

การจำลองดัชนี SET50 แห่งประเทศไทยด้วย แบบจำลองมาร์คอฟซ่อนเร้นผสมแบบเกาส์หลายตัวแปร

Thailand SET50 Simulation with Multivariate Gaussian Mixture Hidden Markov Model

อัครา กิจการเจริญสิน^{1*}
Akara Kijkarncharoensin^{1*}

Received: 26 July 2024
Revised: 25 February 2025
Accepted: 21 May 2025

บทคัดย่อ

การลงทุนในตลาดทุนจำเป็นต้องพิจารณาทั้งอัตราผลตอบแทนและความเสี่ยงที่เกิดกับการลงทุน การจำลองการแจกแจงราคาของสินทรัพย์ในช่วงเวลาข้างหน้าเป็นประโยชน์ต่อนักลงทุนเพราะสามารถทำนายทั้งอัตราผลตอบแทนและความเสี่ยงที่อาจเกิดขึ้น บทความนี้นำเสนอแบบจำลองมาร์คอฟซ่อนเร้นผสมแบบเกาส์ชนิดหลายตัวแปรมาจำลองดัชนี SET50 รายวันตั้งแต่ต้นเดือน ธ.ค. 2564 ถึงปลายเดือน ม.ค. 2567 โดยผลลัพธ์ที่สังเกตได้ซึ่งอยู่ในรูปแบบอนุกรมเวลาเป็นเวกเตอร์สามมิติของอัตราผลตอบแทนที่วัดจากราคาเปิดไปยังราคาปิด ราคาสูงสุดและราคาต่ำสุดภายในวันเดียวกัน งานวิจัยนี้เปรียบเทียบการแจกแจงดัชนีราคาวันสุดท้าย ซึ่งจำลองตามแบบจำลองมาร์คอฟซ่อนเร้นผสมแบบเกาส์ชนิดหลายตัวแปรกับชนิดตัวแปรเดียว ผลการจำลองดัชนีราคาในหนึ่งปีข้างหน้าจำนวน 100,000 เส้นทาง ยืนยันความถูกต้องในผลการทำนายดัชนี SET50 โดยค่า MAPE ของแบบจำลองมาร์คอฟซ่อนเร้นผสมแบบเกาส์ชนิดหลายตัวแปรมีค่าเพียง 1.7409 เทียบกับ 2.7261 ของแบบจำลองชนิดตัวแปรเดียวการทดสอบ Wilcoxon Rank Sum ยืนยันการแจกแจงที่ต่างกันในระดับที่จำลองขึ้นทั้งสองแบบด้วย $p\text{-value} < 0.001$ นอกจากนี้การทดสอบ Kruskal-Wallis แสดงว่าในช่วงข้อมูลที่ทดสอบนั้นแบบจำลองมาร์คอฟซ่อนเร้นผสมแบบเกาส์ปราศจากอคติด้านบวกอย่างที่พบในแบบจำลองบราวน์เรขาคณิตทั้งหนึ่งมิติและหลายมิติ

คำสำคัญ: มาร์คอฟซ่อนเร้น ผสมแบบเกาส์ การจำลอง SET50 บราวน์เรขาคณิต

¹ คณะบริหารธุรกิจและเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีราชมงคลตะวันออก เขตพื้นที่จันทบุรีจันทบุรี กรุงเทพมหานคร 10400

¹ Faculty of Business Administration and Information Technology, Rajamangala University of Technology Tawan-Ok, Chakrabongse Bhuvanarth Campus, Bangkok, Thailand 10400

*ผู้ติดต่อประสานงาน (Corresponding author) e-mail: Akara_Ki@rmutto.ac.th

ABSTRACT

Investment in the capital market necessitates a dual consideration of returns and risks. Asset price distribution simulations for future periods benefit investors as they allow for predictions of both returns and potential risks. This article introduces a Multivariate Gaussian Mixture Hidden Markov Model to simulate the daily SET50 index in Thailand from December 2021 to January 2024. The observed outcomes are presented as three-dimensional time series vectors measuring returns from Open to Close and the day's High and Low prices. This research compared the distributions of the closing price on the last date simulated by the multivariate model against those of a univariate model. Over 100,000 simulated paths, the multivariate model demonstrated superior predictive accuracy with a Mean Absolute Percentage Error (MAPE) of 1.7409 compared to 2.7261 for the univariate model. The Wilcoxon Rank Sum test affirmed the statistical difference in their distributions with a p-value <0.001 . Additionally, the Kruskal-Wallis test indicated that the multivariate model was free from the positive bias observed in both one-dimensional and multidimensional Geometric Brownian Motion models.

Keywords: Hidden Markov; Gaussian Mixture; Simulation; SET50; Geometric Brownian Motion

บทนำ

ตลาดเงินและตลาดทุนเป็นแหล่งเงินออมที่สำคัญเพื่อรับมือกับสังคมผู้สูงอายุ อัตราผลตอบแทนและความผันผวนในราคาสินทรัพย์เสี่ยงจึงส่งผลต่อเงินออมของนักลงทุน ดัชนี SET50 ของตลาดหลักทรัพย์แห่งประเทศไทยปรับตัวลดลงตั้งแต่วิกฤติโควิด-19 และไม่ฟื้นตัวกลับที่ระดับเดิม (21 กุมภาพันธ์ 2568) ดังนั้นการจำลองดัชนี SET50 จึงสำคัญมากในการบริหารความเสี่ยงและเพิ่มผลตอบแทนเพื่อปกป้องเงินออมและเตรียมพร้อมกับสังคมผู้สูงอายุ

การเรียนรู้ถูกนำมาใช้เพื่อหาผลตอบแทนในดัชนี SET50 แห่งประเทศไทย โดยพบว่าหากใช้ราคาห้าวันทำการย้อนหลังเป็นตัวแปรต้นของแบบจำลอง Support Vector Machine (SVM) และข้อมูล Training ครอบคลุมพิสัยของการขึ้น-ลงของดัชนีในช่วงที่พิจารณาแล้วแบบจำลอง SVM จะทำนายดัชนี SET50 ได้อย่างแม่นยำ [1] อย่างไรก็ตามแบบจำลอง SVM เป็นการเรียนรู้ประเภท Supervised learning ความถูกต้องของการทำนายขึ้นอยู่กับโครงสร้างของสมการที่ใช้ นอกจากนี้แบบจำลองให้ผลการทำนายเฉพาะค่าคาดหวัง แต่การป้องกันความเสี่ยงจำเป็นต้องรู้การแจกแจงของดัชนีราคาเพื่อจำลองสถานการณ์ที่เลวร้ายที่สุด

แบบจำลองการเคลื่อนที่ของบราวน์เรขาคณิต (Geometric Brownian Motion, GBM) สามารถใช้จำลองการแจกแจงของดัชนีราคาหุ้น [2] ภายใต้สมมติฐานว่าราคาหุ้นเคลื่อนที่ตาม Stochastic Process และมีการแจกแจงแบบ Log-Normal การเรียนรู้ตามแบบจำลองนี้จึงจัดเป็น Supervised learning อีกชนิดหนึ่งที่บ่งชี้โดยนัยว่าค่าคาดหวังของราคาหุ้นเป็นไปตามฟังก์ชันเอ็กซ์โพเนนเชียลซึ่งหมายความว่าในระยะยาวราคาหุ้นควรเพิ่มขึ้นอย่างต่อเนื่อง

อย่างไรก็ตามดัชนี SET50 ณ วันที่ 21 กุมภาพันธ์ 2568 อยู่ที่ระดับ 795.09 และยังไม่ฟื้นตัวกลับไปสู่ระดับ 1,159.73 ก่อนเกิดวิกฤต COVID-19 แม้เวลาจะผ่านไปกว่าห้าปี ความแม่นยำของการจำลองดัชนี SET50 ผ่าน GBM จึงถูกตั้งคำถาม การศึกษาเพิ่มเติมจึงจำเป็นเพื่อพัฒนาแบบจำลองที่สามารถจำลองดัชนี SET50 ให้สอดคล้องกับค่าที่เกิดขึ้นจริงมากขึ้น

แบบจำลองมาร์คอฟซ่อนเร้น (Hidden Markov Model, HMM) เป็นการเรียนรู้ของเครื่องมือประเภท Unsupervised learning ที่ไม่ขึ้นกับโครงสร้างของสมการที่ใช้เรียนรู้ มีลักษณะที่น่าสนใจคือแบบจำลองพิจารณาว่าผลลัพธ์ที่สังเกต (Observed outcome) ได้มาจากแหล่งกำเนิดสัญญาณหรือสถานะที่ซ่อนอยู่ (Hidden state) หากพิจารณาว่าแหล่งกำเนิดสัญญาณมีมากกว่าหนึ่งทำให้ผลลัพธ์ที่สังเกตได้เกิดจากการผสมของการแจกแจงแบบเกาส์ (Gaussian Mixture) แล้วแบบจำลองเช่นนี้มีชื่อว่าแบบจำลองมาร์คอฟซ่อนเร้นผสมแบบเกาส์ (Gaussian Mixture Hidden Markov Model, gmHMM) [3] ผลลัพธ์ที่สังเกตได้ตามแบบจำลอง gmHMM จึงสุ่มตามการแจกแจงผสมแบบเกาส์

อัครา กิจการเจริญสินและปราณี มณีรัตน์ [4] ประยุกต์ใช้แบบจำลอง gmHMM เพื่อเลียนแบบดัชนี Thailand SET50 โดยปราศจากอคติด้านบวกที่พบใน GBM ได้สำเร็จ ทั้งคู่วัดอัตราผลตอบแทนรายวันจากสัดส่วนราคาเปิดและราคาปิดทำให้ผลลัพธ์ที่สังเกตได้มีการแจกแจงแบบตัวแปรเดี่ยว (Univariate distribution) จากนั้นใช้ gmHMM จำลองการเคลื่อนไหวของดัชนีและเปรียบเทียบการแจกแจงของดัชนีราคาที่จำลองขึ้น 10,000 ครั้งกับแบบจำลอง GBM ชนิดตัวแปรเดี่ยว เนื่องจากการแจกแจงของดัชนีที่จำลองขึ้นไม่ได้เป็นแบบปกติ นักวิจัยจึงใช้การทดสอบแบบ non-parametric ของ Kruskal-Wallis และพบว่าผลการแจกแจงดัชนีราคาสุดท้ายจาก gmHMM แตกต่างจาก GBM ชนิดตัวแปรเดี่ยวอย่างมีนัยสำคัญทางสถิติที่ $p\text{-value} < 0.01$ อย่างไรก็ตาม การจำลองตัวแปรเดี่ยวจากงานวิจัยนี้ไม่ครอบคลุมข้อมูลราคาทั้งหมด ณ สิ้นวันทำการ โดยดัชนี SET50 ในหนึ่งวันประกอบด้วยราคาเปิด (Open) ราคาปิด (Close) ราคาสูงสุด (High) และราคาต่ำสุด (Low) ซึ่งรวมเรียกว่าราคา OHLC การวัดอัตราผลตอบแทนโดยใช้เพียงราคา Open-Close อาจละเลยความสัมพันธ์ (correlation) ระหว่างอัตราผลตอบแทนดังกล่าวกับอัตราผลตอบแทนที่วัดจากราคา Open-High และ Open-Low การวัดอัตราผลตอบแทนที่แม่นยำยิ่งขึ้นจึงควรอยู่ในรูปของเวกเตอร์อัตราผลตอบแทน (Open-Close, Open-High, Open-Low) หรือเวกเตอร์ราคา (Open, High, Low, Close) ซึ่งช่วยสะท้อนโครงสร้างของข้อมูลราคาที่สมบูรณ์ยิ่งขึ้น Gupta & Dhingra [5] ใช้เวกเตอร์อัตราผลตอบแทนในการสร้างแบบจำลอง HMM ขณะที่ Nguyen [6] เลือกใช้เวกเตอร์ของราคาดิบแทน

อย่างไรก็ตามวรรณกรรมที่ผ่านมายังไม่มีการศึกษาเรื่องการจำลองดัชนีราคาผ่านแบบจำลอง HMM แบบหลายตัวแปร งานวิจัยนี้จึงตั้งคำถามว่าหากใช้ข้อมูลทั้งสี่ประเภทในการจำลองดัชนี SET50 ความถูกต้องในแง่ของค่าคาดหวังและความแปรปรวนจะเพิ่มขึ้นหรือไม่เมื่อเปรียบเทียบกับแบบจำลองที่ใช้เพียงราคาเปิดและปิด บทความนี้นำเสนอแนวทางใหม่ในการจำลองดัชนี Thailand SET50 โดยใช้เวกเตอร์อัตราผลตอบแทน (Open-Close, Open-High, Open-Low) แทนการใช้ตัวแปรเดี่ยวแบบ อัครา กิจการเจริญสินและปราณี มณีรัตน์ [4] แบบจำลอง gmHMM ถูกนำมาใช้เพื่อเลียนแบบการเคลื่อนไหวของดัชนี Thailand SET50 และเปรียบเทียบการแจกแจงของดัชนีราคาสุดท้ายกับแบบจำลอง GBM ชนิดหลายตัวแปรซึ่งเหมาะสมกว่ารูปแบบตัวแปรเดี่ยว จำนวนการจำลองถูกเพิ่มจาก 10,000 เป็น 100,000 ครั้งเพื่อเพิ่มความแม่นยำ ขณะที่การทดสอบความแตกต่างของการแจกแจงยังคงใช้ Kruskal-Wallis แบบ non-parametric เช่นเดิม บทความนี้คาดการณ์ว่าแบบจำลอง gmHMM แบบหลายตัวแปรจะให้ผลลัพธ์ที่ถูกต้องกว่ารูปแบบตัวแปรเดี่ยวโดยใช้ค่าคาดหวังของดัชนี SET50 ที่จำลองขึ้นเปรียบเทียบกับค่าที่เกิดขึ้นจริงเป็นตัวชี้วัดความแม่นยำ

เนื้อหาในส่วนต่อไปของบทความคือทฤษฎีที่เกี่ยวข้อง วิธีการทดลอง ผลลัพธ์และอภิปราย และจบบทความด้วยการสรุปผล

ทฤษฎีที่เกี่ยวข้อง

การบริหารความเสี่ยงมุ่งเน้นการจัดการการขาดทุนที่อาจเกิดจากการแจกแจงที่มีโอกาสเกิดเหตุการณ์สุดโต่งสูง (Fat Tail Distribution) ขณะที่การลงทุนให้ความสำคัญกับผลตอบแทน ดังนั้นการจำลองการแจกแจงของราคาสินทรัพย์เสี่ยงจึงมีความสำคัญต่อการตัดสินใจทั้งในการบริหารความเสี่ยงและการลงทุนโดยอ้างอิงจากค่าความน่าจะเป็นการเคลื่อนที่แบบบราวน์เรขาคณิต (Geometric Brownian Motion, GBM) เป็นหนึ่งในวิธีที่ได้รับความนิยมในการจำลองราคาสินทรัพย์เสี่ยง โดยอ้างอิงจากสมมติฐานว่าราคาสินทรัพย์เป็นไปตามกระบวนการสโตแคสติก (Stochastic Process) ตามสมการ (1)

$$dX(t) = \Theta(t)X(t)dt + \sigma(t)X(t)dW(t) \quad (1)$$

สมการ(1) อยู่ในรูปสมการเชิงอนุพันธ์สโตแคสติก(Stochastic Differential Equation, SDE) คอลัมน์เวกเตอร์ของการเปลี่ยนแปลงราคาสินทรัพย์เสี่ยงหรือ $dX(t)$ คือ $(dX_1(t), dX_2(t), dX_3(t))'$ โดยที่ $dX_i(t)$ หมายถึงการเปลี่ยนแปลงราคาในสินทรัพย์ i

บทความนี้กำหนดให้ผลลัพธ์ที่สังเกตได้ ณ เวลา t ใด ๆ (O_t) และลำดับของผลลัพธ์ทั้งหมดที่สังเกตได้ ($\mathbb{O}$) ตั้งแต่ $t = 1, \dots, T$ คือ

$$\mathbb{O} = \{O_t = (r_{OC}, r_{OH}, r_{OL})'_t; t = 1, \dots, T\} \quad (2)$$

โดยนิยามให้ $r_{OC} = \frac{Close-Open}{Open} r_{OH} = \frac{High-Open}{Open} r_{OL} = \frac{Open-Low}{Open}$ และ ' หมายถึง transpose vector แทนสมการ (2) ลงในนิยามของ $dX(t)$ ทำให้ $dX(t)$ คือเวกเตอร์ของอัตราผลตอบแทนวัดจากราคาเปิดหรือ $(r_{OC}, r_{OH}, r_{OL})'_t$

พารามิเตอร์ $\Theta(t)$ และ $\sigma(t)$ หมายถึงเทอมไดรฟต์และเทอมดิฟฟิวชันตามลำดับ ดังนั้น $\Theta(t)$ คือเวกเตอร์ของค่าแนวโน้มของการเคลื่อนไหวราคาสินทรัพย์เสี่ยงในระยะยาว และ $\sigma(t)$ คือเมทริกซ์ความผันผวนระหว่างของ r_{OC}, r_{OH} และ r_{OL} ที่เวลา t ส่วน $dW(t)$ คือเวกเตอร์ของ Weiner process แบบหลายตัวแปรซึ่งทำให้ $dX(t)$ มีค่าสุ่ม

ผลเฉลยของสมการ (1) อยู่ในรูปของฟังก์ชันเอ็กซ์โพเนนเชียล ซึ่งหมายความว่าตามแบบจำลอง GBM ราคาสินทรัพย์เสี่ยงจะเติบโตในลักษณะเอ็กซ์โพเนนเชียลในระยะยาว อย่างไรก็ตามดัชนี SET50 ยังคงไม่ฟื้นตัวสู่ระดับก่อนวิกฤต Covid-19 ทำให้การจำลองดัชนี SET50 จำเป็นต้องได้รับการศึกษาเพิ่มเติม

ในทางตรงกันข้ามแบบจำลอง gmHMM ไม่ได้กำหนดรูปแบบการเคลื่อนที่ของราคาอย่างตายตัวเช่นเดียวกับ GBM แต่พิจารณาว่าค่าของ $\mathbb{O}$ ถูกกำหนดโดยความน่าจะเป็นของแหล่งกำเนิดสัญญาณใน N สถานะซ่อนเร้น (Hidden Regimes) นอกจากนี้ gmHMM ยังใช้การแจกแจงแบบเกาส์ผสมจำนวน M รูปแบบเพื่อจำลองสัญญาณส่งออกในแต่ละสถานะซ่อนเร้นทำให้ gmHMM ไม่มีข้อจำกัดของโครงสร้าง SDE เช่นเดียวกับ GBM ทำให้สามารถจำลองการเปลี่ยนแปลงของราคาสินทรัพย์ได้อย่างยืดหยุ่นมากขึ้น

ในแต่ละช่วงเวลา t นั้น O_t ใน gmHMM จะขึ้นอยู่กับความน่าจะเป็นเริ่มต้นของสถานะ (Initial Probabilities, π) ความน่าจะเป็นการเปลี่ยนสถานะ (Transition Probabilities, A) และการแจกแจงของสัญญาณส่งออก (Emission Distribution, F) นอกจากนี้สำหรับ $\forall i, j = 1, \dots, N$ และ $\forall k = 1, \dots, M$ แล้วเวกเตอร์ $\pi = \{\pi_i\}$ คือ ความน่าจะเป็นที่ระบบจะเริ่มต้นที่สถานะซ่อนเร้น i ก่อนเริ่มกระบวนการเมทริกซ์

$A = \{a_{ij}\}$ คือ ความน่าจะเป็นที่ระบบจะเปลี่ยนจากสถานะซ่อนเร้น i เป็น j และเมทริกซ์แนวทแยง $F = \{f_j(O_t)\}$ คือสัญญาณส่งออกของสถานะซ่อนเร้น j ที่มีการแจกแจงแบบเกาส์ผสม

แบบจำลอง gmHMM กำหนดให้ค่า Likelihood L ของโมเดล λ ที่มีต่อ $\mathbb{O}$ หรือ $L(\mathbb{O}|\lambda)$ มีค่าขึ้นอยู่กับ 1) ค่า Forward probability นับตั้งแต่เวลาเริ่มต้นจนถึง t โดย O_t เกิดจากสถานะ $i = 1, \dots, N$ หรือ $\alpha_t(i)$ 2) ค่า Backward probability นับตั้งแต่เวลา T ย้อนมาถึง $t + 1$ โดย O_t จะเข้าสู่สถานะ $j = 1, \dots, N$ หรือ $\beta_{t+1}(j)$ 3) ค่า Transition probability ซึ่งแสดงถึงความน่าจะเป็นที่จะเปลี่ยนสถานะจาก i ไป j หรือ a_{ij} และ 4) ค่า Density ของ O_t ณ สถานะ j หรือ $f_j(O_t)$ ดังนั้นแสดง $L(\mathbb{O}|\lambda)$ ได้ใน (3)

$$L(\mathbb{O}|\lambda) = \sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) a_{ij} f_j(O_{t+1}) \beta_{t+1}(j) \quad (3)$$

การแจกแจงผสมแบบเกาส์กำหนดให้ $f_j(O_t)$ มาจากการแจกแจงแบบเกาส์ผสมกัน M รูปแบบ ให้ $b_{jk}(O_t) \sim \mathcal{N}(\mu_k, \Sigma_k)$ คือค่า Density ของการแจกแจงแบบเกาส์รูปแบบ $k = 1, \dots, M$ ของสถานะ $j = 1, \dots, N$ โดย μ_k และ Σ_k คือค่า mean และ covariance matrix ของการแจกแจงแบบเกาส์หลายตัวแปรแบบที่ k ให้เวกเตอร์ $C = \{c_{jk}\}$ คือค่าน้ำหนักของ $b_{jk}(O_t)$ ที่มีต่อ $f_j(O_t)$ นั่นคือ c_{jk} แสดงถึงค่าความน่าจะเป็น (p_k) ที่จะได้การแจกแจงแบบเกาส์รูปแบบที่ k จากทั้งหมด M รูปแบบ ดังนั้นค่า Density ของการแจกแจงผสมแบบเกาส์คือ

$$f_j(O_t) = \sum_{k=1}^M c_{jk} b_{jk}(O_t) \quad (4)$$

จาก (3) และ (4) จะได้ว่าพารามิเตอร์ของแบบจำลอง gmHMM คือ

$$\lambda = \{a_{ij}, c_{jk}, \mu_k, \Sigma_k; i, j = 1, \dots, N; k = 1, \dots, M\}$$

การประมาณค่า λ จะกระทำผ่าน Baum-Welch Algorithm [7] ซึ่งเทียบเท่าการประมาณด้วย Maximum likelihood [8] จากนั้น Viterbi algorithm [7] ใช้เพื่อบ่งบอกสถานะที่ซ่อนอยู่โดยนัยของผลลัพธ์ที่สังเกตได้ ณ เวลา t

ผลลัพธ์ที่สังเกตได้ ณ $t + 1$ หรือ O_{t+1} สามารถจำลองขึ้นตามค่าความน่าจะเป็น เมื่อกำหนดให้สินทรัพย์อยู่สถานะซ่อนเร้น i ที่เวลา t ผ่านการสุ่มหาสถานะซ่อนเร้น j และการแจกแจงแบบเกาส์รูปแบบ k อ้างอิงจากค่าความน่าจะเป็น A และ C จากนั้นการแจกแจงของ O_{t+1} จะเป็นตามสมการ (5)

$$O_{t+1} \sim \mathcal{N}(\mu_k, \Sigma_k; \text{state } t + 1 = j, k \sim \text{multi}(M, p_k \in P)) \quad (5)$$

การจำลองตาม (5) เป็นสาระสำคัญที่บทความนี้มุ่งเน้นนำเสนอ งานวิจัยก่อนหน้านี้มุ่งเน้นการซื้อขายหุ้น หรือคัดเลือกหุ้นตามสถานะซ่อนเร้น [5,6] ส่วนงานวิจัยก่อนหน้านี้ [4] เป็นการจำลองแบบตัวแปรเดียวที่ไม่ได้พิจารณาความสัมพันธ์ระหว่างสมาชิกของเวกเตอร์ O_t บทความนี้เป็นงานวิจัยแรกที่ใช้ gmHMM แบบหลายตัวแปรจำลอง O_{t+1} โดยคำนึงถึงความสัมพันธ์ระหว่าง $r_{OC}r_{OH}$ และ r_{OL} การที่ O_{t+1} เป็นตัวแปรสุ่มทำให้การจำลองแบบสุ่มตาม (5) สามารถแสดงการแจกแจงของ O_{t+1} ผ่านกราฟฮิสโตแกรมส่วนค่าเฉลี่ยและการแจกแจงสุดโต่งสูงบ่งบอกถึงผลตอบแทนคาดหวังและความเสี่ยงของการลงทุนตามลำดับ

ดัชนี Mean Absolute Percentage Error (MAPE) ใช้วัดความถูกต้องของค่าคาดหวังของการแจกแจงที่จำลองขึ้นเทียบกับค่าดัชนี SET50 ที่เกิดจริงให้ $r_{OC} = \frac{\text{Open} - \text{Close}}{\text{Open}}$ แล้วจะได้ว่า

$$MAPE = \frac{|r_{OC} - r_{OC}|}{r_{OC}} \times 100 \quad (6)$$

MAPE ที่ต่ำบ่งบอกถึงการทำนายที่แม่นยำหาก gmHMM ชนิดหลายตัวแปรมีความถูกต้องกว่าชนิดตัวแปรเดียวแล้วงานวิจัยคาดหวังว่า gmHMM ชนิดหลายตัวแปรจะมีค่า MAPE น้อยกว่าชนิดหนึ่งตัวแปร

วิธีดำเนินการวิจัย

งานวิจัยนี้พิจารณาว่าตลาดหลักทรัพย์มีสองสถานะหลัก ได้แก่ สถานะตลาดกระทิง (Bull Market) ซึ่งให้ผลตอบแทนสูงและมีความผันผวนต่ำ และสถานะตลาดหมี (Bear Market) ซึ่งให้ผลตอบแทนต่ำแต่มีความผันผวนสูง กล่าวคือแบบจำลองกำหนดจำนวนสถานะซ่อนเร้น (N) = 2

สำหรับการแจกแจงแบบเกาส์ในแต่ละสถานะงานวิจัยพิจารณาว่าแต่ละสถานะประกอบด้วย การแจกแจงแบบเกาส์สองรูปแบบ ซึ่งสะท้อนถึงลักษณะของตลาดกระทิงและตลาดหมีที่ผสมกันจนเกิดเป็นการแจกแจงผสมแบบเกาส์ (Gaussian Mixture Distribution) ดังนั้นแบบจำลองที่ใช้ในงานวิจัยนี้กำหนดจำนวนองค์ประกอบ (M) = 2 เพื่อรองรับการแจกแจงผสมในแต่ละสถานะ

งานวิจัยนี้เปรียบเทียบการแจกแจงของอัตราผลตอบแทน ณ วันสุดท้ายระหว่างแบบจำลอง GBM แบบหลายตัวแปรและ gmHMM โดยมุ่งเน้นที่อัตราผลตอบแทนจากราคาเปิดถึงราคาปิด (r_{oc}) ซึ่งสะท้อนผลตอบแทนจริงจากการถือครองสินทรัพย์ภายในหนึ่งวันทำการและสามารถเปรียบเทียบกับงานวิจัยก่อนหน้าได้ [4]

กระบวนการวิจัยใช้แนวทางเดียวกับ อัครา กิจการเจริญสินและปราณี มณีรัตน์ [4] แต่มีความแตกต่างที่สำคัญดังนี้ 1) ใช้ข้อมูลดัชนี Thailand SET50 รายวันตั้งแต่ 1 ธ.ค. 2564 – 31 ม.ค. 2567 รวม 526 วันทำการ (เพิ่มจาก 510 วัน) ดาวน์โหลดจากโปรแกรม MT4 [9] 2) เพิ่มจำนวนเส้นทางจำลองจาก 10,000 เป็น 100,000 เส้นทาง 3) เปรียบเทียบค่าคาดหวังและความแปรปรวนของดัชนี SET50 จำลองกับค่าที่เกิดขึ้นจริงและ 4) ใช้ GBM แบบสามมิติเป็นเกณฑ์ชี้วัดแทนที่ GBM แบบหนึ่งมิติ การทดลองนี้ใช้วิธีแบ่งข้อมูลฝึกและทดสอบแบบ 50% ซึ่งเป็นแนวทางที่นิยมใช้ในทางปฏิบัติในการสร้าง GBM แบบสามมิติโดยแบ่งข้อมูลรายวันออกเป็นสองส่วนได้แก่ 263 ตัวอย่างสำหรับการฝึกและ 263 ตัวอย่างสำหรับการทดสอบ

สำหรับ gmHMM การสร้างแบบจำลองใช้หน้าต่างเลื่อนขนาด 25 เฟรม โดยในแต่ละช่วงเวลาแบบจำลองจะถูกสร้างจากข้อมูล 25 วันล่าสุด จากนั้นหน้าต่างข้อมูลจะเลื่อนไปข้างหน้าทีละวันและสร้างแบบจำลอง gmHMM ใหม่ กระบวนการนี้ทำซ้ำไปเรื่อยๆ จนกว่าหน้าต่างจะครอบคลุมข้อมูลวันสุดท้าย

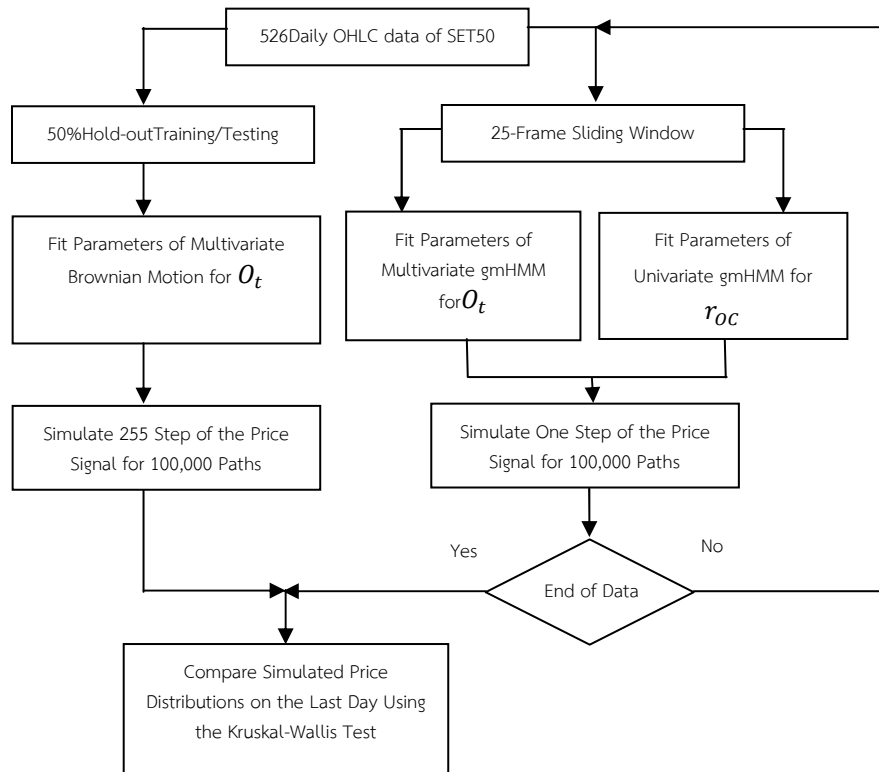


Figure 1 Experimental diagram comparing the distribution of simulated closing prices among multivariate gmHMM, univariate gmHMM, and multivariate GBM Models.

การทดสอบ Kruskal-Wallis และ Wilcoxon Rank Sum ใช้เพื่อตรวจสอบความแตกต่างทางสถิติของการแจกแจงราคาปิด ณ วันทำการสุดท้ายจากการจำลอง 100,000 เส้นทาง โดยเปรียบเทียบระหว่าง GBM สามมิติและ gmHMM แบบหลายตัวแปร การทดสอบนี้เป็นการทดสอบแบบไม่อิงพารามิเตอร์ (Non-Parametric Test) ซึ่งเหมาะสมกับการแจกแจงราคาปิดของ gmHMM ที่ไม่ได้เป็นการแจกแจงแบบปกติ[4] สมมติฐานหลักของการทดสอบคือการแจกแจงของราคาปิดจากทั้งสองแบบจำลองไม่แตกต่างกัน หากการแจกแจงของราคาปิดเป็นการแจกแจงแบบปกติแล้วการทดสอบ Kruskal-Wallis จะให้ผลลัพธ์เทียบเท่ากับการทดสอบ ANOVA กระบวนการทดลองทั้งหมดนี้สรุปได้ดังภาพที่ 1

การจำลองเชิงตัวเลขดำเนินการผ่าน MATLAB เวอร์ชัน R2022a บนคอมพิวเตอร์ส่วนบุคคลโดย Source Code ที่พัฒนาขึ้นใช้การเขียนโปรแกรมเชิงวัตถุ (Object-Oriented Programming - OOP) เครื่องมือที่ใช้ประกอบด้วย Financial Toolbox และ Statistical Toolbox สำหรับการทดสอบการแจกแจง Kruskal-Wallis Test และ Wilcoxon Rank Sum Test กระทำผ่านฟังก์ชัน kruskalwallis และ ranksum ใน MATLAB ตามลำดับ

ผลการวิจัยและอภิปราย

การทดลองพบว่าการแจกแจงของดัชนี SET50 ตามแบบบราวน์เรขาคณิต (GBM) โดย O_t ในสมการ (2) เป็นอัตราผลตอบแทนในสามมิติไม่แตกต่างไปจากการจำลอง GBM ในหนึ่งมิติ [4] ภาพที่ 2 แสดงการจำลองดัชนี SET50 ตามแบบจำลองบราวน์เรขาคณิตในสามมิติโดยอัตราผลตอบแทนคำนวณจาก r_{OC} กราฟด้านซ้ายแสดงการแจกแจงของดัชนี SET50 จำลองเทียบกับค่าที่เกิดขึ้นจริง ส่วนด้านขวาคือฮิสโตแกรมของดัชนีจำลอง SET50 ณ วันสุดท้ายเทียบกับการแจกแจงแบบ Log-Normal ผลลัพธ์ที่ได้ยังแสดงว่าแบบจำลองบราวน์เรขาคณิตในหลายมิติโดยอัตราผลตอบแทนตามสมการ (2) พบปัญหาเรื่องอคติด้านบวก (Positive bias) เช่นเดียวกับการหนึ่งมิติ [4]

อคติเชิงบวกของแบบจำลอง GBM มีสาเหตุมาจากโครงสร้างของสมการ (1) ซึ่งมีผลเฉลยในรูปของฟังก์ชันเอกซ์โพเนนเชียลหากตลาดอยู่ในภาวะขาขึ้น เช่น ระหว่างช่วงต้นปี พ.ศ. 2552 ถึงปลายปี พ.ศ. 2560 แบบจำลอง GBM สามารถจำลองพฤติกรรมของตลาดได้อย่างเหมาะสมอย่างไรก็ตามในช่วงวันที่ 1 ธันวาคม พ.ศ. 2564 ถึง 31 มกราคม พ.ศ. 2567 ซึ่งตลาดอยู่ในภาวะขาลงและไซด์เวย์พบว่า GBM อาจไม่สามารถจำลองพฤติกรรมตลาดได้อย่างแม่นยำ ส่งผลให้มีความจำเป็นต้องศึกษาวิธีการจำลองทางเลือกเพิ่มเติมที่สามารถรองรับสถานะตลาดที่แตกต่างกันได้อย่างมีประสิทธิภาพมากขึ้น

แบบจำลอง gmHMM ชนิดหลายตัวแปรโดย O_t ตามสมการ (2) เพิ่มความถูกต้องของดัชนีจำลอง SET50 มากกว่าชนิดหนึ่งตัวแปร ภาพที่ 3 และ 4 เป็นดัชนีจำลอง SET50 จากแบบจำลอง gmHMM ชนิดหนึ่งตัวแปรและชนิดหลายตัวแปรตามลำดับ กราฟทางซ้ายมีนอกจากแสดงเส้นทางเดินจำลองและดัชนี SET50 จริงแล้วเส้นกราฟสีดำแสดงค่าคาดหวังของดัชนี SET50 ที่อ้างอิงการแจกแจงตามแบบจำลอง gmHMM ส่วนเส้นโค้งระฆังในกราฟทางขวาคือเส้น Density ของการแจกแจง log-Normal ซึ่งได้จากแบบจำลองบราวน์เรขาคณิต ผลการทดลองแสดงว่าความถูกต้องของแบบจำลอง gmHMM ชนิดหลายตัวแปรมากกว่าชนิดตัวแปรเดียว โดยค่าคาดหวังตามเส้นกราฟสีดำของชนิดตัวหลายตัวแปรใกล้เคียงกับเส้นกราฟสีแดงที่เป็นค่าจริงมากกว่าชนิดตัวแปรเดียว นอกจากนี้ค่า MAPE ของชนิดหลายตัวแปรมีค่าเพียง 1.7409 เทียบกับ 2.7261 ของชนิดตัวแปรเดียว

การจำลองด้วย gmHMM สามารถสะท้อนการเปลี่ยนแปลงของราคาที่เกิดขึ้นอย่างฉับพลันได้อย่างแม่นยำตามที่แสดงในภาพที่ 4 ดัชนีราคาจริง (เส้นสีแดง) มีแนวโน้มขาลงแต่มีการปรับตัวขึ้นเป็นระยะในช่วงวันที่ 1 ถึง 121 ทั้งดัชนีราคาจริงและค่าคาดหวังจากการจำลอง (เส้นสีดำ) แสดงแนวโน้มลดลงอย่างต่อเนื่องจากนั้นเกิดการปรับตัวขึ้นจนถึงวันที่ 161 ก่อนกลับเข้าสู่แนวโน้มขาลงอีกครั้งจนถึงวันสุดท้ายของการจำลอง

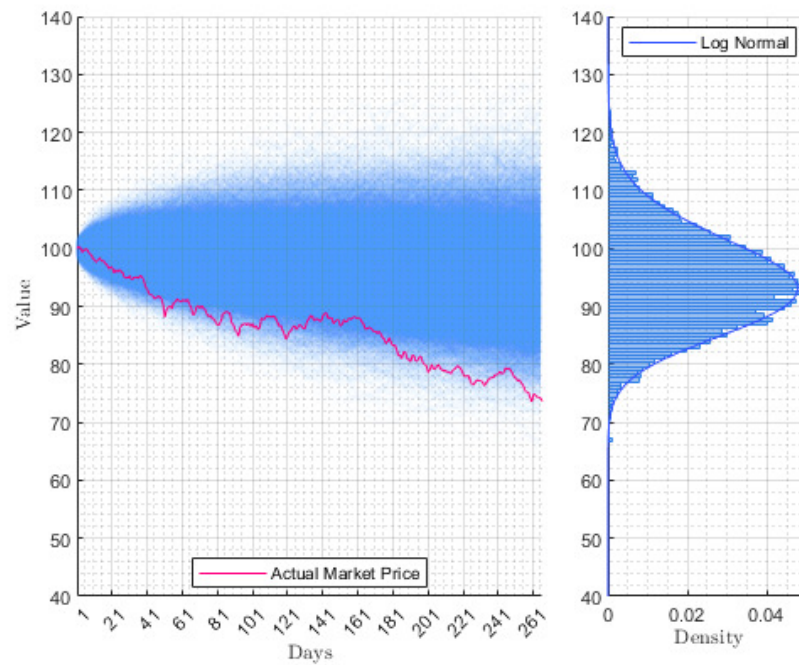


Figure 2 Histogram and distribution of simulated closing prices on the last date of the Thailand SET50 Index using a multidimensional geometric Brownian motion model ($\mu = -0.0578$, $\sigma = 0.0879$) with theoretical P.D.F. overlay.

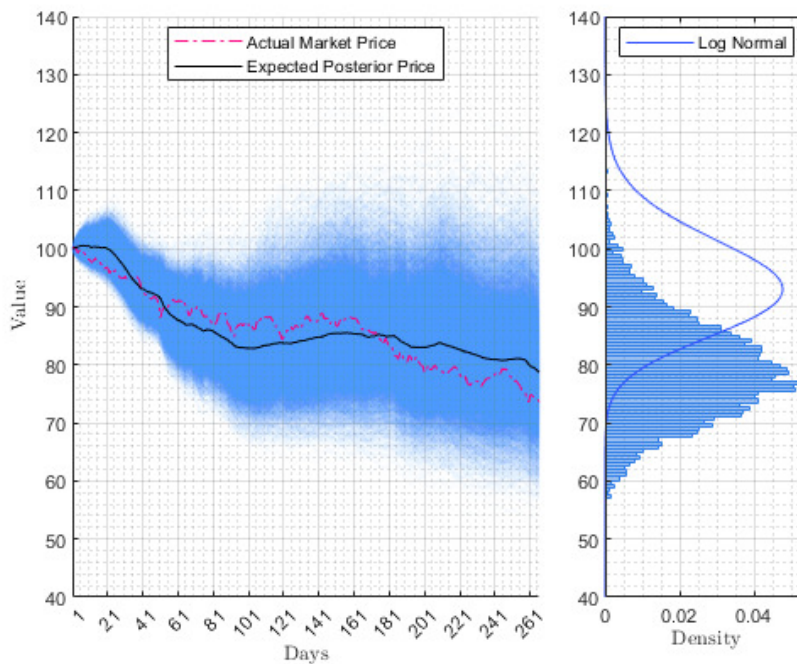


Figure 3 Histogram and distribution of simulated closing prices on the last date of the Thailand SET50 index using a two-state, two-mixture univariate gmHMM with log-Normal P.D.F. overlay ($\mu=-0.0578$, $\sigma=0.0879$) and MAPE of expected vs. actualvalue = 2.7261

ผลลัพธ์ดังกล่าวแสดงให้เห็นว่า gmHMM สามารถจับแนวโน้มตลาดและพฤติกรรมของดัชนีที่มีความผันผวนได้อย่างมีประสิทธิภาพ ซึ่งแตกต่างจากแบบจำลอง GBM ที่สามารถสะท้อนแนวโน้มของตลาดได้อย่างแม่นยำเฉพาะในภาวะขาขึ้นเท่านั้น อย่างไรก็ตามการสร้างแบบจำลอง gmHMM อาศัยกระบวนการที่ต้องใช้ทรัพยากรสูงเนื่องจากใช้ Sliding Window ขนาด 25 เฟรม ซึ่งต้องมีการสร้างแบบจำลองใหม่ทุกครั้งที่หน้าต่างเลื่อน ส่งผลให้ต้นทุนการคำนวณ เพิ่มขึ้นอย่างมาก ในทางตรงกันข้าม GBM สามารถสร้างแบบจำลองเพียงครั้งเดียวและใช้งานได้ทันที ดังนั้นหากตลาดอยู่ในภาวะขาขึ้นเพียงอย่างเดียวการจำลองด้วย gmHMM อาจมีค่าใช้จ่ายที่สูงและเกินความจำเป็น ความแตกต่างในการแจกแจงของราคา ณ วันสุดท้ายที่จำลองขึ้นตามภาพที่ 2 ถึง 4 ถูกนำมาเปรียบเทียบและแสดงผลผ่านกราฟ Box plot ในภาพที่ 5 และยืนยันได้ว่าการแจกแจงของราคาจำลองไม่ใช้การแจกแจงแบบปกติ ดังนั้นการทดสอบแบบไมอิงพารามิเตอร์เพื่อเปรียบเทียบการแจกแจงในราคาจำลองวันสุดท้ายจึงจำเป็น

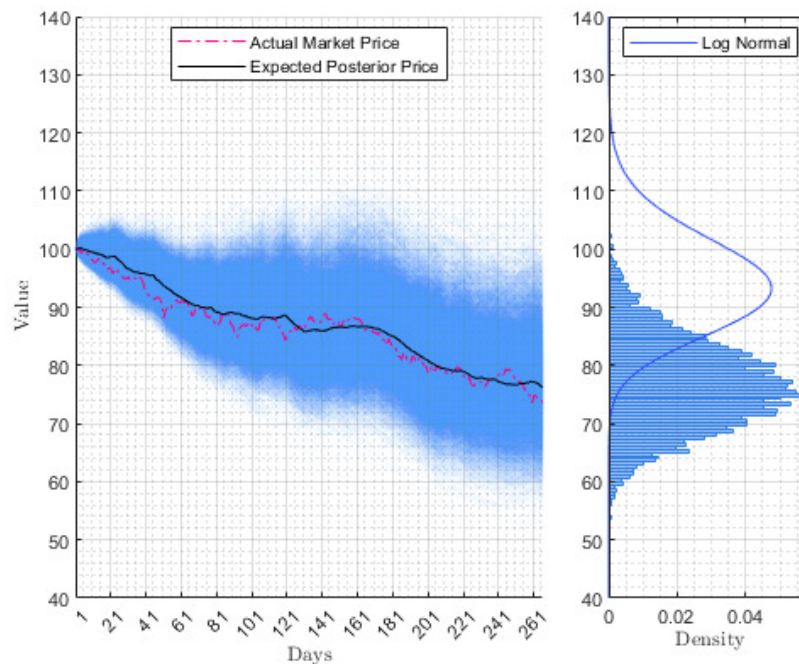


Figure 4 Histogram and distribution of simulated closing prices on the last date of the Thailand SET50 Index using a two-state, two-mixturemultivariategmHMM with log-Normal P.D.F. overlay ($\mu=-0.0578$, $\sigma=0.0879$) and MAPE of expected vs. actual value = 1.7409.

จาก Box Plot ในภาพที่ 5 พบว่าการแจกแจงของราคาจำลองที่ได้จาก GBM ทั้งแบบหนึ่งมิติและสามมิติไม่แตกต่างกัน ในขณะที่การแจกแจงของราคาจำลองที่ได้จาก gmHMM ชนิดตัวแปรเดียวและหลายตัวแปรมีความแตกต่างกัน นอกจากนี้ยังพบว่าการแจกแจงของราคาจำลองที่ได้จาก GBM และ gmHMM มีความแตกต่างกันอย่างชัดเจน ผลการทดสอบด้วย Kruskal-Wallis Test ยืนยันว่าแบบจำลอง gmHMM ชนิดหลายตัวแปรสร้างการแจกแจงของราคาดัชนี SET50 ณ วันสุดท้ายที่แตกต่างจาก gmHMM ชนิดตัวแปรเดียวและไม่พบอคติด้านบวกที่เกิดขึ้นในแบบจำลอง GBM ทั้งแบบหนึ่งมิติและสามมิติโดยมีนัยสำคัญทางสถิติที่ $p\text{-value} < 0.001$ นอกจากนี้ Wilcoxon Rank Sum Test ยังยืนยันความแตกต่างของการแจกแจงใน gmHMM ทั้งสองประเภทด้วย $p\text{-value} < 0.001$ ในขณะที่การทดสอบเดียวกันไม่พบความแตกต่างระหว่าง GBM แบบหนึ่งมิติและสามมิติโดยให้ $p\text{-value} = 0.5489$

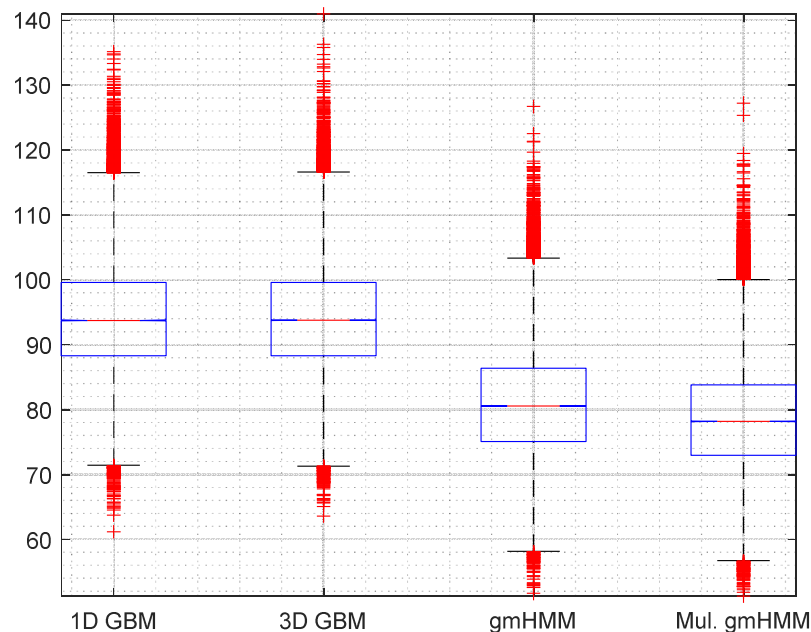


Figure 5 Box plot comparison of non-Normal simulated closing price distribution on the last date of Thailand SET50 index comparing 1D GBM with 3D GBM and univariate gmHMM with multivariate gmHMM.

สรุปผลการวิจัย

งานวิจัยนี้สร้างแบบจำลองของดัชนี Thailand SET50 จำนวนสี่แบบ ได้แก่ แบบจำลองบราวน์เรซาคณิต (GBM) ทั้งหนึ่งและหลายมิติและแบบจำลองมาร์คอฟซ่อนเร้นผสมแบบเกาส์ (gmHMM) ตัวแปรเดียวและหลายตัวแปรผ่านข้อมูล SET50 รายวัน ตั้งแต่วันที่ 1 ธ.ค. 2564 ถึง 31 ม.ค. 2567 ผลการเปรียบเทียบการแจกแจงเชิงสถิติของค่าดัชนีจำลอง ณ วันสุดท้ายเมื่อกำหนดให้ผลลัพธ์ที่สังเกตได้ของ Thailand SET50 ประกอบด้วยราคาเปิด ราคาปิด ราคาสูงสุด และราคาต่ำสุดสามารถสรุปได้ดังนี้

- 1) การเพิ่มมิติของแบบจำลอง GBM จากหนึ่งเป็นสามมิติไม่มีความแตกต่างและทั้งคู่มีอคติเชิงบวก
- 2) แบบจำลอง gmHMM ชนิดหลายตัวแปรจำลองการแจกแจงของดัชนี SET50 ต่างจากชนิดตัวแปรเดียวทั้งคู่ไม่พบอคติเชิงบวก
- 3) แบบจำลอง gmHMM ชนิดหลายตัวแปรจำลองการแจกแจงของดัชนี Thailand SET50 ได้แม่นยำกว่าแบบตัวแปรเดียว โดยค่าคาดหวังเมื่อจำลองดัชนีจำนวน 100,000 เส้นทางแล้วแบบจำลองชนิดหลายตัวแปร มีค่า MAPE ลดลงจาก 2.7261 เป็น 1.7409 เทียบกับชนิดตัวแปรเดียว

งานวิจัยนี้อยู่ในสมมติฐานว่าดัชนี Thailand SET50 เป็นกระบวนการมาร์คอฟ เมื่อสมมติฐานนี้เป็นจริงแบบจำลอง gmHMM ชนิดหลายตัวแปรสามารถคาดการณ์การแจกแจงของดัชนี SET50 ที่ช่วงเวลาถัดไป

นักลงทุนสามารถใช้แบบจำลอง gmHMM บริหารการลงทุนให้ได้ผลตอบแทนตามคาดหวังบนความเสี่ยงที่สามารถรับได้เพื่อพร้อมต่อสังคมผู้สูงวัยที่กำลังมาถึง

กิตติกรรมประกาศ

ขอแสดงความขอบคุณ ผศ.ดร.ศุภเชษฐ์ อินทร์เนตร คณะวิศวกรรมศาสตร์ มหาวิทยาลัยหอการค้าไทย เป็นอย่างยิ่งที่กรุณาตรวจสอบต้นฉบับ โดยให้คำแนะนำเกี่ยวกับการตรวจตัวสะกดและการแก้ไขข้อผิดพลาด ในการพิมพ์ก่อนการส่งต้นฉบับฉบับแรก ซึ่งมีส่วนสำคัญในการปรับปรุงความถูกต้องและความสมบูรณ์ของงานวิจัยนี้

References

- [1] Innet, S., & Kijkarncharoensin, A. (2021). The Machine Learning Model of the Thailand SET 50 Index. In *The 44th Electrical Engineering Conference (EECON44)* (p. 562-265). 17-19 November, 2021, Nan, Thailand. (in Thai)
- [2] Agustini, W. F., et al. (2018). Stock Price Prediction Using Geometric Brownian Motion. *Journal of Physics: Conference Series*, 974, 012047.
- [3] Juang, B. H. (1985). Maximum-Likelihood Estimation for Mixture Multivariate Stochastic Observations of Markov Chains. *AT&T Technical Journal*, 64(6), 1235–1249.
- [4] Kijkarcharoensin, A., & Maneeratn, P. (2024). Simulate the Signal of Stock Price by the Gaussian Mixture Hidden Markov Model. In *The 16th Conference of Electrical Engineering Network 2024 (EENET2024)*. 29-31 May, 2024, Nongkhai, Thailand. (in Thai)
- [5] Gupta, A., & Dhingra, B. (2012). Stock Market Prediction Using Hidden Markov Models. *2012 Students Conference on Engineering and Systems (SCES 2012)* (p. 1-4). 16-18 March, 2012, Uttar Pradesh, India.
- [6] Nguyen, N. (2018). Hidden Markov Model for Stock Trading. *International Journal of Financial Studies*, 6(2), 36.
- [7] Levinson, S. E., et al. (1983). An Introduction to the Application of the Theory of Probabilistic Functions of a Markov. *The Bell System Technical Journal*, 62(4), 1035–1074.
- [8] Liporace, L. (1982). Maximum Likelihood Estimation for Multivariate Observations of Markov Sources. *IEEE Transactions on Information Theory*, 28(5), 729-734.
- [9] Top Trader (2025). *Metatrader 4 for TFEX*. Available from <https://www.toptrader.co.th/download/>. Accessed date: 21 February 2025.