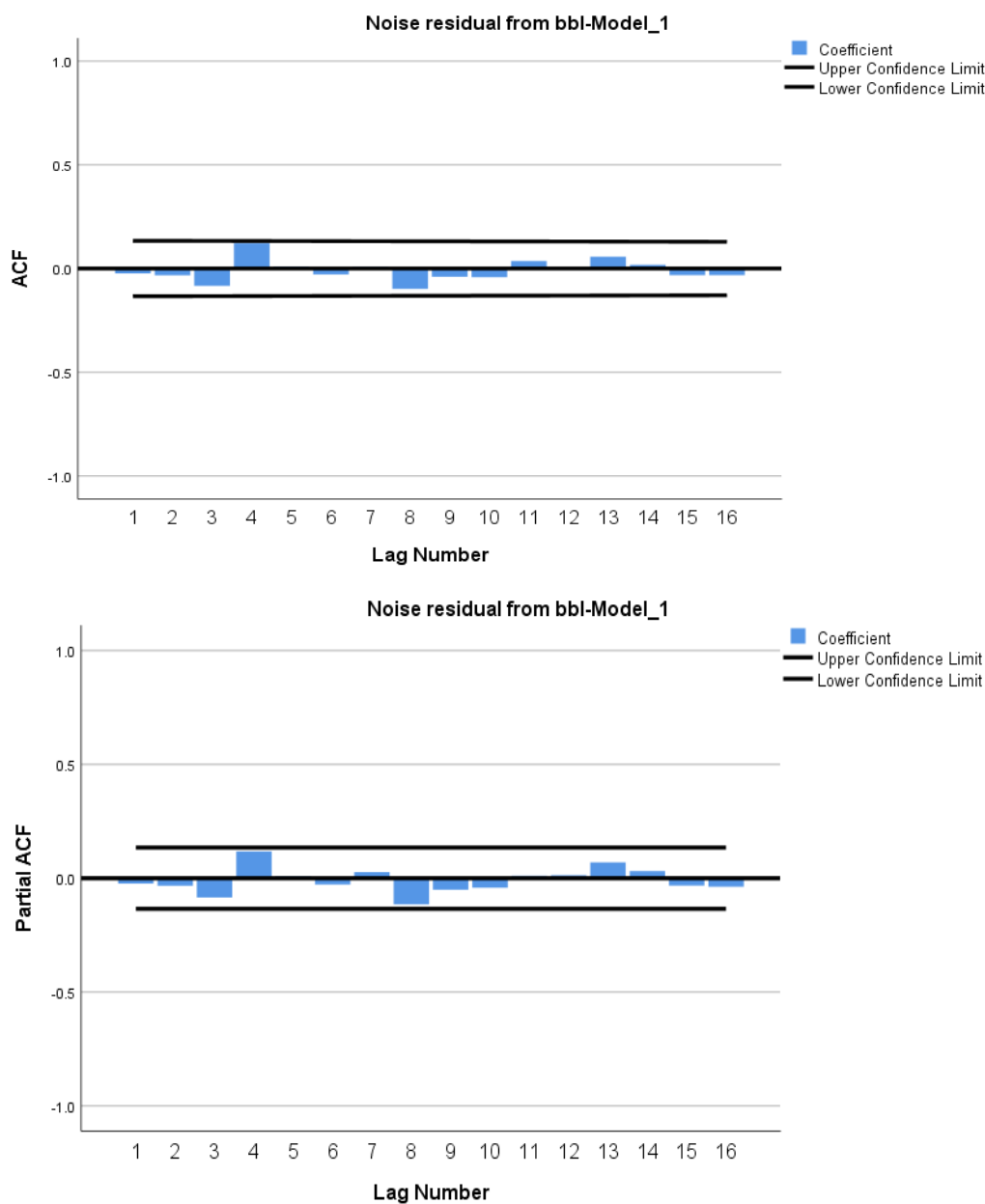
เมื่อตรวจสอบความเหมาะสมของตัวแบบ พบว่าความคลาดเคลื่อนเป็นอิสระต่อกัน (แสดงรายละเอียดในรูปที่ 3) โดยพบว่า ค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเองและสัมประสิทธิ์ในตัวเองบางส่วนของความคลาดเคลื่อนตกอยู่ในขอบเขตความเชื่อมั่นร้อยละ 95 และความคลาดเคลื่อนมีการแจกแจงปกติ (แสดงรายละเอียดในรูปที่ 4) จึงสามารถสรุปได้ว่าตัวแบบมีความเหมาะสม เกณฑ์การเปรียบเทียบความแม่นยำจากการพยากรณ์จากสมการที่ 11 และ 12 หาค่า MAPE เท่ากับ 0.810 และค่า RMSE เท่ากับ 1.585 ตามลำดับ



รูปที่ 1 การเคลื่อนไหวของอนุกรมเวลาราคาหุ้น BBL



รูปที่ 2 ACF และ PACF ของอนุกรมเวลาที่แปลงข้อมูลด้วยการหาผลต่าง $d = 1$



รูปที่ 3 ACF และ PACF ของค่าความคลาดเคลื่อน

One-Sample Kolmogorov-Smirnov Test

		Noise residual from bbl-Model_1
N		222
Normal Parameters ^{a,b}	Mean	.0097
	Std. Deviation	1.64714
Most Extreme Differences	Absolute	.041
	Positive	.032
	Negative	-.041
Test Statistic		.041
Asymp. Sig. (2-tailed)		.200 ^{c,d}

a. Test distribution is Normal.

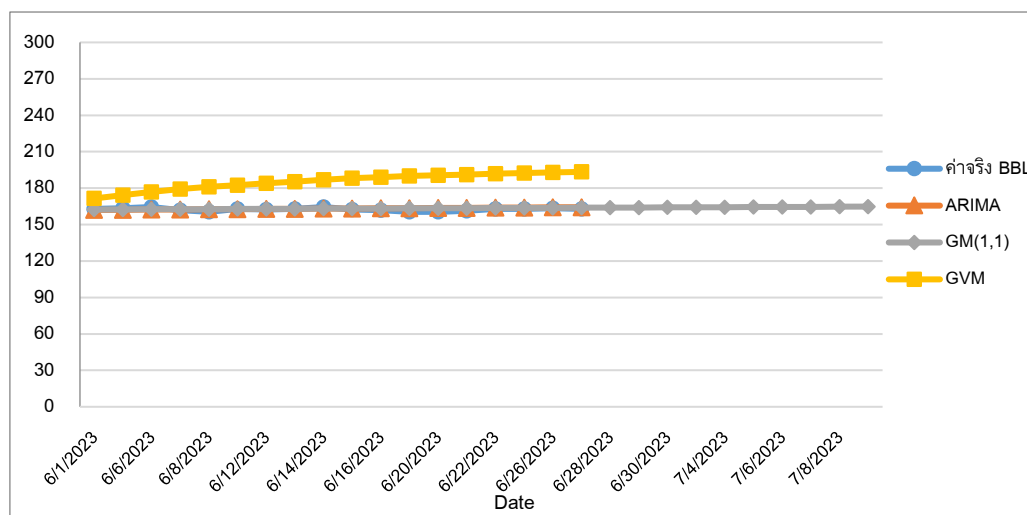
รูปที่ 4 การทดสอบการแจกแจงปกติของค่าความคลาดเคลื่อน

ตารางที่ 1 ค่าพยากรณ์ตัวแบบอาร์มา ตัวแบบ GM(1,1) และตัวแบบเกรย์เวอร์เนอร์ไฮท์ ของข้อมูลชุดที่ 2

วันที่	ค่าจริง	ARIMA	GM(1,1)	GVM
01/06/2023	163.0	162.13	162.41	171.54
02/06/2023	163.5	162.26	162.50	174.28
06/06/2023	164.5	162.39	162.59	176.84
07/06/2023	162.0	162.52	162.68	179.33
08/06/2023	160.5	162.64	162.77	181.08
09/06/2023	163.0	162.77	162.86	182.43
12/06/2023	162.5	162.90	162.95	184.07
13/06/2023	163.0	163.03	163.04	185.47
14/06/2023	164.5	163.16	163.13	186.81
15/06/2023	162.5	163.29	163.22	188.16
16/06/2023	162.0	163.42	163.31	189.17
19/06/2023	160.5	163.55	163.40	190.02
20/06/2023	160.5	163.68	163.49	190.67
21/06/2023	161.5	163.81	163.58	191.27
22/06/2023	163.0	163.93	163.67	191.86
23/06/2023	163.0	164.06	163.76	192.45
26/06/2023	163.5	164.19	163.85	192.96
27/06/2023	163.0	164.32	163.93	193.40
MAPE		0.810	0.724	14.223
RMSE		1.585	1.463	24.114

ตารางที่ 2 ค่าพยากรณ์ตัวแบบ GM(1,1) ช่วงหน้า 10 วัน

วันที่	GM(1,1)
28/06/2023	164.02
29/06/2023	164.11
30/06/2023	164.20
03/07/2023	164.29
04/07/2023	164.39
05/07/2023	164.48
06/07/2023	164.57
07/07/2023	164.66
08/07/2023	164.75
09/07/2023	164.84



รูปที่ 5 การเปรียบเทียบค่าพยากรณ์ตัวแบบอาร์มา ตัวแบบ GM(1,1) และตัวแบบเกรย์เวอร์เนอร์ไฮท์

3.2 ตัวแบบ GM(1,1)

จากการพยากรณ์โดยใช้ตัวแบบ GM(1,1) สามารถประมาณค่า $\hat{x}$ โดยวิธีกำลังสองน้อยที่สุดของตัวแบบ

$$GM(1,1) \text{ จะได้เป็น } \hat{x} = \begin{bmatrix} -0.000548382 \\ 160.6001844 \end{bmatrix}$$

เมื่อนำ $\hat{x}$ ไปแทนในสมการที่ (7) จะได้ตัวแบบการพยากรณ์ราคาหุ้น BBL โดยตัวแบบ GM(1,1) ดังนี้

$$\hat{y}^{(1)}(k+1) = \left(y^{(1)}(1) - \frac{160.6001844}{-0.000548382} \right) e^{-0.000548382k} + \frac{160.6001844}{-0.000548382}$$

เกณฑ์การเปรียบเทียบความแม่นยำจากการพยากรณ์จากสมการที่ (11) และ (12) หาค่า MAPE เท่ากับ 0.724 และค่า RMSE เท่ากับ 1.463 ตามลำดับ

3.3 ตัวแบบเกรย์เวอร์เนอร์

จากการพยากรณ์โดยใช้ตัวแบบเกรย์เวอร์เนอร์สามารถประมาณค่า $\hat{x}$ โดยวิธีกำลังสองน้อยที่สุดของตัว

$$\text{แบบเกรย์เวอร์เนอร์ จะได้เป็น } \hat{x} = \begin{bmatrix} -0.1078711 \\ -0.000015 \end{bmatrix}$$

เมื่อนำ $\hat{x}$ ไปแทนในสมการที่ (10) จะได้ตัวแบบการพยากรณ์ราคาหุ้น BBL โดยตัวแบบเกรย์เวอร์เนอร์ดังนี้

$$\hat{y}^{(1)}(k+1) = \frac{-0.1078711y^{(1)}(0)}{-0.000015y^{(1)}(0) + (-0.1078711 + 0.000015y^{(1)}(0))e^{-0.1078711k}}$$

เกณฑ์การเปรียบเทียบความแม่นยำจากการพยากรณ์จากสมการที่ (11) และ (12) หาค่า MAPE เท่ากับ 14.223 และค่า RMSE เท่ากับ 24.114 ตามลำดับ

จากการศึกษาตัวแบบการพยากรณ์โดยใช้ตัวแบบอาร์มา ในสมการที่ (1) ตัวแบบ GM(1,1) ในสมการที่ (7) และตัวแบบเกรย์เวอร์เนอร์ (GVM) ในสมการที่ (10) สำหรับการพยากรณ์ข้อมูลชุดที่ 2 คือ ราคาหุ้น BBL เดือนมิถุนายน 2566 จำนวน 18 ค่า ได้ค่าพยากรณ์ ค่าร้อยละ ค่าตลาดเคลื่อนสามบุรณ์เฉลี่ยและค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) แสดงดังตารางที่ 1 ผลการเปรียบเทียบพบว่า จากวิธีการพยากรณ์ทั้ง 3 วิธี ตัวแบบ GM(1,1) มีความแม่นยำมากที่สุดเนื่องจากให้ค่าพยากรณ์ที่มีความแตกต่างน้อยที่สุด นอกจากนี้ยังพบว่าตัวแบบ GM(1,1) ให้ค่า MAPE และ RMSE ต่ำที่สุด จึงถือว่าเป็นตัวแบบที่เหมาะสมที่สุดกับข้อมูลที่น่ามาศึกษา

นอกจากนี้จากการตรวจสอบความแม่นยำของตัวแบบพยากรณ์ทั้ง 3 ตัวแบบคือ ตัวแบบอาร์มา ตัวแบบ GM(1,1) และตัวแบบเกรย์เวอร์เนอร์ (GVM) เมื่อทราบค่าตัวแบบพยากรณ์ได้มีค่าร้อยละค่าตลาดเคลื่อนสามบุรณ์เฉลี่ยและค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ต่ำที่สุด จะใช้ตัวแบบพยากรณ์นั้นสำหรับการพยากรณ์ราคาหุ้น BBL เดือนมิถุนายน-กรกฎาคม 2566 ต่อไป แสดงดังรูปที่ 4

4. การอภิปรายและสรุปผลการวิเคราะห์

การวิจัยครั้งนี้ได้นำเสนอวิธีการสร้างและคัดเลือกตัวแบบการพยากรณ์ที่เหมาะสมกับราคาหุ้น BBL โดยใช้อนุกรมเวลารายวัน ได้แบ่งข้อมูลออกเป็น 2 ชุด ชุดที่ 1 ตั้งแต่เดือนกรกฎาคม 2565 ถึงเดือนพฤษภาคม 2566 จำนวน 222 วัน สำหรับการสร้างตัวแบบพยากรณ์ด้วยวิธีการทางสถิติทั้งหมด 3 วิธี ได้แก่ ตัวแบบอาร์มา และตัวแบบเกรย์เวอร์เนอร์ได้ตัวแบบ GM(1,1) และตัวแบบเกรย์เวอร์เนอร์ ข้อมูลชุดที่ 2 เดือนมิถุนายน 2566 จำนวน 18 วัน สำหรับการสร้างตัวแบบพยากรณ์ด้วยวิธีการทางสถิติทั้งหมด 3 วิธี ได้แก่ ตัวแบบอาร์มา และตัวแบบเกรย์เวอร์เนอร์ได้ตัวแบบ GM(1,1) และตัวแบบเกรย์เวอร์เนอร์ (GVM) โดยตรวจสอบความแม่นยำของตัวแบบการพยากรณ์ด้วยร้อยละค่าตลาดเคลื่อนสามบุรณ์เฉลี่ยและค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ผลการวิจัยพบว่า จากวิธีการพยากรณ์ทั้งหมดที่ได้ศึกษาตัวแบบที่มีความแม่นยำมากที่สุด คือ ตัวแบบ GM(1,1) สามารถทำนายค่าราคาหุ้น BBL ได้ใกล้เคียงค่าจริง ทำให้มีความผิดพลาดจากการพยากรณ์ร้อยละ 0.724 (MAPE = 0.724) และมีความผิดพลาดจากการพยากรณ์ 1.463 (RMSE = 1.463) ที่ต่ำที่สุด โดยตัวแบบการพยากรณ์ราคาหุ้น BBL โดยตัวแบบ GM(1,1) ดังนี้

$$\hat{y}^{(1)}(k+1) = \left(y^{(1)}(1) - \frac{160.6001844}{-0.000548382} \right) e^{-0.000548382k} + \frac{160.6001844}{-0.000548382}$$

การศึกษาครั้งต่อไป สามารถที่จะประยุกต์ใช้ตัวแบบเกรย์เวอร์เนอร์ได้ตัวแบบ GM(1,1) และตัวแบบเกรย์เวอร์เนอร์ กับข้อมูลชนิดอื่น ๆ เช่น ข้อมูลทางธุรกิจ ข้อมูลทางทรัพยากรธรรมชาติ นอกจากนี้ยังพบว่าวิธีเกรย์สามารถใช้ในการทำนายข้อมูลที่มีจำนวนข้อมูลน้อย ข้อมูลที่ไม่เพียงพอ หรือข้อมูลที่ไม่สมดุลได้ และตัวแบบเกรย์เวอร์เนอร์ซึ่งเหมาะสมกับพยากรณ์ข้อมูลที่มีลักษณะเติบโต

แบบ Logistic ดังนั้นจึงสามารถนำตัวแบบเกรย์เวอร์เข้าที่ไปประยุกต์ใช้กับข้อมูลที่มีลักษณะดังกล่าว เช่น ปริมาณน้ำระบาย เพื่อให้ค่าพยากรณ์มีความแม่นยำมากขึ้น และควรพิจารณาอนุกรมเวลารายเดือนและรายสัปดาห์ประกอบการสร้างตัวแบบการพยากรณ์ด้วย

กิตติกรรมประกาศ

งานวิจัยนี้ได้รับทุนสนับสนุนจากคณะวิทยาศาสตร์ประยุกต์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ ตามสัญญาเลขที่ 6245110

เอกสารอ้างอิง

- Kayacan, E., Ulutas, B., & Kaynak, O. (2010). Grey system theory-based models in time series prediction. *Expert systems with applications*, 37(2), 1784–1789.
- Shen, C., Du, F., Zhang, L., & Song, H. (2019, February). Application of GM (1, N) model in groundwater mineralization in Cheng'an County. In *IOP Conference Series: Earth and Environmental Science*, 233(4), 042010. IOP Publishing.
- Puripat, C., & Sarikavanij, S. (2018). The comparison of Grey system and the Verhulst model for rainfall and water in Dam prediction. *Advances in Meteorology*, 2018.
- Davis, P. D., Amankwah, G., & Fang, Q. (2019). Predicting the rainfall of ghana using the grey prediction model GM (1, 1) and the Grey verhulst model. *International Research Journal of Engineering and Technology (IRJET)*, 6(8), 1362–1372.
- He, F., & Tao, T. (2014). An Improved Coupling Model of Grey-System and Multivariate Linear Regression for Water Consumption Forecasting. *Polish Journal of Environmental Studies*, 23(4).
- Amphanthong, P., & Busabodhin, P. (2015). Forecasting PM10 in the upper northern area of Thailand with grey system theory. *Burapha Science Journal*, 20(1), 15–24.
- Doungdee, N., & Chiangpradit, M. (2018). Forecasting Crisis Peaks of Economics in Thailand Using Grey Model. *Burapha Science Journal*, 1111–1122.
- Liu, S., & Forrest, J. Y. L. (2010). *Grey systems: theory and applications*. Springer Science & Business Media.

การพัฒนาผลิตภัณฑ์ขนมทองม้วนโดยใช้แป้งรำข้าวปกาอำปูล สกัดน้ำมันทดแทนแป้งสาลี

Development of Kha Nom Tong Muan by using Paka-umpoul Defatted Rice Bran Flour Substituted Wheat Flour (*Oryza sativa* L.)

ภัทรารณณ์ แก้วกูล^{1*}, อภิชากร ดวงแซ่เพ็ญศิริกุล¹ และ ณัฐปภัสร ฉลาดมาก²

Phattaraporn Kaewkool^{1*}, Apichargorn Dangkhapensirikul¹ and Nutpaphat Chaladmak²

¹สาขาวิชาธุรกิจการแปรรูปอาหาร คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏสุรินทร์ ประเทศไทย และ

²สำนักส่งเสริมวิชาการและงานทะเบียน มหาวิทยาลัยราชภัฏสุรินทร์ ประเทศไทย

¹Department of Food Processing Business Faculty of Science and Technology Surindra Rajabhat University Surin, Thailand and

²Academic Promotion and Registration Office, Surindra Rajabhat University, Surin, Thailand

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์ เพื่อศึกษาระดับการทดแทนแป้งสาลีด้วยแป้งรำข้าวปกาอำปูลสกัดน้ำมันที่เหมาะสมในการผลิตผลิตภัณฑ์ขนมทองม้วน เพื่อพัฒนาเป็นขนมขบเคี้ยวที่มีคุณค่าทางโภชนาการสูงขึ้น ทำการศึกษาโดยการวางแผนการทดลองแบบสุ่มสมบูรณ์ (Complete Randomized Design: CRD) ใช้แป้งรำข้าวปกาอำปูลสกัดน้ำมันทดแทนแป้งสาลี 5 ระดับ คือ ร้อยละ 0, 10, 20, 30 และ 40 (สูตรที่ 1, 2, 3, 4 และ 5 ตามลำดับ) นำไปทดสอบคุณภาพทางประสาทสัมผัสด้วยวิธี 9-point hedonic scale scoring test ใช้ผู้ทดสอบทั้งฝึกฝน จำนวน 30 คน พบว่า ผลิตภัณฑ์ขนมทองม้วนสูตรที่ 3 ซึ่งใช้แป้งรำข้าวปกาอำปูลสกัดน้ำมันทดแทนแป้งสาลีในอัตราร้อยละ 20 ได้รับคะแนนความชอบโดยรวมมากที่สุด (7.82 คะแนน) และเมื่อนำไปเปรียบเทียบกับคุณภาพทางกายภาพ เคมี และจุลชีววิทยากับสูตรพื้นฐาน (สูตรที่ 1) พบว่า ค่าสีมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับความเชื่อมั่น ร้อยละ 95 โดยการเพิ่มปริมาณแป้งรำข้าวปกาอำปูลสกัดน้ำมันมีผลทำให้ค่าความสว่าง (L^*) ของผลิตภัณฑ์มีค่าลดลง ผลิตภัณฑ์มีสีเข้มเพิ่มขึ้น และค่าปริมาณน้ำอิสระ (a_w) มีค่าเพิ่มขึ้นจาก 0.389 เป็น 0.429 จากการตรวจสอบคุณภาพทางเคมีพบว่า ปริมาณไขมัน เถ้า และใยอาหารสูงขึ้น ร้อยละ 23.01 2.18 และ 2.16 ตามลำดับ และจากการตรวจวิเคราะห์สารแอมไมโอไรซานอล และกิจกรรมต้านอนุมูลอิสระ โดยวิธี DPPH พบว่า ไม่สามารถตรวจพบในสูตรพื้นฐาน ในขณะที่ผลิตภัณฑ์ขนมทองม้วนสูตรที่ 3 พบว่ามีค่าเท่ากับ 0.031 มิลลิกรัมต่อ 100 กรัม และ 19.05 มิลลิกรัมต่อ 100 กรัม ตามลำดับ เมื่อตรวจสอบคุณภาพทางจุลินทรีย์ พบว่า มีเชื้อโคลิฟอร์ม <3.0 เอ็มพีเอ็น/ กรัม และ ยีสต์และรา <10 โคโลนี/ กรัม อยู่ในเกณฑ์มาตรฐานผลิตภัณฑ์ชุมชนขนมไทย (มผช.1/2552)

คำสำคัญ: การทดแทนแป้งสาลี, ขนมทองม้วน, แป้งรำข้าวปกาอำปูลสกัดน้ำมัน

ABSTRACT

The objective of this research is to develop the appropriate formula of kha nom tong muan from paka-umpoul defatted rice bran flour to develop as a snack with higher nutritional value. This study was done by planning a complete randomized design (CRD). The amount of paka-umpoul defatted rice bran flour substituted at 5 levels of 0, 10, 20, 30, and 40 % (formula 1, 2, 3, 4, and 5 respectively). It was tested for sensory quality using the 9-point hedonic scale scoring test using 30 semi-trained testers. It was found that the Thong Muan dessert product (formula 3) used paka-umpoule defatted rice bran flour substituted with wheat flour at the rate of 20 % received the highest overall liking score (7.82). The comparison between formula 3 and the basic formula (formula 1) in the physical, chemical, and microbiological quality found that the color values were significantly different at the 95% confidence level. By increasing the amount of defatted rice bran flour resulted in the brightness value (L^*) decreased and the product became darker in color. In addition, the free water content (a_w) increased from 0.389 to 0.429. The chemical quality of fat, ash, and fiber was higher by 23.01%, 2.18%, and 2.16%, respectively. Analysis of gamma-oryzanol and total antioxidant activity by DPPH assay was not found in the basic formula while was found in formula 3 at 0.031 mg per 100 and 19.05 mg per 100 grams, respectively. An examination of microbiological quality of formula 3 was found that there were coliforms <3.0 MPN/g. and yeast and mold <10 cfu/g. in Thai Dessert Community Product Standard Criteria (TCPS.1/2009).

KEYWORDS: Wheat flour substitution, Kha nom tong muan, Paka-umpoul defatted rice bran flour

*Corresponding Author: p_kaewkool@yahoo.com

1. บทนำ

ขนมทองม้วนเป็นขนมไทยประเภทแป้งหรืออบชนิดหนึ่ง มีกลิ่นหอมจากไข่ กะทิ และน้ำตาล เนื้อสัมผัสกรอบ ร่วน มีรสหวานพอเหมาะ เป็นที่นิยมของผู้บริโภคทุกเพศ ทุกวัย สะดวกต่อการพกพา และมีอายุการเก็บรักษา ยาวนาน ปัจจุบันมีการผลิตขนมทองม้วนเป็นสินค้าระดับ ชุมชนถึงระดับอุตสาหกรรมมากขึ้น ขนมทองม้วนให้ คุณค่าทางโภชนาการ เช่น โปรตีน คาร์โบไฮเดรต และ ไขมัน แต่ใยอาหารมีน้อย เนื่องจากวัตถุดิบหลักที่ใช้เป็น แป้งสาลี ทำให้ขนมทองม้วนมีคุณค่าทางโภชนาการต่ำ ปัจจุบันผู้บริโภคใส่ใจสุขภาพมากขึ้น จึงมีการวิจัยวัตถุดิบ ทดแทนแป้งสาลีที่เป็นวัตถุดิบหลักของขนมทองม้วน เช่น การใช้แป้งข้าวลินเเหล็ก (Sarawut, 2021) การใช้ข้าว กล้องมันปู (Sunan et al., 2016) การใช้แป้งข้าวกล้อง งอกพันธุ์พื้นเมืองสุรินทร์ (Phattaraporn et al., 2015) และการใช้แป้งข้าวสาลี (Supavadee, 2015) เป็นต้น โดยวัตถุดิบที่นำมาทดแทนแป้งสาลีนั้นควรเป็นวัตถุดิบที่มี คุณค่าทางโภชนาการสูงและหาง่ายในท้องถิ่น

ข้าว (*Oryza sativa* L.) เป็นสินค้าเกษตรที่เป็นพืช หลักของการส่งออกและการบริโภคภายในประเทศ ความสำคัญของข้าวนอกจากเป็นพืชเศรษฐกิจแล้วยังเป็น สินค้าทางวัฒนธรรม ข้าวแต่ละสายพันธุ์จะมีคุณค่าอยู่ใน ตัวเอง (Theerarat et al., 2019) กระบวนการสีข้าวเปลือก นอกจากจะได้เมล็ดข้าวขาวเป็นผลผลิตหลักแล้ว ยังมีรำ ข้าวเป็นผลผลิตพลอยได้ประมาณร้อยละ 10.5 ของข้าว ทั้งเมล็ด (Onanong, 2007) รำข้าวมีคุณค่าทาง โภชนาการสูง แต่การนำไปใช้ประโยชน์ยังมีน้อยมีเพียง บางส่วนเท่านั้นที่นำไปใช้ในการผลิตน้ำมันรำข้าว รำข้าว สกัดน้ำมันที่เหลือจากการกระบวนการผลิตน้ำมันรำข้าว ยังคงคุณค่าทางโภชนาการและมีความเสถียรค่อนข้างสูง มี รายงานวิจัย องค์ประกอบทางเคมีของรำข้าวสกัดน้ำมันแ ต่ละสายพันธุ์ 100 กรัม มีองค์ประกอบหลักไม่ต่างกัน คือ ไขมัน 1.85-13.73 โปรตีน 14.01-18.30 คาร์โบไฮเดรต 45.31-60.80 เถ้า 6.34-10.78 และ ใยอาหาร 5.78-13.51 (Ladda., 2017, Sukuntaros., 2017, Naiyana & Rewadee., 2002, Saunder., 1990 & Youssef., 1974) รำข้าวสกัดน้ำมันมีใยอาหารสูง เป็น แหล่งโปรตีน ช่วยป้องกันโรคเมะเร็งลำไส้ใหญ่ ช่วยรักษา อาการท้องผูก ช่วยควบคุมระดับน้ำตาลและไขมันในเลือด เหมาะกับการนำมาพัฒนาเป็นอาหารสำหรับกลุ่มแพ ้กลูเตน เด็ก และผู้สูงอายุ (Onanong, 2004) นอกจากนี้ ยังพบสารออกฤทธิ์ทางชีวภาพ (Bioactive compounds)

(Ferronato et al., 2021) ซึ่งเป็นสารสำคัญที่ช่วยลด อัตราเสี่ยงของโรคหลอดเลือดแข็งตัวตีตันโดยช่วยยับยั้ง การจับตัวของเกล็ดเลือดทำให้เลือดสามารถไหลเวียนใน ร่างกายได้ดีขึ้น ช่วยควบคุมให้ระดับคอเลสเตอรอลใน ร่างกายเป็นไปอย่างปกติ ช่วยลดระดับไตรกลีเซอไรด์ ช่วย บำรุงสายตา และเป็นสารต้านอนุมูลอิสระช่วยชะลอวัย (Saunder., 1990) จากการออกเก็บข้อมูลอาหารตาม วัฒนธรรมในเขตพื้นที่จังหวัดสุรินทร์ ของสาขาวิชา วิทยาศาสตร์และเทคโนโลยีการอาหาร มหาวิทยาลัยราช ภัฏสุรินทร์ พบว่า กลุ่มเกษตรกรบางพื้นที่รวมกลุ่มปลูก และอนุรักษ์ข้าวพันธุ์พื้นเมืองสุรินทร์มีความต้องการที่จะ แปรรูปผลิตภัณฑ์อาหารจากข้าวและพลอยได้ให้ม ีความหลากหลายและสร้างจุดเด่นให้ผลิตภัณฑ์

จากเหตุผลที่กล่าวมาผู้วิจัยจึงเล็งเห็นความสำคัญของ การใช้ประโยชน์จากรำข้าวพื้นเมืองสุรินทร์สกัดน้ำมัน: รำ ข้าวเปลือก (Paka-umpoul rice bran) ให้มากที่สุด ซึ่ง รำข้าวเปลือกเป็นรำข้าวเจ้าที่มีค่าดัชนีน้ำตาลต่ำ เหมาะ สำหรับผู้ป่วยเบาหวาน และผู้ที่ต้องการควบคุมน้ำหนัก (ศูนย์ข้อมูลตลาดข้าวเฉพาะ กรมการข้าว, ม.ป.ป) ด้วย การนำมาเป็นวัตถุดิบหลักทดแทนแป้งสาลีในผลิตภัณฑ์ ขนมทองม้วน เพื่อเพิ่มคุณค่าทางโภชนาการให้กับ ผลิตภัณฑ์ขนมทองม้วน และยังเป็นส่งเสริมใช้รำข้าวป ากาเปลือกสกัดน้ำมันเป็นวัตถุดิบเชิงกลยุทธ์ (Strategic ingredient) สร้างจุดขายที่แตกต่างเพิ่มมูลค่าสินค้าเชิง วัฒนธรรมของพื้นที่ เป็นทางเลือกแก่ผู้บริโภคที่รักสุขภาพ และชื่นชอบขนมไทย ดังนั้นวัตถุประสงค์ของงานวิจัยนี้จึง ศึกษาปริมาณรำข้าวเปลือกสกัดน้ำมันทดแทนแป้งสาลี ที่เหมาะสมในการผลิตขนมทองม้วนที่มีคุณค่าทาง โภชนาการที่ดีขึ้นและเป็นที่ยอมรับของผู้บริโภค

2. วิธีการทดลอง

1. วัตถุดิบ

รำข้าวพื้นเมืองสุรินทร์: รำข้าวเปลือก ได้จากตำบล คอโค อำเภอมืองสุรินทร์ จังหวัดสุรินทร์ ผ่านการสีไม่เกิน 24 ชั่วโมง นำมาสกัดน้ำมันแบบบีบเย็นด้วยเครื่องสกัด น้ำมันแบบสกรูเพรส (ยี่ห้อ Bosmall รุ่น M 92 ประเทศ จีน) สภาวะการสกัดที่อุณหภูมิห้อง จากนั้นนำรำข้าวป ากาเปลือกที่ผ่านการสกัดน้ำมันแล้วบดด้วยเครื่องบ ีบละเอียด (ยี่ห้อ Philips รุ่น HR 2225 ประเทศจีน) ร่อนผ่านตะแกรงขนาด 60 เมช จะได้แป้งรำข้าวเปลือก สกัดน้ำมันบรรจุถุงพอยล์แบบสุญญากาศและเก็บรักษาที่ อุณหภูมิ 5 - 8 องศาเซลเซียส เพื่อใช้ในการทดลองต่อไป

2. การพัฒนาผลิตภัณฑ์ขนมทองม้วนแป้งรำข้าวปกาอำปิลสกัดน้ำมันทดแทนแป้งสาลีบางส่วน

2.1 สูตรพื้นฐานผลิตภัณฑ์ขนมทองม้วน สูตรพื้นฐานในการผลิตขนมทองม้วนงานวิจัยนี้ใช้สูตรของ Phattaraporn et al. (2015) ซึ่งเป็นสูตรที่ผู้บริโภครู้จักให้การยอมรับมาแล้ว โดยมีส่วนผสม ดังนี้ แป้งสาลี 150 กรัม ไข่ไก่ 50 กรัม กะทิ 480 กรัม น้ำตาลทราย 150 กรัม และ เกลือป่น 5 กรัม

2.2 การศึกษาปริมาณแป้งรำข้าวปกาอำปิลสกัดน้ำมันทดแทนแป้งสาลีบางส่วนในผลิตภัณฑ์ขนมทองม้วน ปัจจัยที่ทำการศึกษาคือ ร้อยละแป้งรำข้าวปกาอำปิลสกัดน้ำมันทดแทนแป้งสาลีในผลิตภัณฑ์ขนมทองม้วน 5 ระดับ คือ ร้อยละ 0, 10, 20, 30 และ 40 ของน้ำหนักแป้งสาลี วางแผนการทดลองแบบสุ่มสมบูรณ์ (CRD) โดยส่วนผสมอื่น ๆ คงที่ ดังตารางที่ 1

2.3 วิธีการทำผลิตภัณฑ์ขนมทองม้วนแป้งรำข้าวปกาอำปิลสกัดน้ำมันทดแทนแป้งสาลี



1. จัดส่วนผสมทุกอย่างให้อย่างเหมาะสม 2. นำส่วนผสมทั้งหมดมาผสมให้เข้ากัน 3. ตักใส่พิมพ์ให้มีความร้อน

3. การวิเคราะห์คุณภาพทางประสาทสัมผัส (HE 631001)¹ การทดสอบคุณภาพทางประสาทสัมผัส โดยใช้ผู้ทดสอบชิมทั้งฝึกฝน จำนวน 30 คน เสรีตัวอย่างผลิตภัณฑ์ขนมทองม้วนตัวอย่างละ 1 ชิ้น ในภาชนะสีขาว ทดสอบความชอบ แบบ 9 point hedonic scale scoring test โดยให้ผู้ทดสอบชิมผลิตภัณฑ์ขนมทองม้วนและให้คะแนนความชอบของคุณลักษณะที่ตรงตามความรู้สึก ได้แก่ สี กลิ่น ลักษณะปรากฏ รสชาติ เนื้อสัมผัส และความชอบโดยรวม นำมาวิเคราะห์ความแปรปรวนของข้อมูลโดยใช้ Analysis of Variance (ANOVA) และเปรียบเทียบความแตกต่างของคะแนนเฉลี่ยด้วย Duncan's New Multiple Range test (DMRT) ที่ระดับความเชื่อมั่นร้อยละ 95 ด้วยโปรแกรมสำเร็จรูป (SPSS for Windows) จากนั้นนำผลิตภัณฑ์ขนมทองม้วนแป้งรำข้าวปกาอำปิลสกัดน้ำมันทดแทนแป้งสาลีที่ได้รับคะแนนความชอบโดยรวมมากที่สุดทำการวิเคราะห์คุณภาพด้านต่าง ๆ เปรียบเทียบกับผลิตภัณฑ์ขนมทองม้วนสูตรพื้นฐาน

4. การวิเคราะห์คุณภาพทางเคมี

4.1 วิเคราะห์องค์ประกอบทางเคมี ได้แก่ ความชื้น เถ้า โปรตีน ไขมัน คาร์โบไฮเดรต และใยอาหาร จากวิธีของ AOAC (2012)

4.2 วิเคราะห์สารแกมมา-โอโรซานอล ทำโดยการเทียบกับกราฟมาตรฐานสารแกมมา-โอโรซานอลที่ละลายในตัวทำละลาย (โดยดัดแปลงจากวิธีของ Perretti et al. 2003) ซึ่งตัวอย่างผลิตภัณฑ์ จำนวน 10 กรัม ใส่ขวดรูปชมพู่ ขนาด 250 มิลลิลิตร สกัดด้วยเมทานอล ปริมาตร 50 มิลลิลิตร เขย่าด้วยเครื่องเขย่า ที่อุณหภูมิ 50 องศาเซลเซียส เป็นเวลา 24 ชั่วโมง กรองสารละลายด้วยกระดาษกรองเบอร์ 1 ปรับปริมาตรให้ครบ 50 มิลลิลิตร จะได้สารสกัดตัวอย่างผลิตภัณฑ์ จากนั้นนำไปวิเคราะห์หาปริมาณสารแกมมา-โอโรซานอล ด้วยเครื่อง Spectrophotometer (ยี่ห้อ Perkin elmer รุ่น Lampda 365 ประเทศสหรัฐอเมริกา) วัดค่าการดูดกลืนแสงที่ 315 นาโนเมตร คำนวณหาปริมาณของสารแกมมา-โอโรซานอลต่อน้ำหนักแห้งของตัวอย่างแป้งรำข้าวปกาอำปิลสกัดน้ำมัน

4.3 วิเคราะห์กิจกรรมต้านอนุมูลอิสระ (antioxidant activity) ด้วยวิธี DPPH โดยดัดแปลงจากวิธีของ Kwaw et al. (2018) ซึ่งสารสกัดตัวอย่างผลิตภัณฑ์ จากข้อ 4.2 ปริมาตร 0.1 มิลลิลิตร ผสมกับสารละลาย DPPH (2,2-Diphenyl-1-picrylhydrazyl) ที่มีความเข้มข้น 0.1 มิลลิโมลลาร์ ปริมาตร 3 มิลลิลิตร (A_{sample}) เขย่าให้เข้ากัน บ่มในที่มืดเป็นเวลา 30 นาที ที่อุณหภูมิห้อง ตัวอย่างควบคุม (A_{control}) ใช้เมทานอลแทนสารสกัดตัวอย่างผลิตภัณฑ์ วัดค่าการดูดกลืนแสงที่ความยาวคลื่น 517 นาโนเมตร ด้วยเครื่อง Spectrophotometer ใช้ (ยี่ห้อ Perkin elmer รุ่น Lampda 365 ประเทศสหรัฐอเมริกา) คำนวณกิจกรรมต้านอนุมูลอิสระในหน่วยร้อยละของการยับยั้ง (%Inhibition)

ตารางที่ 1 สัดส่วนปริมาณแป้งรำข้าวปกาอำปิลสกัดน้ำมันทดแทนแป้งสาลีในผลิตภัณฑ์ขนมทองม้วน

ส่วนผสม	ปริมาณแป้งรำข้าวปกาอำปิลสกัดน้ำมันทดแทนแป้งสาลี				
	ในผลิตภัณฑ์ขนมทองม้วน (ร้อยละ)				
	0 (สูตรที่ 1)	10 (สูตรที่ 2)	20 (สูตรที่ 3)	30 (สูตรที่ 4)	40 (สูตรที่ 5)
แป้งสาลี	150	135	120	45	60
แป้งรำข้าวปกาอำปิลสกัดน้ำมัน	-	15	30	105	90

ตารางที่ 2 ลักษณะทางประสาทสัมผัสของผลิตภัณฑ์ขนมทองม้วนแป้งรำข้าวปกาอำปิลสกัดน้ำมันทดแทนแป้งสาลี

ระดับการทดแทน (ร้อยละ)	รูปภาพ	ลักษณะทางประสาทสัมผัส
0 (สูตรที่ 1)		- มีสีเหลืองอมน้ำตาล - ผิวหน้าเรียบ - มีกลิ่นหอมกะทิ - มีรสหวาน - เมื่อกัดมีความกรอบร่วน
10 (สูตรที่ 2)		- มีสีเหลืองอมน้ำตาลเล็กน้อย - ผิวหน้าเรียบ มีการกระจายตัวของแป้งรำข้าวเล็กน้อย - มีกลิ่นหอมกะทิ - มีรสหวาน - เมื่อกัดมีความกรอบร่วน
20 (สูตรที่ 3)		- มีสีน้ำตาลอ่อน - ผิวหน้าเรียบ มีการกระจายตัวของแป้งรำข้าวเล็กน้อย - มีกลิ่นหอมกะทิ - มีรสหวาน - เมื่อกัดมีความกรอบร่วน เวลาเคี้ยวแน่นเนื้อ
30 (สูตรที่ 4)		- มีสีน้ำตาล - ผิวหน้าเรียบสม่ำเสมอ มีการกระจายตัวของแป้งรำข้าวทั่วแผ่น - มีกลิ่นหอมกะทิ - มีรสหวาน - เมื่อกัดมีความกรอบ หักง่าย เวลาเคี้ยวแน่นเนื้อ
40 (สูตรที่ 5)		- มีน้ำตาลเข้มมากขึ้น - ผิวหน้าเรียบสม่ำเสมอ มีการกระจายตัวของแป้งรำข้าวทั่วแผ่น - มีกลิ่นหอมกะทิ - มีรสหวาน - เมื่อกัดมีความกรอบ หักง่าย เวลาเคี้ยวแน่นเนื้อขึ้น

ตารางที่ 3 คุณภาพเชิงประสาทสัมผัสของผลิตภัณฑ์ขนมทองม้วนแป้งรำข้าวปกาอำปิลสกัดน้ำมันทดแทนแป้งสาลี โดยการประเมินด้วย 9-point hedonic scale scoring test

ปัจจัย	ปริมาณแป้งรำข้าวปกาอำปิลสกัดน้ำมันทดแทนแป้งสาลี (ร้อยละ)				
	0 (สูตรที่ 1)	10 (สูตรที่ 2)	20 (สูตรที่ 3)	30 (สูตรที่ 4)	40 (สูตรที่ 5)
สี	7.78 ^a ±0.91	7.48 ^{ab} ±0.98	7.73 ^{ab} ±0.90	7.44 ^{ab} ±0.91	7.27 ^b ±0.89
กลิ่น ^{ns}	7.43±0.93	7.46±0.98	7.59±0.93	7.21±0.98	7.10±0.98
ลักษณะปรากฏ ^{ns}	7.59±0.93	7.40±0.98	7.50±0.98	7.37±0.98	7.36±0.98
รสชาติ ^{ns}	7.67±0.95	7.37±0.95	7.67±0.95	7.23±0.95	7.22±0.95
เนื้อสัมผัส ^{ns}	7.77±0.95	7.49±0.95	7.71±0.95	7.52±0.95	7.31±0.95
ความชอบโดยรวม ^{ns}	7.78±0.91	7.56±0.98	7.82±0.98	7.51±0.95	7.38±0.97

หมายเหตุ ตัวอักษรที่แตกต่างกันในแนวนอน แสดงค่าเฉลี่ยมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ ความเชื่อมั่นร้อยละ 95

^{ns} ไม่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับความเชื่อมั่นร้อยละ 95

ตารางที่ 4 ผลคุณลักษณะทางเคมี กายภาพ และจุลชีววิทยาของผลิตภัณฑ์ขนมทองม้วนแป้งรำข้าวปกอำปิลสกัดน้ำมันทดแทนแป้งสาลี

คุณภาพทางเคมี		สูตรพื้นฐาน (สูตรที่ 1)	ผลิตภัณฑ์ขนมทองม้วนแป้งรำข้าวปกอำปิลสกัดน้ำมันทดแทนแป้งสาลี ร้อยละ 20 (สูตรที่ 3)
ความชื้น ^{ns} (ร้อยละ)		2.20±0.98	2.54±0.98
โปรตีน ^{ns} (ร้อยละ)		8.10±0.95	8.65±0.95
ไขมัน (ร้อยละ)		20.99±0.95 ^a	23.01±0.95 ^b
เถ้า ^a (ร้อยละ)		0.97±0.98 ^a	2.18±0.98 ^b
คาร์โบไฮเดรต (ร้อยละ)		67.74±0.94 ^a	63.62±0.94 ^b
ใยอาหาร (ร้อยละ)		0.11±0.95 ^a	2.16±0.98 ^b
พลังงาน (กิโลแคลอรี)		492.27±0.98 ^a	496.17±0.98 ^b
สารแกมมา-โอโรซานอล (มิลลิกรัม/ 100 กรัม)		nd	0.031±0.05
กิจกรรมต้านอนุมูลอิสระ (มิลลิกรัม/ 100 กรัม)		nd	19.05±0.03
คุณภาพทางกายภาพ			
ค่าสี	L*	42.48±0.01 ^a	40.26±0.01 ^b
	a*	7.63±0.01 ^a	9.36±0.01 ^b
	b*	21.45±0.01 ^a	23.69±0.01 ^b
	ค่า a _w	0.389±0.98 ^a	0.429±0.98 ^b
คุณภาพทางจุลชีววิทยา			
เชื้อโคลิฟอร์ม (MPN/ กรัม)		<3.0	<3.0
ยีสต์และรา (โคโลนี/ กรัม)		<10	<10

หมายเหตุ ตัวอักษรที่แตกต่างกันในแนวนอน แสดงค่าเฉลี่ยมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับความเชื่อมั่นร้อยละ 95

^{ns} ไม่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับความเชื่อมั่นร้อยละ 95nd ตรวจแล้วไม่พบ

5. การวิเคราะห์คุณภาพทางกายภาพ

5.1 ค่าสี นำตัวอย่างผลิตภัณฑ์มาทำการวิเคราะห์ค่าสี โดยใช้เครื่องวัดสี (ยี่ห้อ Hunter Lab รุ่น Color Flex ประเทศสหรัฐอเมริกา) วางหัววัดกึ่งกลางขึ้นผลิตภัณฑ์เครื่องจะรายงานค่าสีที่ได้จากการวัดระบบ CIE L*a*b* a*

5.2 ปริมาณน้ำอิสระ (Water activity, a_w) วัดปริมาณน้ำอิสระของตัวอย่างผลิตภัณฑ์โดยใช้เครื่องวัดปริมาณน้ำอิสระในอาหาร (ยี่ห้อ Aqualab รุ่น Series 3 TE ประเทศสหรัฐอเมริกา) ควบคุมอุณหภูมิการวัดที่ 25 องศาเซลเซียส

6. การวิเคราะห์คุณภาพทางจุลชีววิทยา

หาจำนวนเชื้อโคลิฟอร์ม ยีสต์และรา ของตัวอย่างผลิตภัณฑ์ โดยใช้วิธีการเพาะเชื้อจากวิธีของ AOAC (2012)

3. ผลและวิจารณ์ผลการวิจัย

1. ผลการพัฒนาผลิตภัณฑ์ขนมทองม้วนแป้งรำข้าวปกอำปิลสกัดน้ำมันทดแทนแป้งสาลี ผลการพัฒนาผลิตภัณฑ์ขนมทองม้วนโดยการศึกษาปริมาณแป้งรำข้าวปกอำปิลสกัดน้ำมันทดแทนแป้งสาลี เลือกใช้สูตรพื้นฐานจากงานวิจัยของ Phattaraporn et al. (2015) ใช้ปริมาณแป้งรำข้าวปกอำปิลสกัดน้ำมันทดแทนแป้งสาลี 5 ระดับ คือ ร้อยละ 0, 10, 20, 30 และ 40 (สูตรที่ 1, 2, 3, 4

และ 5 ตามลำดับ) ได้ผลดังแสดงในตารางที่ 2 จากนั้นนำผลิตภัณฑ์ที่ได้ทดสอบทางประสาทสัมผัส โดยใช้ผู้ทดสอบกึ่งฝึกฝน จำนวน 30 คน ทำการทดสอบความชอบคุณลักษณะด้านสี กลิ่น ลักษณะปรากฏ รสชาติ เนื้อสัมผัส และความชอบโดยรวม ได้ผลดังแสดงในตารางที่ 3

จากตารางที่ 2 พบว่า ปริมาณแป้งรำข้าวปกอำปิลสกัดน้ำมันที่ใช้ทดแทนแป้งสาลีในผลิตภัณฑ์ขนมทองม้วนเมื่อมีปริมาณเพิ่มขึ้นจะส่งผลต่อลักษณะปรากฏของผลิตภัณฑ์ขนมทองม้วน โดยมีสีเข้มขึ้นเปลี่ยนจากสีเหลืองอมน้ำตาลเป็นสีน้ำตาลเข้มขึ้น ผิวหน้าเรียบสม่ำเสมอ มีการกระจายตัวของแป้งรำข้าวกระจายทั่วทั้งแผ่น เมื่อกัดมีความกรอบ หักง่าย และแน่นเนื้อมากขึ้น อาจเนื่องจากใยอาหารในแป้งรำข้าวปกอำปิลสกัดน้ำมันที่ทดแทนแป้งสาลีในผลิตภัณฑ์ขนมทองม้วนมีปริมาณมากขึ้นจึงขัดขวางการพอกตัวของเม็ดแป้งส่งผลให้ผลิตภัณฑ์มีความพองตัวได้น้อยลงเกิดความแน่นเนื้อ (Somwang L.,2019)

จากตารางที่ 3 พบว่า คะแนนเฉลี่ยการทดสอบความชอบด้านสีที่ผู้บริโภคมีต่อผลิตภัณฑ์ขนมทองม้วนแป้งรำข้าวปกอำปิลทดแทนแป้งสาลีทุกสูตร ผู้บริโภคให้ค่าคะแนนสูตรพื้นฐาน (สูตรที่ 1) มากที่สุด (7.78 คะแนน) มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับความเชื่อมั่นร้อยละ 95 เมื่อพิจารณาค่าคะแนนเฉลี่ยด้านความชอบโดยรวม พบว่า ผลิตภัณฑ์ขนมทองม้วนแป้งรำข้าวปกอำปิลทดแทนแป้งสาลีในอัตรา

ร้อยละ 20 (สูตรที่ 3) ได้รับคะแนนความชอบโดยรวมมากที่สุด (7.82 คะแนน) แม้ได้รับคะแนนความชอบโดยรวมไม่แตกต่างจากสูตรพื้นฐาน (สูตรที่ 1) แต่เมื่อพิจารณาเปรียบเทียบเฉพาะผลิตภัณฑ์ขนมทองม้วนที่มีการใช้แป้งรำข้าวปกาอำปิลสกัดน้ำมันทดแทนแป้งสาลี (สูตรที่ 2, 3, 4 และ 5) พบว่า สูตรที่ 3 ได้รับคะแนนความชอบด้านสี กลิ่น ลักษณะปรากฏ รสชาติ และเนื้อสัมผัส มากที่สุด ผู้วิจัยจึงเลือกสูตรที่ 3 เป็นสูตรที่เหมาะสมในการพัฒนาผลิตภัณฑ์ขนมทองม้วนแป้งรำข้าวปกาอำปิลสกัดน้ำมันทดแทนแป้งสาลีบางส่วน เพื่อศึกษาคุณลักษณะทางเคมี กายภาพ และจุลชีววิทยาในขั้นตอนต่อไป

2. 2. ผลการศึกษาคุณลักษณะทางเคมี กายภาพ และจุลชีววิทยา ผลการศึกษาคุณลักษณะทางเคมี กายภาพ และจุลชีววิทยา ของผลิตภัณฑ์ขนมทองม้วนแป้งรำข้าวปกาอำปิลสกัดน้ำมันทดแทนแป้งสาลีร้อยละ 20 (สูตรที่ 3) เปรียบเทียบกับสูตรพื้นฐาน (สูตรที่ 1) ได้ผลดังแสดงในตารางที่ 4

จากตารางที่ 4 ผลคุณภาพทางเคมีของผลิตภัณฑ์ขนมทองม้วนแป้งรำข้าวปกาอำปิลสกัดน้ำมันทดแทนแป้งสาลีร้อยละ 20 (สูตรที่ 3) พบว่า ผลิตภัณฑ์ปริมาณ 100 กรัม ให้พลังงาน 496.17 กิโลแคลอรี โดยมีความชื้น โปรตีน ไขมัน เถ้า คาร์โบไฮเดรต และใยอาหาร เท่ากับ 2.54, 8.65, 23.01, 2.18, 63.62 และ 2.16 ตามลำดับ เมื่อเปรียบเทียบกับสูตรพื้นฐาน (สูตรที่ 1) จะเห็นว่าปริมาณ ไขมัน เถ้า และใยอาหาร มีค่าเพิ่มขึ้นแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ที่ระดับความเชื่อมั่นร้อยละ 95 ทั้งนี้ อาจเป็นผลจากในรำข้าวสกัดน้ำมัน 100 กรัม จะมีปริมาณ ไขมัน 1.85–13.73 เถ้า 6.34–10.78 และ ใยอาหาร 5.78–13.51 (Ladda., 2017, Sukuntaros., 2017, Naiyana and Rewadee., 2002, Saunder., 1990 and Youssef., 1974) ซึ่งสูงกว่าในแป้งข้าวสาลี ที่มีปริมาณ ไขมัน 1.73–3.78 เถ้า 1.23–3.54 และ ใยอาหาร 0.59 – 1.31 (Fabian., et al, 2021 & Chopra., 2018) ดังนั้น เมื่อนำแป้งรำข้าวปกาอำปิลสกัดน้ำมันมาทดแทนแป้งสาลี จึงทำให้ผลิตภัณฑ์ขนมทองม้วนมีคุณค่าทางโภชนาการเพิ่มขึ้น สอดคล้องกับรายงานการวิจัยที่การใช้รำข้าวมาทดแทนแป้งสาลีในผลิตภัณฑ์ต่างๆ แล้วส่งผลให้ผลิตภัณฑ์นั้นๆ มีคุณค่าทางโภชนาการที่ดีขึ้น (Saengkhuae., 2019, Sujitta., 2018 & Jittra., 2016) และยังพบสารแอมมา-โอไรซานอล 0.031 มิลลิกรัม/ 100 กรัม และกิจกรรมต้านอนุมูลอิสระ 19.05 มิลลิกรัม/ 100 กรัม โดยสารแอมมา-โอไรซานอลและกิจกรรมต้าน

อนุมูลอิสระ เป็นสารออกฤทธิ์ทางชีวภาพ (Ferronato et al., 2021) จึงแสดงให้เห็นว่าผลิตภัณฑ์ขนมทองม้วนแป้งรำข้าว ปกาอำปิลสกัดน้ำมันทดแทนแป้งสาลีมีสมบัติยับยั้ง อนุมูลอิสระ สามารถลดความเสี่ยงต่างๆ ต่อโรค โดยเฉพาะโรคเรื้อรังที่สัมพันธ์กับอาหารและมีส่วนช่วยชะลอวัย (Moongngarm., et al, 2012)

คุณภาพทางกายภาพด้านค่าสี (L^*) ค่าความสว่าง (a^*) ค่าสีแดง และ (b^*) ค่าสีเหลือง และ ค่า a_w ของผลิตภัณฑ์ขนมทองม้วนแป้งรำข้าวปกาอำปิลสกัดน้ำมันทดแทนแป้งสาลี ร้อยละ 20 (สูตรที่ 3) เทียบกับสูตรพื้นฐาน (สูตร 1) มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับความเชื่อมั่น ร้อยละ 95 โดยปริมาณแป้งรำข้าวปกาอำปิลสกัดน้ำมันมีผลทำให้ค่าความสว่าง (L^*) มีค่าลดลงจาก 42.48 เป็น 40.26 ผลิตภัณฑ์ที่มีสีเข้มเพิ่มขึ้น สอดคล้องกับภาพและผลในตารางที่ 3 และ 4 เนื่องจากแป้งรำข้าวปกาอำปิลสกัดน้ำมันมีลักษณะเป็นผงละเอียดสีน้ำตาลอ่อนเมื่อเทียบกับแป้งสาลีที่มีสีขาว เมื่อได้รับความร้อนจึงทำให้สีของผลิตภัณฑ์มีแนวโน้มสีน้ำตาลเข้มขึ้น (Saunders, 1990) ค่า a_w ของผลิตภัณฑ์ขนมทองม้วนแป้งรำข้าวปกาอำปิลสกัดน้ำมันทดแทนแป้งสาลี ร้อยละ 20 (สูตรที่ 3) เทียบกับสูตรพื้นฐาน (สูตรที่ 1) มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับความเชื่อมั่น ร้อยละ 95 มีค่าเฉลี่ยเท่ากับ 0.429 และ 0.389 ซึ่งจะเห็นว่า ค่า a_w มีค่าสูงขึ้นตามปริมาณแป้งรำข้าวสกัดน้ำมันที่ใช้เพราะในรำข้าวสกัดน้ำมันมีไขมัน ร้อยละ 1.85–13.73 ส่งผลให้ผลิตภัณฑ์มีคุณสมบัติในการอุ้มน้ำเพิ่มขึ้น (Sorat, W., & Pakawadee, P., 2017) ดังนั้นผลิตภัณฑ์จึงจัดให้เป็นอาหารกลุ่มที่มีความชื้นต่ำสามารถเก็บรักษาได้นาน เนื่องจากมีค่า a_w ต่ำกว่า 0.6 ซึ่งอยู่ในระดับที่เชื้อจุลินทรีย์ไม่สามารถเจริญเติบโตได้ (Wiwat, 2010) และคุณภาพทางจุลชีววิทยาของผลิตภัณฑ์ขนมทองม้วนทั้งสองสูตรมีจำนวนโคลิฟอร์ม น้อยกว่า 3.0 MPN/ กรัม จำนวนยีสต์และราน้อยกว่า 10 โคโลนี/ กรัม ซึ่งอยู่ในมาตรฐานผลิตภัณฑ์ชุมชนขนมไทย (มผช. 1/2552)

4. สรุปผลการทดลอง

การพัฒนาผลิตภัณฑ์ขนมทองม้วนแป้งรำข้าวปกาอำปิลสกัดน้ำมันทดแทนแป้งสาลี สามารถสรุปได้ว่า แป้งรำข้าวปกาอำปิลสกัดน้ำมัน ร้อยละ 20 ของน้ำหนักแป้งสาลี เป็นสัดส่วนที่ความเหมาะสมที่สุด เนื่องจากเป็นปริมาณที่ใช้แป้งรำข้าวปกาอำปิลสกัดน้ำมันทดแทนแป้ง

สาธิตได้มากที่สุด โดยที่ผู้ทดสอบความชอบให้คะแนนความชอบโดยรวมมากที่สุด (7.82 คะแนน) นอกจากนี้ผลิตภัณฑ์ยังมีองค์ประกอบหลักสูงขึ้น คือ ไขมัน เกล็ด และใยอาหาร และพบสารออกฤทธิ์ทางชีวภาพ คือ สารแกมมา-โอไรซานอล 0.031 มิลลิกรัม/ 100 กรัม และกิจกรรมต้านอนุมูลอิสระ 19.05 มิลลิกรัม/ 100 กรัม ซึ่งไม่พบในผลิตภัณฑ์ขนมทองม้วนสูตรพื้นฐาน จึงสรุปว่าแป้งรำข้าวเปลือกอัดน้ำมันที่นำมาใช้ทดแทนแป้งสาลีสามารถเพิ่มคุณค่าทางโภชนาการให้กับผลิตภัณฑ์ขนมทองม้วนได้

กิตติกรรมประกาศ

ขอขอบคุณสำนักวิจัยและพัฒนา มหาวิทยาลัยราชภัฏสุรินทร์ที่ให้การสนับสนุนทุนวิจัย ประจำปีงบประมาณ พ.ศ. 2563

เอกสารอ้างอิง

- Gani, A., Wani, S. M., Masoodi, F. A., & Hameed, G. (2012). Whole-grain cereal bioactive compounds and their health benefits: A review. *J Food Process Technol*, 3(3), 146–56. <https://doi.org/10.4172/2157-7110.1000146>
- AOAC. (2012). *Official Method of Analysis* (19th ed.). Washington DC, USA.
- Chopra, N., Rani, R., & Singh, A. (2018). Physico-nutritional and sensory properties of cookies formulated with quinoa, sweet potato and wheat flour blends. *Current Research in Nutrition and Food Science Journal*, 6(3), 798–806.
- Ugwuona, F., Ukom, A., Ejinkonye, B., Obeta, N., & Ojinnaka, M. (2023). Exploring the possibilities of some selected flour blends for the development of bakery products: Comparison with some physicochemical and functional properties of wheat flour. *Food Science and Technology International*, 29(2), 105–114.
- Ferronato, A. N., Rossi, R., Pinto, L. M. N., & Garavaglia, J. (2021). Development of a freeze-dried symbiotic obtained from rice bran. *Biotechnology Reports*, 30, e00636.
- Singthong, J. (2016). Functional Properties and Utilization of Rice Bran Dietary Fiber in Bread. *Journal of Science and Technology, Ubon Ratchathani University*, 18(2), 63–73.
- Khongsak, S. (2010). Simple Process to Increase Yield and Improve Qualities of Cold Pressed Rice Bran Oil. *Area Based Development Research Journal*, 2(6), 43–52. (in Thai)
- Kwaw, E., Ma, Y., Tchabo, W., Apaliya, M. T., Wu, M., Sackey, A. S., & Tahir, H. E. (2018). Effect of lactobacillus strains on phenolic profile, color attributes and antioxidant activities of lactic-acid-fermented mulberry juice. *Food chemistry*, 250, 148–154.
- Ladda, S. W. (2017). Rice Bran Protein. *Food*, 43(7), 31–38.
- Moongarm, A., Daomukda, N., & Khumpika, S. (2012). Chemical compositions, phytochemicals, and antioxidant capacity of rice bran, rice bran layer, and rice germ. *Apcbee Procedia*, 2, 73–79.
- Naiyana, B., & Rewadee, C. (2002). *Rice Bran Oil, a Healthy Alternative for Thai People* (1st ed.). Bangkok, Thailand: Odeon Store Publishing.
- Onanong, N. (2007). *Rice: Science and Technology* (2th ed.). Bangkok, Thailand: Kasetsart University.
- Perretti, G., Miniati, E., Montanari, L., & Fantozzi, P. (2003). Improving the value of rice by-products by SFE. *The Journal of supercritical fluids*, 26(1), 63–71.
- Phattaraporn, K., Pradub, R., Wattanachai, Y., Worasitkulya, T., & Kitti, N. (2015). *Economic Value added of Local Rice (Oryza sativa Linn.) of Surin Province by Application of Germinated Brown Rice in Thai Desert and Cosmetics*. (Research report). Faculty of Science and Technology, Surindra Rajabhat University. Surin. (in Thai)
- Jamuna, P., & Ramanatham, G. (1995). Proximate composition and protein quality of stabilized rice bran. *Journal of Food Science and Technology*, 32(5), 416–419.
- Saengkhae, S. (2019). Development of Bread with Riceberry Rice Bran. *RMUTP Research Journal*, 13(2), 186–195.
- Sarawut, J. (2021). The Study of the Starch Production Process and the Ratios of Sinlek Rice Flour in the Development of Fortified Cereal Sinlek Rice Crispy Crepe. In *The 9th National Suan Sunandha Conference 2021, 17–18 June 2021*, (pp.720–729). Suan Sunandha Rajabhat University, Bangkok. (in Thai)
- Saunders, R. M. (1990). The properties of rice bran as a foodstuff. *Cereal foods world*, 35(7), 632–636.
- Somwang, L., Paramee, N., Sujinda, B., & Pattama, C. (2019). Effect of Substitution of Wheat Flour with Sangyod Brown Rice Flour on Physicochemical and Sensory Qualities in Snack Product. *Khon Kaen Agriculture Journal*. 47(SUPPL. 1), 679 – 684.
- Inket, S. W., & Phugan, P. (2017). Development of Cracker Supplemented with Jasmine Rice Bran Powder. *SDU Research Journal Sciences and Technology*, 10(1), 121–136.
- Sujitta, R., Kanokan, J., & Obcheuy, W. (2018). Development of sandwich bread formulation substituted wheat flour with riceberry rice bran. *Research Journal Phranakhon Rajabhat: Science and Technology*, 13(1), 123–138.
- Sukuntaro, T. (2017). *Rice Bran Protein with Functional and Biological Properties*. (Research report). Kasetsart University Research and Development Institute. Kasetsart University, Bangkok. (in Thai)
- Sunan, B., Sarinee, S., & Roongtip, C. (2016). *Development of Thong Moul (Thai Desserts) from Mun-Pu Brown Rice and Jerusalem Artichoke supplemented with Colorant of Plant*. (Research report). Division of Food Technology, Faculty of Agricultural Technology, Rajabhat Maha Sarakham University. Maha Sarakham. (in Thai)
- Supavadee, R. (2015). *Product Development of Soft Thai Rice Flour Pancake Rood (Thong Muan Sod) form Sinin Rice Flour*. (Research report) Division of Food Science and Technology, Faculty of Agricultural Technology and Agro-Industry, Rajamangala University of Technology Suvarnabhumi, Bangkok. (in Thai)
- Thai Industrial Standards Institute. (2009). *Thai Community Product Standard: Thai Dessert (TCPS: 1 / 2 5 5 2)*. Bangkok, Thailand: Thai Industrial Standards Institute.
- Theerarat, C., Sumran, P., Naphaporn, W., Ketjit, K., & Somchai, C. (2019). A Quality of Local Rice Seed Varieties in Maha Sarakham Province and Neighboring Provinces at Different Storability. *Prawarun Agricultural Journal*, 16(1), 91–104.
- Wiwat, W. (2010). Water Activity and Shelf Life of Dry Food. *Food*, 40 (4), 277–281.
- Youssef, A. M., El-Fouly, M. M., & El-Baz, F. K. (1974). Isolation and chemical composition of protein concentrates from soya-bean, rice-bran and protelan. *Qualitas Plantarum*, 24, 71–84.