

การเพิ่มความแม่นยำในการพยากรณ์ข้อมูลอนุกรมเวลาด้วยตัวแบบ

ผสม ARIMA-ANN และการวิเคราะห์การถดถอย

Improving Prediction Accuracy of Time Series Data Using

ARIMA-ANN Hybrid Model and Regression Analysis

นัท กุลวานิช*

ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

แขวงวังใหม่ เขตปทุมวัน กรุงเทพมหานคร 10330

Nat Kulvanich*

Department of Statistics, Faculty of Commerce and Accountancy, Chulalongkorn University,

Wangmai, Pathumwan, Bangkok 10330

บทคัดย่อ

การใช้วิธีการวิเคราะห์การถดถอยในการสร้างตัวแบบผสมแบบใหม่ระหว่างตัวแบบ ARIMA กับตัวแบบเครือข่ายประสาทเทียม ที่มีชื่อเรียกว่าตัวแบบผสม ARIMA-ANN-REG ให้ความแม่นยำในการพยากรณ์ดีกว่าตัวแบบผสมระหว่างตัวแบบ ARIMA กับตัวแบบเครือข่ายประสาทเทียมแบบดั้งเดิม ที่มีชื่อเรียกว่าตัวแบบผสม ARIMA-ANN งานวิจัยนี้ศึกษาและเปรียบเทียบความแม่นยำของค่าพยากรณ์ที่ได้จากตัวแบบ ARIMA และตัวแบบผสมระหว่าง ARIMA กับเครือข่ายประสาทเทียม (ARIMA-ANN) รวมทั้งตัวแบบผสมระหว่าง ARIMA กับเครือข่ายประสาทเทียม โดยอาศัยวิธีการวิเคราะห์การถดถอย (ARIMA-ANN-REG) โดยศึกษาเปรียบเทียบทั้งในส่วนของข้อมูลจำลองและข้อมูลจริง ข้อมูลจำลองที่ใช้ในการศึกษานั้นสร้างมาจากตัวแบบ ARIMA จำนวนทั้งสิ้น 8 ตัวแบบ และข้อมูลจริงนั้นได้มีการนำข้อมูลจริงทั้งหมด 6 ชุดข้อมูล ในประเทศไทยมาใช้ในการเปรียบเทียบ ผลการศึกษาพบว่าตัวแบบผสม ARIMA-ANN-REG มีความแม่นยำในการพยากรณ์สูงที่สุดทั้งชุดข้อมูลจริงและชุดข้อมูลจำลอง

คำสำคัญ : ARIMA-ANN; ตัวแบบผสม; การวิเคราะห์การถดถอย; ข้อมูลอนุกรมเวลา

Abstract

The new hybrid model using linear regression methods to combine both ARIMA and Artificial Neural Networks (ARIMA-ANN-REG Hybrid Model) achieves better prediction accuracy than the traditional hybrid model combining ARIMA and ANN (ARIMA-ANN Hybrid Model). This research

*ผู้รับผิดชอบบทความ : nat@cbs.chula.ac.th

examines and compares the prediction accuracy among ARIMA model, ARIMA-ANN model, and ARIMA-ANN-REG model by using both real and simulated data for the comparison of prediction accuracy. The simulated data are generated from 8 ARIMA processes and the real data are comprised of six real datasets in Thailand. The results suggest that ARIMA-ANN-REG model has the highest prediction accuracy in both real and simulated data.

Keywords: ARIMA-ANN; hybrid model; regression analysis; time series data

1. บทนำ

ปัจจุบันมีบริษัทหรือองค์กรต่าง ๆ เกิดขึ้นมากมาย ทำให้เกิดการแข่งขันที่สูงขึ้นเมื่อเปรียบเทียบกับในอดีตที่ผ่านมา และเป็นที่ทราบกันดีว่าธุรกิจหรือองค์กรที่ประสบความสำเร็จจำเป็นต้องมีการวางแผนอนาคตที่ดี โดยทั่วไปมักมีการใช้ความรู้ทางสถิติมาประยุกต์เพื่อใช้พยากรณ์เหตุการณ์ในอนาคต ซึ่งการวิเคราะห์อนุกรมเวลา (time series analysis) นั้นเป็นหนึ่งในวิธีการสถิติที่ใช้ข้อมูลในอดีตที่ถูกเก็บรวบรวมตามลำดับเวลามาสร้างสมการหรือตัวแบบเชิงคณิตศาสตร์แทนลักษณะการเคลื่อนไหวของข้อมูล เพื่อใช้สำหรับพยากรณ์ข้อมูลในอนาคต [1]

วิธีการวิเคราะห์อนุกรมเวลามีหลายวิธีสามารถแบ่งเป็น 2 กลุ่ม กลุ่มแรก คือ วิธีการแบบเชิงเส้น (linear approach) ซึ่งมีข้อสมมติว่าลักษณะความสัมพันธ์ของค่าสังเกตปัจจุบันเป็นฟังก์ชันเชิงเส้นของค่าสังเกตในอดีต และวิธีการสถิติที่เป็นที่นิยมในกลุ่มนี้ คือ การพยากรณ์ด้วยวิธีของ Box-Jenkins ซึ่งเกิดในปี ค.ศ. 1970 โดยใช้ค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเองและค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเองบางส่วนมาสร้างตัวแบบสำหรับการพยากรณ์ที่เรียกว่า autoregressive integrated moving average (ARIMA) ซึ่งเป็นตัวแบบที่เหมาะสมกับการใช้พยากรณ์อนุกรมเวลาในส่วนที่เป็นฟังก์ชันเชิงเส้นตรงได้ดี แต่อาจไม่สามารถอธิบายลักษณะแบบไม่ใช่เชิงเส้นตรงที่มี

อยู่ในค่าสังเกตของข้อมูลอนุกรมเวลาได้ดีนัก [2]

ในเวลาต่อมา มีการพัฒนาวิธีการวิเคราะห์อนุกรมเวลากลุ่มที่สอง เพื่อใช้ในการศึกษาความสัมพันธ์ที่ไม่ใช่เชิงเส้นของอนุกรมเวลา (non-linear approach) และวิธีทางสถิติที่เป็นที่นิยมในกลุ่มนี้ คือ ตัวแบบเครือข่ายประสาทเทียม (artificial neuron network, ANN) ซึ่งเป็นตัวแบบที่มีความแม่นยำสูงในการพยากรณ์ข้อมูลอนุกรมเวลาในส่วนที่ไม่เป็นฟังก์ชันเชิงเส้นตรงและมีความยืดหยุ่นในการใช้งาน เนื่องจากไม่จำเป็นต้องมีข้อสมมติเบื้องต้น (assumption) ในการสร้างตัวแบบ

ทั้งนี้ข้อมูลอนุกรมเวลาส่วนใหญ่ก็มีโครงสร้างที่มีส่วนประกอบทั้งแบบเชิงเส้นตรง และแบบที่ไม่ใช่เชิงเส้นตรง ดังนั้นการวิเคราะห์อนุกรมเวลาด้วยตัวแบบ ARIMA หรือตัวแบบ ANN เพียงตัวแบบเดียวอาจมีข้อจำกัด เนื่องจากไม่สามารถอธิบายลักษณะการเคลื่อนไหวและลักษณะการผันแปรของข้อมูลอนุกรมเวลาเหล่านั้นอย่างครบถ้วน Zhang [3] จึงได้เสนอตัวแบบผสม (hybrid model) ระหว่างตัวแบบ ARIMA และตัวแบบ ANN เพื่อใช้พยากรณ์ข้อมูลที่มีส่วนประกอบทั้งสองให้มีความแม่นยำมากขึ้น และได้มีผู้สนใจศึกษาการพยากรณ์อนุกรมเวลาด้วยตัวแบบผสมดังกล่าวไว้หลายงานวิจัยด้วยกัน เช่น Faruk [4] Meth และคณะ [5] Koutroumanidis และคณะ [6] Ebrahimi [7] พงษ์ศิริ และคณะ [8] ชญานิน และ นัท

[9] และ ชยกร สมศิลา และคณะ [10] ทั้งนี้งานวิจัยดังกล่าวรายงานผลสรุปที่เหมือนกันว่าการใช้ตัวแบบผสมระหว่างตัวแบบ ARIMA และตัวแบบ ANN ให้ความแม่นยำที่สูงกว่าการใช้ตัวแบบ ARIMA เพียงตัวแบบเดียว

ต่อมา Khairalla และคณะ [11] ได้พัฒนาวิธีการผสมตัวแบบ ARIMA และ ANN วิธีใหม่ โดยอาศัยวิธีการวิเคราะห์การถดถอย (regression analysis) เพื่อรวมค่าพยากรณ์ของทั้งสองตัวแบบเข้าด้วยกัน และใช้ข้อมูลอัตราแลกเปลี่ยนเงินสกุลยูโร (SDG-EURO) เพื่อทดสอบความแม่นยำของตัวแบบดังกล่าว ซึ่งได้ผลสรุปว่าตัวแบบผสมระหว่าง ARIMA กับ ANN โดยใช้วิธีวิเคราะห์การถดถอยให้ความแม่นยำในการพยากรณ์ที่สูงกว่าตัวแบบผสมของ Zhang [3]

การศึกษาครั้งนี้ผู้วิจัยสนใจที่จะขยายขอบเขตงานวิจัยของ Khairalla และคณะ [11] เพื่อเปรียบเทียบความแม่นยำของตัวแบบผสมระหว่าง ARIMA กับ ANN โดยใช้วิธีวิเคราะห์การถดถอยและตัวแบบผสมระหว่าง ARIMA กับ ANN ของ Zhang [3] โดยใช้ชุดข้อมูลอนุกรมเวลาที่จำลองด้วยตัวแบบ ARIMA ทั้งหมด 8 ตัวแบบ และใช้ชุดข้อมูลจริงทั้งหมด 6 ชุดข้อมูล โดยใช้เกณฑ์รากของค่าคลาดเคลื่อนกำลังสอง (RMSE) เป็นเครื่องมือในการเปรียบเทียบตัวแบบตัวแบบใดที่มีค่า RMSE ต่ำสุด จะเป็นตัวแบบที่ให้ผลดีที่สุด ผลการศึกษาพบว่าตัวแบบผสม ARIMA-ANN-REG มีความแม่นยำในการพยากรณ์สูงที่สุดในชุดข้อมูลจริงและชุดข้อมูลจำลอง งานวิจัยนี้จึงเกิดประโยชน์ประโยชน์ทั้งในทางทฤษฎีและทางปฏิบัติ เนื่องจากผู้วิเคราะห์ข้อมูลสามารถนำผลวิจัยที่ได้นี้ไปใช้ในการวางแผนหรือใช้เป็นแนวทางในการเลือกวิธีการพยากรณ์สำหรับข้อมูลอนุกรมเวลาในอนาคต

2. วิธีการวิจัย

2.1 ศึกษาตัวแบบและทฤษฎีที่เกี่ยวข้อง

2.1.1 ตัวแบบ integrated autoregressive moving average (ARIMA)

ตัวแบบ ARIMA พัฒนาขึ้นในปี ค.ศ. 1970 เป็นวิธีการวิเคราะห์อนุกรมเวลาที่ได้รับความนิยมมากวิธีหนึ่ง โดยมีข้อสมมติว่าค่าของข้อมูลปัจจุบันเป็นฟังก์ชันเชิงเส้นของค่าสังเกตและค่าความคลาดเคลื่อนสุ่มในอดีต โดยมีรูปแบบสมการดังนี้

$$\phi(B)\nabla^d Y_t = \theta(B)w_t \quad (1)$$

เมื่อ Y_t แทนค่าสังเกตของข้อมูลอนุกรมเวลา ณ เวลา t ; B แทนตัวดำเนินการถอยหลัง (backward operator) โดยที่ $B^s Y_t = Y_{t-s}$; $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ แทนตัวดำเนินการสหสัมพันธ์ลำดับที่ p ; $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$ แทนตัวดำเนินการเคลื่อนที่เฉลี่ยลำดับที่ q ; $\phi_1, \phi_2, \dots, \phi_p$ แทนสัมประสิทธิ์ตัวดำเนินการสหสัมพันธ์ โดยที่ $\phi_p \neq 0$; $\theta_1, \theta_2, \dots, \theta_q$ แทนสัมประสิทธิ์ตัวดำเนินการเคลื่อนที่เฉลี่ย โดยที่ $\theta_q \neq 0$; $\nabla^d = (1 - B)^d$ แทนการหาผลต่าง (differencing) อันดับที่ d ; w_t คือ gaussian white noise ที่มีค่าเฉลี่ยเป็น 0 และความแปรปรวน σ_w^2

2.1.2 ตัวแบบเครือข่ายประสาทเทียม (artificial neural network model, ANN)

ตัวแบบ ANN เป็นตัวแบบทางคณิตศาสตร์ที่จำลองขึ้นโดยอาศัยหลักการทำงานของสมองมนุษย์ โดยการประมวลผลของเครือข่ายประสาทเทียมจะเกิดขึ้นในหน่วยประมวลผลย่อย (node) โดยเมื่อมีข้อมูลนำเข้า (input data) เข้ามายังโครงข่าย ข้อมูลนำเข้าแต่ละค่าจะถูกนำมาคูณกับค่าถ่วงน้ำหนัก (weight) แล้วนำมารวมกัน เพื่อส่งผ่านไปยังฟังก์ชันการแปลง (transfer function) หรือฟังก์ชันกระตุ้น (activation function) และส่งข้อมูลผลลัพธ์ไปยังหน่วยประมวลผลย่อยถัดไป

ทั้งนี้ ตัวแบบ ANN สามารถนำมาประยุกต์ใช้ในการวิเคราะห์อนุกรมเวลา โดยเครือข่ายประสาทเทียมเพอร์เซ็ปตรอนหลายชั้น (multi-layer perceptron, MLP) จัดเป็นตัวแบบ ANN ที่นิยมใช้ในการพยากรณ์ข้อมูลอนุกรมเวลา โดยมีโครงสร้าง 3 ชั้น คือ input layer, hidden layer และ output layer ซึ่งแต่ละชั้นจะประกอบด้วย node ที่มีจำนวนต่างกันออกไป และ node ใน input layer จะเชื่อมโยงเข้ากับทุก node ในชั้นถัดไปตามลำดับจนถึง output layer ซึ่งสามารถเขียนในรูปสมการได้ดังนี้ [12]

$$Y_t = \alpha + \sum_{j=1}^n \alpha_j f(\sum_{i=1}^m \beta_{ij} Y_{t-1} + \beta_j) + \varepsilon_t \quad (2)$$

เมื่อ Y_t แทนค่าสังเกตของข้อมูลอนุกรมเวลา ณ เวลา t ; α แทนเวกเตอร์ของค่าถ่วงน้ำหนักระหว่างหน่วยประมวลผลย่อยจำนวน n หน่วยใน hidden layer กับหน่วยประมวลผลย่อยใน output layer; β_{ij} แทนค่าถ่วงน้ำหนักระหว่างหน่วยประมวลผลย่อยแต่ละหน่วยใน input layer ซึ่งมีทั้งหมด m หน่วย กับหน่วยประมวลผลย่อยใน hidden layer โดยที่ $i = 1, 2, \dots, m$ และ $j = 1, 2, \dots, n$; f แทนฟังก์ชันกระตุ้นโดยใช้ sigmoid logistic function หรือ $f(x) = 1/(1 + e^{-x})$; ε_t แทนค่าคลาดเคลื่อน ณ เวลา t

2.1.3 ตัวแบบผสม ARIMA และ ANN (hybrid model, ARIMA-ANN)

Zhang [3] ได้เสนอตัวแบบอนุกรมเวลาแบบผสม ซึ่งมีข้อสมมติว่าข้อมูลอนุกรมเวลาที่นำมาวิเคราะห์นั้นมี 2 องค์ประกอบ คือ องค์ประกอบที่เป็นเส้นตรง (linear component) และองค์ประกอบที่ไม่เป็นเส้นตรง (non-linear component) ทั้งนี้ Zhang [3] ได้นำตัวแบบ ARIMA ซึ่งเป็นฟังก์ชันเชิงเส้นตรงในเทอมของค่าพารามิเตอร์มารวมกับตัวแบบ ANN ซึ่งใช้สำหรับพยากรณ์ข้อมูลส่วนที่ไม่เป็นฟังก์ชันเชิงเส้นตรงในเทอมของค่าพารามิเตอร์ โดยตัวแบบ

ผสมดังกล่าวมีขั้นตอนการสร้าง ดังแสดงในรูปที่ 1 และสามารถแสดงได้ด้วยสมการเชิงคณิตศาสตร์ ดังนี้

$$Y_t = F_t^{ARIMA} + F_t^{ANN} + \varepsilon_t \quad (3)$$

เมื่อ Y_t แทนค่าสังเกตของข้อมูลอนุกรมเวลา ณ เวลา t ; F_t^{ARIMA} แทนข้อมูลส่วนที่เป็นฟังก์ชันเชิงเส้นตรงในเทอมของค่าพารามิเตอร์ ที่พยากรณ์ด้วยตัวแบบ ARIMA; F_t^{ANN} แทนข้อมูลส่วนที่ไม่เป็นฟังก์ชันเชิงเส้นตรงในเทอมของค่าพารามิเตอร์ ที่พยากรณ์ด้วยตัวแบบ ANN; ε_t แทนค่าคลาดเคลื่อน ณ เวลา t

โดยมีขั้นตอนการสร้างตัวแบบผสมอย่างละเอียดดังนี้

ขั้นตอนที่ 1 : วิเคราะห์ข้อมูลอนุกรมเวลาด้วยตัวแบบ ARIMA เพื่อพยากรณ์ข้อมูลส่วนที่เป็นฟังก์ชันเชิงเส้นตรงในเทอมของค่าพารามิเตอร์ คือ \hat{F}_t^{ARIMA}

ขั้นตอนที่ 2 : คำนวณค่าส่วนเหลือ (residual) จากตัวแบบ ARIMA โดยที่ $\varepsilon_t = Y_t - \hat{F}_t^{ARIMA}$

ขั้นตอนที่ 3 : นำค่าส่วนเหลือที่ได้จากตัวแบบ ARIMA ไปพยากรณ์ข้อมูลส่วนที่ไม่เป็นฟังก์ชันเชิงเส้นตรงในเทอมของค่าพารามิเตอร์ คือ \hat{F}_t^{ANN} ด้วยตัวแบบ ANN

ขั้นตอนที่ 4 : คำนวณหาค่าพยากรณ์รวม (forecasting) จากสมการ $\hat{Y}_t = \hat{F}_t^{ARIMA} + \hat{F}_t^{ANN}$

2.1.4 การใช้ตัวแบบผสม ARIMA และ ANN ในงานวิจัยต่าง ๆ

งานวิจัยที่มีข้อมูลประเภทอนุกรมเวลาเกิดขึ้นอยู่เสมอ หลาย ๆ งานวิจัยก็ต้องการเทคนิคการพยากรณ์ที่ให้ความแม่นยำสูง ดังนั้นจึงมีผู้วิจัยนำตัวแบบอนุกรมเวลาแบบผสม ARIMA-ANN ไปใช้เปรียบเทียบกับตัวแบบอนุกรมเวลา ARIMA สามารถแสดงตัวอย่างงานวิจัยดังกล่าวดังในตารางที่ 1

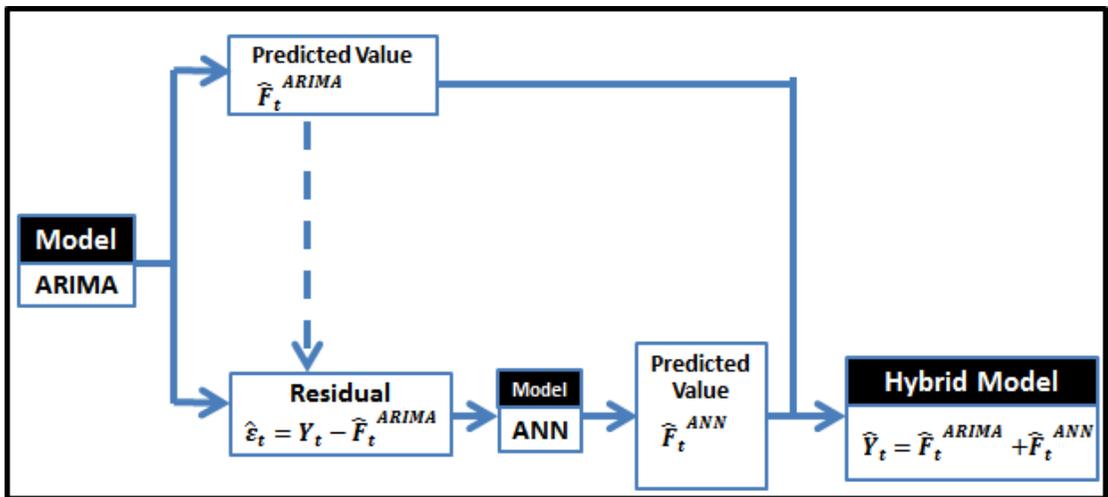


Figure 1 ARIMA-ANN hybrid model

Table 1 Previous studies on ARIMA-ANN hybrid model in time-series analysis

Research	Time-series data
Faruk [4]	water quality
Meth <i>et al.</i> [5]	Indian stock
Koutroumanidis <i>et al.</i> [6]	fuel wood prices in Greece
Ebrahimi [7]	styrene price
Siripanich <i>et al.</i> [8]	closing price of PTT
Somsila <i>et al.</i> [10]	air quality in Thailand

ตารางที่ 1 เห็นได้ชัดว่ามีการนำตัวแบบผสม ARIMA-ANN ไปใช้ในการพยากรณ์ข้อมูลอนุกรมเวลาที่ต่างกัน และทุกงานวิจัยดังกล่าวพบว่าตัวแบบผสม ARIMA-ANN ให้ความแม่นยำมากกว่าตัวแบบ ARIMA

2.1.5 ตัวแบบผสม ARIMA-ANN โดยใช้วิธีวิเคราะห์การถดถอย (hybrid model, ARIMA-ANN-REG)

ทั้งนี้ตัวแบบผสมของ Zhang [3] นั้นจะต้องมีการนำค่าพยากรณ์ที่ได้จากตัวแบบ ARIMA ไปใช้ในการหาค่าพยากรณ์จากตัวแบบ ANN โดย

Khairalla และคณะ [11] ได้เสนอตัวแบบอนุกรมเวลาแบบผสมระหว่างตัวแบบ ARIMA กับตัวแบบ ANN มีการใช้วิธีวิเคราะห์การถดถอย (regression analysis) เพื่อรวมค่าพยากรณ์จากตัวแบบ ARIMA และค่าพยากรณ์จากตัวแบบ ANN เข้าด้วยกัน ซึ่งไม่จำเป็นต้องทำแยกเป็นสองขั้นตอนต่อเนื่องกัน ดังแสดงในรูปที่ 2 และสามารถแสดงได้ด้วยสมการเชิงคณิตศาสตร์ ดังนี้

$$Y_t = w_1 F_t^{ARIMA} + w_2 F_t^{ANN} + \varepsilon_t \quad (4)$$

เมื่อ Y_t แทนค่าสังเกตของข้อมูลอนุกรมเวลา ณ เวลา t ; F_t^{ARIMA} แทนข้อมูลส่วนที่เป็นฟังก์ชันเชิงเส้นตรง

ในเทอมของค่าพารามิเตอร์ พยากรณ์ด้วยตัวแบบ ARIMA; F_t^{ANN} แทนข้อมูลส่วนที่ไม่เป็นฟังก์ชันเชิงเส้นตรงในเทอมของค่าพารามิเตอร์ พยากรณ์ด้วยตัวแบบ ANN; w_1 แทนค่าถ่วงน้ำหนักของข้อมูลส่วนที่

เป็นฟังก์ชันเชิงเส้นตรง คำนวณได้จากการวิเคราะห์การถดถอย; w_2 แทนค่าถ่วงน้ำหนักของข้อมูลส่วนที่ไม่เป็นฟังก์ชันเชิงเส้นตรง คำนวณได้จากการวิเคราะห์การถดถอย; ε_t แทน ค่าคลาดเคลื่อน ณ เวลา t

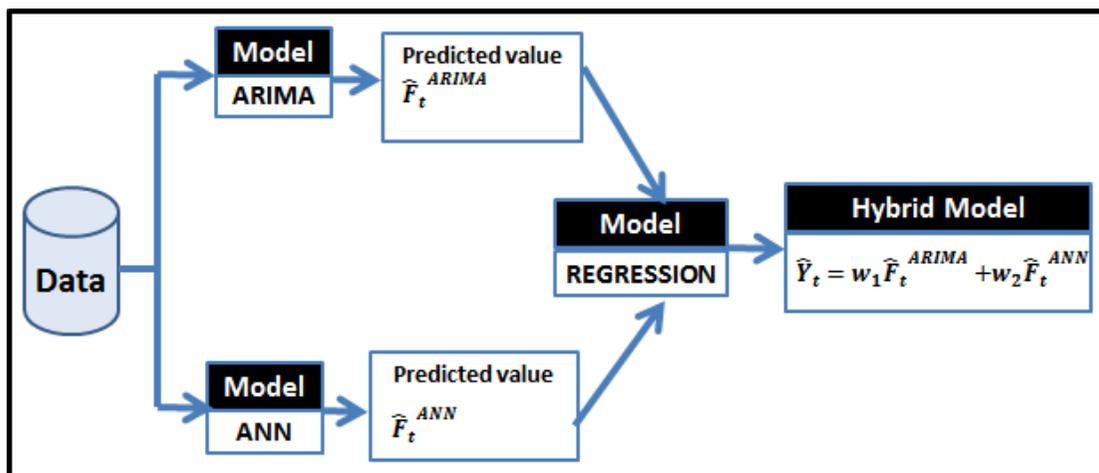


Figure 2 ARIMA-ANN-REG hybrid model

โดยมีขั้นตอนการสร้างตัวแบบผสมอย่างละเอียดดังนี้

ขั้นตอนที่ 1 : วิเคราะห์ข้อมูลอนุกรมเวลาด้วยตัวแบบ ARIMA เพื่อพยากรณ์ข้อมูลส่วนที่เป็นฟังก์ชันเชิงเส้นตรงในเทอมของค่าพารามิเตอร์ คือ F_t^{ARIMA}

ขั้นตอนที่ 2 : วิเคราะห์ข้อมูลอนุกรมเวลาด้วยตัวแบบ ANN เพื่อพยากรณ์ข้อมูลส่วนที่ไม่เป็นฟังก์ชันเชิงเส้นตรงในเทอมของค่าพารามิเตอร์ คือ F_t^{ANN}

ขั้นตอนที่ 3 : นำค่า F_t^{ARIMA} และ F_t^{ANN} มาเป็นตัวแปรต้นในสมการถดถอย และให้ค่าสังเกตของข้อมูล Y_t เป็นตัวแปรตาม

ขั้นตอนที่ 4 : ใช้ค่าสัมประสิทธิ์การถดถอยมาเป็นค่าถ่วงน้ำหนัก w_1 และ w_2 และคำนวณหาค่าพยากรณ์รวม (forecasting) จากสมการ $\hat{Y}_t = w_1 F_t^{ARIMA} + w_2 F_t^{ANN}$

ทั้งนี้งานวิจัยของ Khairalla และคณะ [11] ได้นำตัวแบบผสมโดยใช้วิธีการวิเคราะห์การถดถอย (ARIMA-ANN-REG) ไปใช้ในข้อมูลอัตราแลกเปลี่ยนเงินสกุลยูโร (SDG-EURO) และพบว่าตัวแบบดังกล่าวมีความแม่นยำกว่าตัวแบบผสม ARIMA-ANN และตัวแบบ ARIMA

2.2 การศึกษาเปรียบเทียบในชุดข้อมูลจำลอง

ได้ศึกษาและเปรียบเทียบในชุดข้อมูลจำลอง โดยมีขั้นตอนดังต่อไปนี้

ขั้นตอนที่ 1 : จำลองชุดข้อมูลอนุกรมเวลาทั้งหมด 8 ตัวแบบ ประกอบด้วยตัวแบบดังต่อไปนี้ ARIMA(0,1,1), ตัวแบบ ARIMA(0,1,2), ตัวแบบ ARIMA(1,1,0), ตัวแบบ ARIMA(1,1,1), ตัวแบบ ARIMA(1,1,2), ตัวแบบ ARIMA(2,1,0), ตัวแบบ ARIMA(2,1,1) และตัวแบบ ARIMA(2,1,2) โดยมี

กำหนดค่าพารามิเตอร์ในแต่ละตัวแบบดังแสดงในตารางที่ 2 ซึ่งในแต่ละตัวแบบจะจำลองภายใต้สถานการณ์เดียวกันจำนวน 100 รอบ และในแต่ละรอบของการจำลองจะจำลองขนาดข้อมูลหรือระยะเวลาในแต่ละตัวแบบ (N) 300 จุดเวลา

ขั้นตอนที่ 2 : แบ่งข้อมูลออกเป็นชุดข้อมูลฝึก (training set) จำนวน 210 จุดเวลา (คิดเป็น 70 %) สำหรับใช้ในการสร้างตัวแบบ และชุดข้อมูลทดสอบ (test set) จำนวน 90 จุดเวลา (คิดเป็น 30 %) เพื่อใช้ประเมินความแม่นยำของตัวแบบ โดยตัวแบบที่ดีจะมีค่าเกณฑ์รากของค่าเฉลี่ยของ RMSE ที่ต่ำ

ขั้นตอนที่ 3 : นำข้อมูลฝึกมาสร้างตัวแบบ ARIMA และคัดเลือกตัวแบบที่เหมาะสมที่สุด โดยพิจารณาจากค่า AIC จากนั้นพยากรณ์ในชุดข้อมูลฝึก

และชุดข้อมูลทดสอบด้วยตัวแบบที่คัดเลือก ค่าพารามิเตอร์ RMSE ในแต่ละรอบของการจำลอง และบันทึกค่าเฉลี่ย RMSE จากการจำลอง 100 รอบ

ขั้นตอนที่ 4 : นำข้อมูลฝึกมาสร้างตัวแบบ MLP ประกอบด้วยชั้นข้อมูลนำเข้า 1 ชั้นภายใต้จำนวน 1 node ชั้นซ่อน 1 ชั้นภายใต้จำนวน 1 ถึง 5 node โดยใช้ sigmoid logistic function เป็นฟังก์ชันกระตุ้น และชั้นผลลัพธ์ 1 ชั้นภายใต้จำนวน 1 node โดยใช้ linear function เป็นฟังก์ชันกระตุ้น โดยใช้เทคนิคการเรียนรู้แบบแพร่ย้อนกลับ (resilient back-propagation) จากนั้นคัดเลือกตัวแบบ ANN ที่เหมาะสมที่สุด ซึ่งพิจารณาจากค่า RMSE ที่ต่ำที่สุด และนำไปพยากรณ์ข้อมูลส่วนที่ไม่เป็นฟังก์ชันเชิงเส้นตรงด้วยตัวแบบที่คัดเลือก

Table 2 The 8 ARIMA models and their parameters

Model	Parameters
ARIMA(0,1,1)	$\theta_1 = -0.5$
ARIMA(0,1,2)	$\theta_1 = 0.5, \theta_2 = -0.25$
ARIMA(1,1,0)	$\phi_1 = 0.5$
ARIMA(1,1,1)	$\phi_1 = 0.5, \theta_1 = -0.25$
ARIMA(1,1,2)	$\phi_1 = 0.5, \theta_1 = 0.5, \theta_2 = -0.25$
ARIMA(2,1,0)	$\phi_1 = 0.5, \phi_2 = -0.5$
ARIMA(2,1,1)	$\phi_1 = 0.5, \phi_2 = -0.5, \theta_1 = -0.25$
ARIMA(2,1,2)	$\phi_1 = 0.25, \phi_2 = -0.5, \theta_1 = 0.5, \theta_2 = -0.25$

ขั้นตอนที่ 5 : ในข้อมูลจำลองแต่ละรอบ นำค่าส่วนที่เหลือ (residual) ที่ได้จากการพยากรณ์ด้วยตัวแบบ ARIMA ในชุดข้อมูลฝึกสอนมาสร้างตัวแบบ MLP โดยมีสมบัติเหมือนในขั้นตอนที่ 4 จากนั้นคำนวณค่าพยากรณ์รวม (total forecasting) ด้วยการรวมข้อมูลส่วนที่เป็นฟังก์ชันเชิงเส้นตรงในเทอมของ

ค่าพารามิเตอร์ที่ได้จากตัวแบบ ARIMA ในขั้นตอนที่ 3 และข้อมูลส่วนที่ไม่เป็นฟังก์ชันเชิงเส้นตรงในเทอมของค่าพารามิเตอร์ที่ได้จากการตัวแบบเครือข่ายประสาทเทียมในขั้นตอนนี้เข้าด้วยกัน คำนวณหาค่า RMSE ในแต่ละรอบ และบันทึกค่าเฉลี่ย RMSE ของตัวแบบ ARIMA-ANN จากการจำลอง 100 รอบ

ขั้นตอนที่ 6 : ในข้อมูลจำลองแต่ละรอบ นำค่าพยากรณ์ได้จากการพยากรณ์ด้วยตัวแบบ ARIMA ในขั้นตอนที่ 3 และค่าพยากรณ์ได้จากการพยากรณ์ด้วยตัวแบบ MLP ในขั้นตอนที่ 4 มาเป็นตัวแปรต้นในการวิเคราะห์การถดถอย และใช้สัมประสิทธิ์การถดถอยที่ได้มาเป็นค่าถ่วงน้ำหนักของตัวแบบ ARIMA และ ANN จากนั้นคำนวณค่าพยากรณ์รวม (total forecasting) ด้วยตัวแบบสมการถดถอยดังกล่าว คำนวณหาค่า RMSE ในแต่ละรอบ และบันทึกค่าเฉลี่ย RMSE ของตัวแบบ ARIMA-ANN-REG จากการจำลอง 100 รอบ

ขั้นตอนที่ 7 : เปรียบเทียบความแม่นยำของ

ตัวแบบ ARIMA ตัวแบบ ARIMA-ANN และตัวแบบ ARIMA-ANN-REG โดยใช้เกณฑ์รากของค่าคลาดเคลื่อนกำลังสองเป็นเครื่องมือในการเปรียบเทียบตัวแบบ ซึ่งตัวแบบใดที่มีค่าเฉลี่ยของ RMSE ต่ำสุดจะเป็นตัวแบบที่ให้ผลดีที่สุด

2.3 การศึกษาเปรียบเทียบในชุดข้อมูลจริง

ได้ศึกษาและเปรียบเทียบในชุดข้อมูลจริง โดยมีขั้นตอนดังต่อไปนี้

ขั้นตอนที่ 1 : เก็บรวบรวมชุดข้อมูลอนุกรมเวลาจำนวน 6 ชุดข้อมูล จากแหล่งต่าง ๆ ดังนี้ ตลาดหลักทรัพย์แห่งประเทศไทย กรมการค้าภายใน และสำนักงานเศรษฐกิจอุตสาหกรรม ดังแสดงในตารางที่ 3

Table 3 The 6 real datasets and their related ARIMA models

Time-series dataset	Periods	Time points (<i>N</i>)	ARIMA model
Egg price (monthly)	2008-2018	132	ARIMA(0,1,1)
Import value (monthly)	2011-2017	84	ARIMA(1,1,0)
Pork price (monthly)	2008-2018	132	ARIMA(0,1,2)
Instant noodle price (monthly)	2001-2015	180	ARIMA(1,1,2)
HMPRO stock price (daily)	2018	80	ARIMA(2,1,1)
SIRI stock price (daily)	2018	80	ARIMA(1,1,1)

ขั้นตอนที่ 2 : แบ่งชุดข้อมูลอนุกรมเวลาแต่ละชุดเป็นชุดข้อมูลฝึก (training set) จำนวน 70 % ของจำนวนจุดเวลา เพื่อสำหรับใช้ในการสร้างตัวแบบ และส่วนที่เหลือ 30 % จัดเป็นชุดข้อมูลทดสอบ (test set) เพื่อใช้ประเมินความแม่นยำของตัวแบบ

ขั้นตอนที่ 3 : นำข้อมูลฝึกสอนมาสร้างตัวแบบ ARIMA และคัดเลือกตัวแบบที่เหมาะสมที่สุด โดยพิจารณาจากค่า AIC จากนั้นพยากรณ์ในชุดข้อมูลฝึกและชุดข้อมูลทดสอบด้วยตัวแบบที่คัดเลือก คำนวณค่า RMSE ของแต่ละชุดข้อมูลอนุกรมเวลา

ขั้นตอนที่ 4 : นำข้อมูลฝึกสอนมาสร้างตัวแบบ MLP ภายใต้ขอบเขตงานวิจัยที่ใช้ในชุดข้อมูลจำลอง และนำตัวแบบดังกล่าวพยากรณ์ข้อมูลส่วนที่ไม่เป็นฟังก์ชันเชิงเส้นตรงด้วยตัวแบบที่คัดเลือก

ขั้นตอนที่ 5 : แต่ละชุดข้อมูลข้อมูลอนุกรมเวลา นำค่าส่วนที่เหลือ (residual) ที่ได้จากการพยากรณ์ด้วยตัวแบบ ARIMA ในชุดข้อมูลฝึกมาสร้างตัวแบบ MLP ภายใต้ขอบเขตงานวิจัยที่ใช้ในชุดข้อมูลจำลอง จากนั้นคำนวณค่าพยากรณ์รวม (forecasting) ด้วยการรวมข้อมูลส่วนที่เป็นฟังก์ชันเชิงเส้นตรงใน

เทอมของค่าพารามิเตอร์ที่ได้จากตัวแบบ ARIMA ใน
ขั้นตอนที่ 3 และข้อมูลส่วนที่ไม่เป็นฟังก์ชันเชิงเส้นตรง
ในเทอมของค่าพารามิเตอร์ที่ได้จากการตัวแบบ ANN
ในขั้นตอนนี้เข้าด้วยกัน คำนวณค่า RMSE ของตัวแบบ

ARIMA-ANN ในแต่ละชุดข้อมูลอนุกรมเวลา
ขั้นตอนที่ 6 : แต่ละชุดข้อมูลข้อมูลอนุกรม
เวลานำค่าพยากรณ์ได้จากการพยากรณ์ด้วยตัวแบบ
ARIMA ในขั้นตอนที่ 3 และค่าพยากรณ์ได้จากการ

Table 4 The results of ARIMA, ARIMA-ANN, and ARIMA-ANN-REG on the simulated data

Processes	Models	Averages of RMSE	
		Training set	Test set
ARIMA(0,1,1)	1. ARIMA	5.8892	7.0866
	2. ARIMA-ANN	4.1285	5.3765
	3. ARIMA-ANN-REG	4.0893	5.3208
ARIMA(0,1,2)	1. ARIMA	4.8533	5.7596
	2. ARIMA-ANN	4.5314	5.3201
	3. ARIMA-ANN-REG	3.9795	5.1278
ARIMA(1,1,0)	1. ARIMA	6.8269	9.6872
	2. ARIMA-ANN	6.3847	9.3756
	3. ARIMA-ANN-REG	6.0138	8.9825
ARIMA(1,1,1)	1. ARIMA	4.2479	5.6880
	2. ARIMA-ANN	4.1118	5.5557
	3. ARIMA-ANN-REG	3.9880	5.2110
ARIMA(1,1,2)	1. ARIMA	4.7777	6.8972
	2. ARIMA-ANN	4.6540	6.2221
	3. ARIMA-ANN-REG	4.2854	6.1857
ARIMA(2,1,0)	1. ARIMA	4.0852	5.8882
	2. ARIMA-ANN	4.0627	5.7924
	3. ARIMA-ANN-REG	4.0618	5.6824
ARIMA(2,1,1)	1. ARIMA	3.8528	5.6444
	2. ARIMA-ANN	3.7007	5.0192
	3. ARIMA-ANN-REG	3.6954	5.0009
ARIMA(2,1,2)	1. ARIMA	4.8142	5.5221
	2. ARIMA-ANN	4.8066	5.3218
	3. ARIMA-ANN-REG	4.7240	5.1179

Table 5 The results of ARIMA, ARIMA-ANN, and ARIMA-ANN-REG on the real datasets

Time-series dataset	Models	Averages of RMSE	
		Training set	Test set
Egg price	1. ARIMA	6.8939	10.2570
	2. ARIMA-ANN	6.6501	9.2222
	3. ARIMA-ANN-REG	6.4448	9.1543
Import value	1. ARIMA	17.2887	20.2123
	2. ARIMA-ANN	15.2008	19.8787
	3. ARIMA-ANN-REG	14.1119	18.6450
Pork price	1. ARIMA	8.8899	11.5281
	2. ARIMA-ANN	8.5210	10.8933
	3. ARIMA-ANN-REG	8.4443	10.2256
Instant noodle price	1. ARIMA	5.8993	8.9872
	2. ARIMA-ANN	4.9777	8.0005
	3. ARIMA-ANN-REG	4.2368	7.9892
HMPRO stock price	1. ARIMA	3.6684	5.8722
	2. ARIMA-ANN	3.5972	5.2121
	3. ARIMA-ANN-REG	3.0821	5.1009
SIRI stock price	1. ARIMA	4.8957	6.2121
	2. ARIMA-ANN	3.8280	5.8883
	3. ARIMA-ANN-REG	3.7263	5.4214

พยากรณ์ด้วยตัวแบบ MLP ในขั้นตอนที่ 4 มาเป็นตัวแปรต้นในการวิเคราะห์การถดถอย และใช้สัมประสิทธิ์การถดถอยที่ได้มาเป็นค่าถ่วงน้ำหนักของตัวแบบ ARIMA และ ANN หลังจากนั้นคำนวณค่าพยากรณ์รวม (forecasting) ด้วยตัวแบบสมการถดถอยดังกล่าว คำนวณหาค่า RMSE ของตัวแบบ ARIMA-ANN-REG ในแต่ละชุดข้อมูลอนุกรมเวลา

3. ผลการวิจัย

งานวิจัยนี้นำเสนอผลความแม่นยำของตัวแบบ ARIMA ตัวแบบ ARIMA-ANN และตัวแบบ ARIMA-

ANN-REG โดยแบ่งเป็น 2 ส่วน ส่วนที่ 1 เปรียบเทียบความแม่นยำของค่าพยากรณ์ โดยใช้ชุดข้อมูลจำลอง และส่วนที่ 2 เปรียบเทียบความแม่นยำของค่าพยากรณ์ โดยใช้ชุดข้อมูลจริง 6 ชุดอนุกรมเวลา โดยผลการเปรียบเทียบสำหรับทั้งสองส่วนแสดงในตารางที่ 4 และ 5 ตามลำดับ

ตารางที่ 4 พบว่าชุดข้อมูลการจำลองด้วยตัวแบบอนุกรมเวลา ARIMA จำนวน 8 ตัวแบบ สรุปได้ว่าตัวแบบ ARIMA-ANN-REG มีความแม่นยำในการพยากรณ์สูงสุด รองลงมา คือ ตัวแบบ ARIMA-ANN ส่วนตัวแบบ ARIMA มีความแม่นยำในการพยากรณ์ต่ำ

ที่สุด

ตารางที่ 5 พบว่าชุดข้อมูลจริง จำนวน 6 ชุด ข้อมูล สรุปได้ว่าตัวแบบ ARIMA-ANN-REG มีความแม่นยำในการพยากรณ์สูงสุด รองลงมา คือ ตัวแบบ ARIMA-ANN ส่วนตัวแบบ ARIMA มีความแม่นยำในการพยากรณ์ต่ำที่สุด ซึ่งผลการเปรียบเทียบที่ได้จากข้อมูลจริงสอดคล้องกับผลการเปรียบเทียบในชุดข้อมูลจำลอง

4. สรุป

งานวิจัยนี้นำเสนอผลการวิจัยจากการเปรียบเทียบความแม่นยำของตัวแบบ ARIMA ตัวแบบ ARIMA-ANN และตัวแบบ ARIMA-ANN-REG โดยใช้ชุดข้อมูลจริงและชุดข้อมูลอนุกรมเวลาที่จำลองด้วยตัวแบบ ARIMA และใช้เกณฑ์รากของค่าคลาดเคลื่อนกำลังสอง (RMSE) ในการตัดสินใจ สรุปได้ว่าตัวแบบ ARIMA-ANN-REG มีความแม่นยำในการพยากรณ์สูงสุด รองลงมา คือ ตัวแบบ ARIMA-ANN ส่วนตัวแบบ ARIMA มีความแม่นยำในการพยากรณ์ต่ำที่สุด ซึ่งให้ผลที่สอดคล้องกันทั้งในชุดข้อมูลจริงและชุดข้อมูลจำลอง ดังนั้นการพิจารณาเลือกใช้ตัวแบบผสมเพื่อใช้ในการพยากรณ์จัดว่าเป็นอีกทางเลือกที่น่าสนใจ สามารถช่วยลดความผิดพลาดในการพยากรณ์เมื่อเปรียบเทียบกับการใช้ตัวแบบ ARIMA เพียงตัวแบบเดียว แต่ทั้งนี้ตัวแบบผสมก็มีข้อจำกัดในเรื่องระยะเวลาที่ต้องใช้ในการประมวลผลที่นานกว่าเมื่อเปรียบเทียบกับตัวแบบเดี่ยว

5. ข้อเสนอแนะ

งานวิจัยนี้เลือกใช้ตัวแบบในการเปรียบเทียบความแม่นยำ 3 ตัวแบบ คือ ตัวแบบ ARIMA ตัวแบบ ARIMA-ANN และตัวแบบ ARIMA-ANN-REG ผู้ที่สนใจสามารถศึกษาเพิ่มเติมจากงานวิจัยนี้โดยนำหลักการ

ของ Khairalla และคณะ [11] ไปประยุกต์ใช้กับตัวแบบผสมชนิดอื่น เช่น ตัวแบบผสมระหว่าง ARIMA กับซัพพอร์ตเวกเตอร์แมชชีน (ARIMA-SVM) และผู้ที่สนใจข้อมูลอนุกรมเวลาที่มีฤดูกาลอาจลองพิจารณาตัวแบบ seasonal integrated autoregressive moving average (SARIMA) เพิ่มเติม และนำตัวแบบดังกล่าวมาสร้างเป็นตัวแบบผสมได้ หรือนำหลักการของการทำ validation มาพิจารณา เช่น การสุ่มหลาย ๆ ครั้งของข้อมูลชุดฝึกฝนกับชุดทดสอบ การแบ่งด้วยร้อยละแบบอื่น การทำ K-fold validation และแนวคิดของ Bagging, Boosting มาร่วมพิจารณา ซึ่งจะทำให้งานวิจัยสมบูรณ์มากขึ้น

6. References

- [1] Taesombat, S. , 2010, Quantitative Forecasting, Kasetsart University Press, Bangkok, 487 p. (in Thai)
- [2] Shumway, R.H. and Stoffer, D.S. , 2010, Time Series Analysis and Its Applications with R Examples, 3rd Ed., Springer, Berlin, 576 p.
- [3] Zhang, P.G., 2003, Time series forecasting using a hybrid ARIMA and neural network model, Neurocomputing 50: 159-175.
- [4] Faruk, D.O., 2010, A hybrid neural network and ARIMA model for water quality time series prediction, Eng. Appl. Artif. Intell. 23: 586-594.
- [5] Meth, N., Saxena, V.P. and Pardasani, K.R., 2010, A comparison between hybrid approaches of ANN and ARIMA for Indian stock trend forecasting, Bus. Intell. J. 3: 23-43.

- [6] Koutroumanidis, T. , Ioannou, K. and Arabatzis, G. , 2009, Predicting fuel wood prices in Greece with the use of ARIMA models artificial neural networks and a hybrid ARIMA-ANN model, *Energy Pol.* 37: 3627-3634.
- [7] Ebrahimi, A., 2019, Time series forecasting of styrene price using a hybrid ARIMA and neural network model, *Indep. J. Manag. Prod.* 10: 915-933.
- [8] Siripanich, P., Nillaporn, P. and Trakarnta Ierngsuk, S., 2007, Time Series Forecasting Using a Combined ARIMA and Artificial Neural Network Model of Styrene Price Using a Hybrid ARIMA and Neural Network Model, *Operations Research Network* 2007, 7 p. (in Thai)
- [9] Boonmana, C. and Kulvanich, N., 2017, A comparative prediction accuracy of hybrid time series models, *Thai Sci. Technol. J.* 25(2): 177-190. (in Thai)
- [10] Somsila, C. , Chiewchanwattana, S. and Sunat, K. , 2010, Hybrid Model for Air Quality Data Prediction Case Study: Air Quality Data in Thailand, *The 11th Graduate Research Conference*, 12 p. (in Thai)
- [11] Khairalla, M., Xu-Ning and AL-Jallad, N.T., 2017, Hybrid forecasting scheme for financial time- series data using neural network and statistical methods, *IJACSA* 8: 319-327.
- [12] Zhang, P. G. , 2007, A neural network ensemble method with jittered training data for time series forecasting, *Inform. Sci.* 177: 5329-5346.