บทความวิจัย

วิธีการพยากรณ์ข้อมูลล่วงหน้าหนึ่งคาบเวลาสำหรับตัวแบบอัตตสหสัมพันธ์อันดับที่หนึ่ง เมื่อข้อมูลมีค่าสูญหาย : การศึกษาเชิงจำลอง

THE ONE-STEP-AHEAD PREDICTION METHODS FOR FIRST-ORDER AUTOREGRESSIVE MODEL WITH MISSING VALUES : A SIMULATION STUDY

วราฤทธิ์ พานิชกิจโกศลกุล
ผู้ช่วยศาสตราจารย์ประจำภาควิชาคณิตศาสตร์และสถิติ
คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์
E-mail: wararit@mathstat.sci.tu.ac.th

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบวิธีการพยากรณ์ข้อมูลล่วงหน้าหนึ่งคาบเวลาสำหรับตัวแบบอัตตสหสัมพันธ์อันดับที่หนึ่ง เมื่อข้อมูลมีค่าสูญหายแบบสุ่มด้วยวิธีการพยากรณ์ 3 วิธี คือ วิธีการพยากรณ์โดยใช้วิธีกำลังสองน้อยที่สุดแบบค่าเฉลี่ยเวียนเกิด (RM) วิธีการพยากรณ์โดยใช้วิธีกำลังสองน้อยที่สุดแบบมัธยฐานเวียนเกิด (RMD) และวิธีการพยากรณ์โดยใช้วิธีกำลังสองน้อยที่สุดแบบมัธยฐานเวียนเกิดปรับปรุง (IRMD) กำหนดค่าพารามิเตอร์ของตัวแบบอัตตสหสัมพันธ์อันดับที่หนึ่ง (ρ) 9 ระดับ คือ ระดับต่ำ 0.1-0.3 ระดับปานกลาง คือ 0.4-0.6 และระดับสูง คือ 0.7-0.9 ร้อยละของค่าสูญหายเท่ากับ ร้อยละ 5 และ 10 ขนาดตัวอย่าง 4 ระดับ คือ 25, 50, 100 และ 250 ในการวิจัยครั้งนี้ใช้เทคนิคการจำลองแบบมอนติคาร์โล และทำการทดลองซ้ำๆ กัน 100,000 ครั้งในแต่ละสถานการณ์ เพื่อคำนวณค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของค่าพยากรณ์ (PMSE) ผลการวิจัยสรุปได้ดังนี้ สำหรับทุกระดับของร้อยละของค่าสูญหาย วิธี RM ให้ค่า PMSE ต่ำที่สุด เมื่อตัวอย่างมีขนาดใหญ่ (n = 250) หรือตัวอย่างมีขนาดเล็กและปานกลาง (n = 25, 50 และ 100) และค่า ρ มีค่าน้อย (0.1 ถึง 0.5) วิธี RMD ให้ค่า PMSE ต่ำที่สุด เมื่อตัวอย่างมีขนาดเล็ก (n = 25) และค่ามีค่าน้อยและปานกลาง (0.2 ถึง 0.7) และ วิธี IRMD ให้ค่า PMSE ต่ำที่สุด เมื่อตัวอย่างมีขนาดเล็ก (n = 25) และค่า ρ มีค่ามาก (0.8 ถึง 0.9)

คำสำคัญ : การพยากรณ์ล่วงหน้าหนึ่งคาบเวลา ตัวแบบ AR(1) ค่าสูญหาย

ABSTRACT

The objective of this research is to compare the one-step-ahead prediction methods for first-order autoregressive model with missing values which missing at random. These methods are the prediction method using recursive mean OLS method (RM), using recursive median OLS method (RMD), and using improved recursive median OLS method

(IRMD). There are 9 values for the parameter of the first-order autoregressive model (ρ): low level (ρ = 0.1-0.3), medium level (ρ = 0.4-0.6) and high level (ρ = 0.7-0.9). The percentages of missing values are 5% and 10%. The sample sizes are 25, 50, 100, and 250. This research uses the Monte Carlo simulation technique. The experiment was repeated 100,000 times for each condition to calculate the prediction mean squared error (PMSE). Results of the research are as follows: For all percentages of missing values, a RM method provides the lowest PMSE when = 250, and when n = 25, 50 and 100 and ρ = 0.1-0.5. The PMSE of a RMD method is the lowest when n = 25 and ρ = 0.2-0.7. A IRMD method provides the lowest PMSE when n = 25 and ρ = 0.8-0.9.

KEYWORDS : One-step-ahead prediction, AR(1) model, Missing values

บทนำ

การพยากรณ์อนุกรมเวลา (Time Series Forecasting) เป็นวิธีการพยากรณ์วิธีหนึ่งที่นิยมใช้กันมาก วิธีนี้จะใช้ข้อมูลในอดีต โดยจะศึกษาถึงลักษณะการเปลี่ยนแปลงของข้อมูลเมื่อเวลาเปลี่ยนไปว่ามีลักษณะเช่นไร และทำการกำหนดรูปแบบของการแปรเปลี่ยนที่เหมาะสมที่สุดสำหรับข้อมูลนั้น โดยจะอยู่ในรูปของความสัมพันธ์กับเวลา การพยากรณ์อนุกรมเวลามีอยู่ด้วยกันหลายวิธี เช่น เทคนิคการทำให้เรียบ (Smoothing Techniques) การกรองแบบปรับได้ (Adaptive Filtering) วิธีอนุกรมเวลาแบบคลาสสิค (Classical Time Series Methods) และวิธีอนุกรมเวลาแบบบ็อกซ์-เจนกินส์ (Box-Jenkins Methods) เป็นต้น

นอกเหนือจากการเลือกใช้วิธีการพยากรณ์ที่เหมาะสมและการกำหนดตัวแบบอนุกรมเวลาแล้ว วิธีการประมาณค่าพารามิเตอร์ของตัวแบบอนุกรมเวลาก็เป็นอีกปัจจัยหนึ่งที่ส่งผลถึงความแม่นยำและเชื่อถือได้ของค่าพยากรณ์ (วราฤทธิ์ พานิชกิจโกศลกุล, 2545: 1) วิธีการประมาณค่าพารามิเตอร์ของตัวแบบอนุกรมเวลามีอยู่หลายวิธี เช่น วิธีกำลังสองน้อยที่สุด (Ordinary Least Squares Method: OLS) วิธีการประมาณความควรจะเป็นสูงสุด (Maximum Likelihood Estimation Method: MLE) เป็นต้น โดยแต่ละวิธีการจะมีหลักการในการหาตัวประมาณค่าแตกต่างกันไป อีกทั้งยังให้ผลลัพธ์ที่แตกต่างกันออกไปด้วย ซึ่งวิธีการประมาณค่าพารามิเตอร์นั้นได้มีผู้ศึกษาจากอดีตถึงปัจจุบัน ภายใต้ข้อสมมติเบื้องต้นที่แตกต่างกันออกไป อาทิ So and Shin (1999) ได้พัฒนาวิธีกำลังสองน้อยที่สุดแบบค่าเฉลี่ยเวียนเกิด (Recursive Mean OLS Method) สำหรับตัวแบบอัตตสหสัมพันธ์อันดับที่หนึ่ง (First-Order Autoregressive Model: AR(1)) Vougas (2000) ได้พัฒนาวิธีการประมาณจีเอ็มเอ็ม (Generalized Method of Moment Method) สำหรับตัวแบบ AR(1) สอาด นิวิศพงศ์ (2004) ได้ปรับปรุงวิธีการประมาณจีเอ็มเอ็มของวาวกัส โดยใช้แนวคิดของ So and Shin (1999) วราฤทธิ์ พานิชกิจโกศลกุล (2552) ได้พัฒนาวิธีประมาณค่าพารามิเตอร์ของตัวแบบ AR(1) เมื่อข้อมูลมีค่าสูญหาย

หลังจากประมาณค่าพารามิเตอร์ด้วยวิธีที่เหมาะสมแล้ว ขั้นต่อไปจะหาค่าพยากรณ์ล่วงหน้าหนึ่งคาบเวลา (One-Step-Ahead Prediction Value) ซึ่งข้อมูลในอดีตก็เป็นอีกปัจจัยหนึ่งที่ส่งผลถึงความถูกต้องแม่นยำและความเชื่อถือได้ของค่าพยากรณ์ เนื่องจากการวิเคราะห์อนุกรมเวลาจะอาศัยข้อมูลในอดีตที่ผ่านมา แต่ผู้วิเคราะห์อาจประสบปัญหาเกี่ยวกับข้อมูล เช่น ข้อมูลบางค่าสูญหาย (Missing Values) ทำให้ผลการวิเคราะห์และการพยากรณ์คลาดเคลื่อน หรืออาจนำไปสู่การตัดสินใจหรือวางแผนที่ผิดพลาดได้ ดังนั้นวัตถุประสงค์ของการวิจัยครั้งนี้ คือ การเปรียบเทียบวิธีการพยากรณ์ข้อมูลล่วงหน้าหนึ่งคาบเวลา เมื่อข้อมูลมีค่าสูญหาย โดยประยุกต์ใช้วิธีการประมาณค่าพารามิเตอร์ของตัวแบบ AR(1) เมื่อข้อมูลมีค่าสูญหาย ซึ่งเสนอโดยวราฤทธิ์ พานิชกิจโกศลกุล (2552) และวิธีกำลังสองน้อยที่สุดแบบค่าเฉลี่ยเวียนเกิดซึ่งเสนอโดย So and Shin (1999)

ข้อมูลอนุกรมเวลาที่ศึกษาในครั้งนี้เป็นอนุกรมเวลาตัวแบบอัตตสหสัมพันธ์อันดับที่หนึ่ง (AR(1)) เมื่อข้อมูลมีค่าสูญหาย

จากนั้นทำการเปรียบเทียบประสิทธิภาพของทุกวิธีด้วยค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าพยากรณ์ (Prediction Mean Square Error: PMSE)

ขอบเขตของการวิจัย

1. อนุกรมเวลา เป็นอนุกรมเวลาตัวแบบอัตตสหสัมพันธ์อันดับที่หนึ่ง AR(1) เขียนตัวแบบได้ดังนี้ (Wei, 2006: 34)

$$Y_t - \mu = \rho(Y_{t-1} - \mu) + a_t$$

ในที่นี้กำหนดค่าเฉลี่ยของอนุกรมเวลา (μ) เท่ากับ 0

2. กำหนดค่าพารามิเตอร์ (ρ) ของตัวแบบ 9 ระดับ คือ 0.1 ถึง 0.9 เพิ่มขึ้นครั้งละ 0.1

3. การแจกแจงของความคลาดเคลื่อนสุ่ม a_t มีการแจกแจงแบบปกติ (Normal Distribution) โดยมีฟังก์ชันความหนาแน่นของความน่าจะเป็น

$$f(a) = \frac{1}{\sigma_a\sqrt{2\pi}}\exp(-\frac{1}{2\sigma_a^2}(a-\mu_a)^2)$$

ในที่นี้กำหนดให้ $\mu_a = 0$ และ $\sigma_a = 1$

4. ขนาดตัวอย่าง (n) ที่ใช้ในการศึกษา มี 4 ระดับ คือ 25, 50, 100 และ 250

5. ร้อยละของค่าสูญหาย (p) เท่ากับ ร้อยละ 5 และ 10 และตำแหน่งของค่าสูญหายเป็นไปอย่างสุ่ม ซึ่งอยู่ระหว่างเวลาที่ $t=2$ ถึง $t=n-1$

6. โปรแกรมที่ใช้ในการวิจัยครั้งนี้ทั้งหมดเขียนด้วยโปรแกรม R เวอร์ชัน 2.8.1 ซึ่งทำการทดลองซ้ำ 100,000 ครั้ง ในแต่ละสถานการณ์ของการทดลอง

วิธีดำเนินการวิจัย

ในการวิจัยครั้งนี้มีวิธีดำเนินการวิจัยดังนี้

1. จำลองค่าความคลาดเคลื่อนสุ่ม a_t

การสร้างการแจกแจงของความคลาดเคลื่อนสุ่มที่ต้องการศึกษา มีการแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1

2. จำลองข้อมูลอนุกรมเวลา AR(1) ที่มีค่าสูญหาย

สร้าง Y_0 ให้มีการแจกแจงแบบปกติ ซึ่งมีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ $\frac{\sigma_a^2}{1-\rho^2} = \frac{1}{1-\rho^2}$ และสร้าง $a_t\, ; t = 1, 2, ..., n$ ให้มีการแจกแจงแบบปกติ ซึ่งมีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 จากนั้นสร้าง $Y_t\, ; t = 1, 2, ..., n$ ให้มีรูปแบบความสัมพันธ์ คือ

$$Y_t = \rho Y_{t-1} + a_t$$

เมื่อสร้างข้อมูลอนุกรมเวลา AR(1) แล้ว ต่อไปจะทำการสร้างข้อมูลอนุกรมเวลา AR(1) ที่มีค่าสูญหาย โดยสุ่มเลือกตำแหน่งของค่าสูญหายระหว่างเวลาที่ $t=2$ ถึง $t=n-1$ จากนั้นบันทึกค่า Y_t ในคาบเวลานั้นให้เป็นค่าสูญหาย

3. ประมาณค่าสูญหาย

สมมติว่า Y_t เป็นค่าสูญหาย ประมาณค่า Y_t ด้วย $E(Y_T \mid Y_1, Y_2, ..., Y_{T-1}) = Y_{T-1}$ (Shin and Sarkar, 1996) นั่นคือ ค่าประมาณของ Y_T คือ Y_{T-1}

4. ประมาณค่าพารามิเตอร์ของข้อมูลอนุกรมเวลาทั้ง 3 วิธี

1) วิธีกำลังสองน้อยที่สุดแบบค่าเฉลี่ยเวียนเกิด So and Shin (1999) ได้พัฒนาวิธีกำลังสองน้อยที่สุดแบบค่าเฉลี่ยเวียนเกิด โดยมีหลักการเช่นเดียวกับวิธีกำลังสองน้อยที่สุด คือการทำให้ผลรวมค่ากำลังสองของความคลาดเคลื่อนภายใต้พารามิเตอร์ของตัวแบบมีค่าต่ำที่สุด แต่วิธีกำลังสองน้อยที่สุดแบบค่าเฉลี่ยเวียนเกิดจะประมาณค่าเฉลี่ยของอนุกรมเวลา (μ) ด้วยค่าเฉลี่ยเวียนเกิด (Recursive Mean: $\bar{Y}_t$) แทนค่าเฉลี่ยเลขคณิต (Arithmetic Mean: $\bar{Y}$) ตัวประมาณกำลังสองน้อยที่สุดแบบค่าเฉลี่ยเวียนเกิด คือ

$$\hat{\rho}_{RM} = \frac{\sum_{t=2}^{n}(Y_t - \bar{Y}_{t-1})(Y_{t-1} - \bar{Y}_{t-1})}{\sum_{t=2}^{n}(Y_{t-1} - \bar{Y}_{t-1})^2}$$

โดยที่ $\bar{Y}_t = \frac{\sum_{i=1}^{t} Y_i}{t}$

2) วิธีกำลังสองน้อยที่สุดแบบมัธยฐานเวียนเกิด วราฤทธิ์ พานิชกิจโกศลกุล (2552) ได้พัฒนาวิธีกำลังสองน้อยที่สุด

แบบมัธยฐานเวียนเกิดโดยใช้แนวคิดของ So and Shin (1999) กล่าวคือ วิธีนี้มีหลักการเช่นเดียวกับวิธีกำลังสองน้อยที่สุด และประมาณค่าเฉลี่ยของอนุกรมเวลา (μ) ด้วยค่ามัธยฐานเวียนเกิด (Recursive Median: $\tilde{Y}_t$) แทนค่าเฉลี่ยเลขคณิต (Arithmetic Mean: $\bar{Y}$) ตัวประมาณกำลังสองน้อยที่สุดแบบมัธยฐานเวียนเกิด คือ

$$\hat{\rho}_{IRMD} = \frac{\sum_{t=2}^{n}(Y_t-\tilde{Y}_t)(Y_{t-1}-\tilde{Y}_{t-1})}{\sum_{t=2}^{n}(Y_{t-1}-\tilde{Y}_{t-1})^2}$$

โดยที่ $\tilde{Y}_t = median\{Y_1, Y_2, \ldots, Y_t\}$

3) **วิธีกำลังสองน้อยที่สุดแบบมัธยฐานเวียนเกิดปรับปรุง** วราฤทธิ์ พานิชกิจโกศลกุล (2552) ยังได้ปรับปรุงประสิทธิภาพของวิธีกำลังสองน้อยที่สุดแบบมัธยฐานเวียนเกิด โดยนำค่ามัธยฐานเวียนเกิด มาหาค่าเฉลี่ยเวียนเกิดอีกครั้งหนึ่ง ซึ่งตัวประมาณกำลังสองน้อยที่สุดแบบมัธยฐานเวียนเกิดปรับปรุง ($\ddot{Y}_t$) คือ

$$\hat{\rho}_{RMD} = \frac{\sum_{t=2}^{n}(Y_t-\ddot{Y}_t)(Y_{t-1}-\ddot{Y}_{t-1})}{\sum_{t=2}^{n}(Y_{t-1}-\ddot{Y}_{t-1})^2}$$

โดยที่ $\ddot{Y}_t = \frac{\sum_{i=1}^{t}\tilde{Y}_i}{t}$ และ $\tilde{Y}_t = median\{Y_1, Y_2, \ldots, Y_t\}$

5. **คำนวณค่าพยากรณ์ล่วงหน้าหนึ่งคาบเวลาของตัวแบบ** AR(1)

1) **วิธีการพยากรณ์โดยใช้วิธีกำลังสองน้อยที่สุดแบบค่าเฉลี่ยเวียนเกิด** ค่าพยากรณ์ล่วงหน้าหนึ่งคาบเวลา คือ

$$\hat{Y}_{n+1}^{RM} = \bar{Y} + \hat{\rho}_{RM}(Y_n - \bar{Y})$$

2) **วิธีการพยากรณ์โดยใช้วิธีกำลังสองน้อยที่สุดแบบมัธยฐานเวียนเกิด** ค่าพยากรณ์ล่วงหน้าหนึ่งคาบเวลา คือ

$$\hat{Y}_{n+1}^{RMD} = \bar{Y} + \hat{\rho}_{RMD}(Y_n - \bar{Y})$$

3) **วิธีการพยากรณ์โดยใช้วิธีกำลังสองน้อยที่สุด**แบบมัธยฐานเวียนเกิดปรับปรุง ค่าพยากรณ์ล่วงหน้าหนึ่งคาบเวลา คือ

$$\hat{Y}_{n+1}^{IRMD} = \bar{Y} + \hat{\rho}_{IRMD}(Y_n - \bar{Y})$$

6. **คำนวณค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าพยากรณ์ (PMSE) แล้วทำการเปรียบเทียบ**

$$\mathrm{PMSE} = \frac{\sum_{i=1}^{M}(Y_{n+1}-\hat{Y}_{n+1}^{K})^2}{M}$$

เมื่อ $\hat{Y}_{n+1}^{K}$ แทน ค่าพยากรณ์ล่วงหน้าหนึ่งคาบเวลา

K แทน วิธีการพยากรณ์ ซึ่งในที่นี้คือ RM, RMD และ IRMD

M แทน จำนวนรอบของการทำซ้ำ

ผลการวิจัย

ในการนำเสนอผลการวิจัยเพื่อความสะดวกจะใช้สัญลักษณ์ต่างๆ แทนความหมาย ดังนี้

RM หมายถึง วิธีการพยากรณ์โดยใช้วิธีกำลังสองน้อยที่สุดแบบค่าเฉลี่ยเวียนเกิด

RMD หมายถึง วิธีการพยากรณ์โดยใช้วิธีกำลังสองน้อยที่สุดแบบมัธยฐานเวียนเกิด

IRMD หมายถึง วิธีการพยากรณ์โดยใช้วิธีกำลังสองน้อยที่สุดแบบมัธยฐานเวียนเกิดปรับปรุง

PMSE หมายถึง ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าพยากรณ์

ผลการเปรียบเทียบประสิทธิภาพของวิธีการพยากรณ์ข้อมูลล่วงหน้าหนึ่งคาบเวลาทั้ง 3 วิธี นำเสนอดังตารางที่ 1 และ 2 ในที่นี้จะพิจารณาค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าพยากรณ์ (PMSE) จำแนกตามขนาดตัวอย่าง ดังนี้

เมื่อขนาดตัวอย่างเท่ากับ 25

เมื่อร้อยละของค่าสูญหาย เท่ากับ ร้อยละ 5 สรุปได้ดังนี้ ในกรณีที่ค่า ρ มีค่าเท่ากับ 0.1 วิธี RM ให้ค่า PMSE ต่ำที่สุด ส่วนกรณีที่ค่า ρ มีค่าตั้งแต่ 0.2 ถึง 0.6 วิธี RMD ให้ค่า PMSE ต่ำที่สุด และกรณีที่ค่า ρ มีค่าตั้งแต่ 0.7 ถึง 0.9 วิธี IRMD ให้ค่า PMSE ต่ำที่สุด

เมื่อร้อยละของค่าสูญหาย เท่ากับ ร้อยละ 10 สรุปได้ดังนี้ ในกรณีที่ค่า ρ มีค่าตั้งแต่ 0.1 ถึง 0.7 วิธี RM ให้ค่า PMSE ต่ำที่สุด ส่วนกรณีที่ค่า ρ มีค่าเท่ากับ 0.8 และ 0.9 วิธี

IRMD ให้ค่า PMSE ต่ำที่สุด

เมื่อขนาดตัวอย่างเท่ากับ 50

เมื่อร้อยละของค่าสูญหาย เท่ากับ ร้อยละ 5 สรุปได้ดังนี้ ในกรณีที่ค่า ρ มีค่าตั้งแต่ 0.1 ถึง 0.5 วิธี RM ให้ค่า PMSE ต่ำที่สุด ส่วนกรณีที่ค่า ρ มีค่าเท่ากับ 0.6 และ 0.7 วิธี RMD ให้ค่า PMSE ต่ำที่สุด และกรณีที่ค่า ρ มีค่าเท่ากับ 0.8 และ 0.9 วิธี IRMD ให้ค่า PMSE ต่ำที่สุด

เมื่อร้อยละของค่าสูญหาย เท่ากับ ร้อยละ 10 สรุปได้ดังนี้ ในกรณีที่ค่า ρ มีค่าตั้งแต่ 0.1 ถึง 0.8 วิธี RM ให้ค่า PMSE ต่ำที่สุด ส่วนกรณีที่ค่า ρ มีค่าเท่ากับ 0.9 วิธี IRMD ให้ค่า PMSE ต่ำที่สุด

เมื่อขนาดตัวอย่างเท่ากับ 100

เมื่อร้อยละของค่าสูญหาย เท่ากับ ร้อยละ 5 สรุปได้ดังนี้ ในกรณีที่ค่า ρ มีค่าตั้งแต่ 0.1 ถึง 0.7 วิธี RM ให้ค่า PMSE ต่ำที่สุด ส่วนกรณีที่ค่า ρ มีค่าเท่ากับ 0.8 วิธี RMD ให้ค่า PMSE ต่ำที่สุด และกรณีที่ค่า ρ มีค่าเท่ากับ 0.9 วิธี IRMD ให้ค่า PMSE ต่ำที่สุด

เมื่อร้อยละของค่าสูญหาย เท่ากับ ร้อยละ 10 สรุปได้ดังนี้ ในกรณีที่ค่า ρ มีค่าตั้งแต่ 0.1 ถึง 0.8 วิธี RM ให้ค่า PMSE ต่ำที่สุด ส่วนกรณีที่ค่า ρ มีค่าเท่ากับ 0.9 วิธี IRMD ให้ค่า PMSE ต่ำที่สุด

เมื่อขนาดตัวอย่างเท่ากับ 250

เมื่อร้อยละของค่าสูญหาย เท่ากับ ร้อยละ 5 และร้อยละ 10 สรุปได้ดังนี้ วิธี RM ให้ค่า PMSE ต่ำที่สุด ในทุกระดับของค่า ρ (0.1 ถึง 0.9)

เมื่อพิจารณาในภาพรวมถึงร้อยละของค่าสูญหาย พบว่า เมื่อข้อมูลมีค่าร้อยละของค่าสูญหายเพิ่มขึ้น ค่า PMSE จะมีค่าเพิ่มสูงขึ้นตามไปด้วย ส่วนวิธีการที่ให้ค่า PMSE ต่ำที่สุดของตารางที่ 1 และ 2 ไม่แตกต่างกันมากนัก

ตัวอย่างการประยุกต์ใช้ผลการวิจัย

ในส่วนนี้จะนำเสนอตัวอย่างการประยุกต์ใช้ผลการวิจัย ดังนี้

ในกรณีนี้จะวิเคราะห์การเปลี่ยนแปลงในทรัพย์สินทางธุรกิจ โดยพิจารณาที่อัตรารายปี (Annual Rates) ลักษณะข้อมูลเป็นแบบรายไตรมาส จำนวน 60 ค่า ระหว่างไตรมาสแรกของปี 1955 ถึงไตรมาสที่ 4 ของปี 1969 ข้อมูลนำมาจาก Pankratz (1983) ในที่นี้สมมติว่าข้อมูลตัวที่ 30 เป็นข้อมูลสูญหาย ประมาณค่าข้อมูลสูญหายด้วยข้อมูลตัวที่ 29 กราฟอนุกรมเวลา กราฟฟังก์ชันอัตตสหสัมพันธ์ (ACF) และกราฟฟังก์ชันอัตตสหสัมพันธ์บางส่วน (PACF) แสดงดังภาพที่ 1-3

จากการพิจารณากราฟ PACF ในภาพที่ 3 พบว่า มีลักษณะคล้ายกับ PACF ที่มีค่าเท่ากับศูนย์ตั้งแต่ Lag ที่ 2 เป็นต้นไป ดังนั้น ตัวแบบที่เป็นไปได้คือ ตัวแบบอัตตสหสัมพันธ์อันดับที่หนึ่ง หรือ ตัวแบบ AR(1)

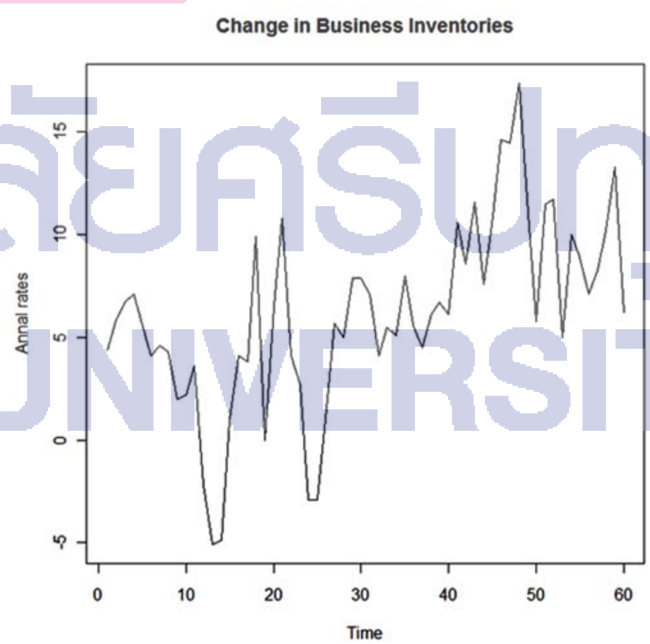
ขั้นตอนถัดไปประมาณค่าเฉลี่ย ($\overline{Y}$) และประมาณค่าพารามิเตอร์ ด้วยวิธีการประมาณ 3 วิธี ได้ผลลัพธ์ดังนี้

$$\overline{Y} = 6.128333$$

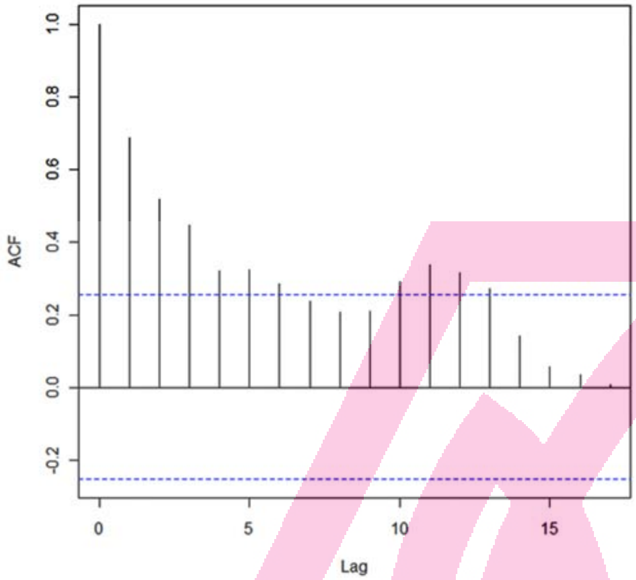
$$\hat{\rho}_{RM} = 0.691341$$

$$\hat{\rho}_{RMD} = 0.674593$$

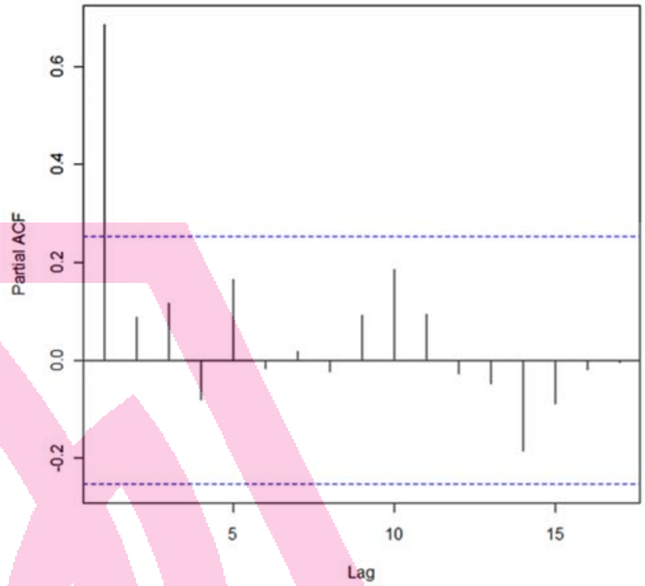
$$\hat{\rho}_{IRMD} = 0.727667$$



ภาพที่ 1 กราฟอนุกรมเวลา



ภาพที่ 2 กราฟฟังก์ชันอัตตสหสัมพันธ์



ภาพที่ 3 กราฟฟังก์ชันอัตตสหสัมพันธ์บางส่วน

จากนั้นคำนวณค่าพยากรณ์ล่วงหน้าหนึ่งคาบเวลา โดยใช้วิธีการที่นำเสนอ 3 วิธี คือ

$$\hat{Y}_{61}^{RM} = \bar{Y} + \hat{\rho}_{RM}(Y_{60} - \bar{Y})$$
$$= 6.128333 + 0.691341(6.2 - 6.128333)$$
$$= 6.177879$$

$$\hat{Y}_{61}^{RMD} = \bar{Y} + \hat{\rho}_{RMD}(Y_{60} - \bar{Y})$$
$$= 6.128333 + 0.674593(6.2 - 6.128333)$$
$$= 6.176679$$

$$\hat{Y}_{61}^{IRMD} = \bar{Y} + \hat{\rho}_{IRMD}(Y_{60} - \bar{Y})$$
$$= 6.128333 + 0.727667(6.2 - 6.128333)$$
$$= 6.180483$$

สรุปผลการวิจัย

ผลการวิจัยครั้งนี้ เมื่อพิจารณาโดยรวมสรุปได้ดังนี้

วิธีการพยากรณ์โดยใช้วิธีกำลังสองน้อยที่สุดแบบค่าเฉลี่ยเวียนเกิด (วิธี RM) ให้ค่า PMSE ต่ำที่สุด เมื่อตัวอย่างมีขนาดใหญ่ (n = 250) หรือตัวอย่างมีขนาดเล็กและปานกลาง (n = 25, 50 และ 100) และค่า ρ มีค่าน้อย (0.1 ถึง 0.5)

วิธีการพยากรณ์โดยใช้วิธีกำลังสองน้อยที่สุดแบบมัยฐานเวียนเกิด (วิธี RMD) ให้ค่า PMSE ต่ำที่สุด เมื่อตัวอย่างมีขนาดเล็ก (n = 25) และค่า ρ มีค่าน้อยและปานกลาง (0.2 ถึง 0.7)

วิธีการพยากรณ์โดยใช้วิธีกำลังสองน้อยที่สุดแบบมัยฐานเวียนเกิดปรับปรุง (วิธี IRMD) ให้ค่า PMSE ต่ำที่สุด เมื่อตัวอย่างมีขนาดเล็ก (n = 25) และค่า ρ มีค่ามาก (0.8 ถึง 0.9)

ข้อเสนอแนะ

จากผลการวิจัยพบว่าวิธีการพยากรณ์ข้อมูลล่วงหน้าหนึ่งคาบเวลาที่ให้ค่า PMSE ต่ำที่สุดจะขึ้นอยู่กับค่าพารามิเตอร์ (ρ) แต่ในทางปฏิบัติเราจะไม่ทราบค่าพารามิเตอร์ ก่อนทำการประมาณค่าพารามิเตอร์ได้ ดังนั้นเพื่อให้สามารถเลือกวิธีการพยากรณ์ข้อมูลที่เหมาะสมได้ ควรทำการประมาณค่าพารามิเตอร์เบื้องต้น (Preliminary Estimation) ซึ่งจะทำให้ทราบค่าประมาณพารามิเตอร์เบื้องต้น จากนั้นสามารถเลือกใช้วิธีการพยากรณ์ข้อมูลได้อย่างเหมาะสมและมีประสิทธิภาพในสถานการณ์ต่างๆ ต่อไปได้ ซึ่งสูตรในการประมาณค่าพารามิเตอร์เบื้องต้นโดยวิธีกำลังสองน้อยที่สุด มีดังนี้

$$\hat{\rho} = \frac{\sum_{t=2}^{n}(Y_t - \bar{Y})(Y_{t-1} - \bar{Y})}{\sum_{t=2}^{n}(Y_{t-1} - \bar{Y})^2}$$

ตารางที่ 1 ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าพยากรณ์ล่วงหน้าหนึ่งคาบเวลา (PMSE) ของตัวแบบ AR(1) เมื่อร้อยละของค่าสูญหาย (p) เท่ากับร้อยละ 5

n	ρ	PMSE					
		RM	(se)	RMD	(se)	IRMD	(se)
25	0.1	1.0941*	(0.0046)	1.0950	(0.0047)	1.1164	(0.0048)
	0.2	1.0895	(0.0048)	1.0894*	(0.0048)	1.1077	(0.0048)
	0.3	1.0966	(0.0049)	1.0961*	(0.0049)	1.1100	(0.0049)
	0.4	1.0992	(0.0052)	1.0979*	(0.0052)	1.1081	(0.0053)
	0.5	1.0961	(0.0048)	1.0955*	(0.0048)	1.1022	(0.0048)
	0.6	1.1036	(0.0049)	1.1032*	(0.0049)	1.1035	(0.0050)
	0.7	1.1131	(0.0049)	1.1140	(0.0049)	1.1094*	(0.0048)
	0.8	1.1204	(0.0046)	1.1222	(0.0046)	1.1109*	(0.0046)
	0.9	1.1283	(0.0056)	1.1325	(0.0056)	1.1135*	(0.0054)
50	0.1	1.0422*	(0.0046)	1.0441	(0.0047)	1.0623	(0.0048)
	0.2	1.0452*	(0.0046)	1.0467	(0.0047)	1.0622	(0.0047)
	0.3	1.0557*	(0.0047)	1.0573	(0.0048)	1.0708	(0.0048)
	0.4	1.0662*	(0.0054)	1.0666	(0.0054)	1.0785	(0.0053)
	0.5	1.0683*	(0.0045)	1.0685	(0.0045)	1.0784	(0.0045)
	0.6	1.0735	(0.0045)	1.0732*	(0.0045)	1.0804	(0.0044)
	0.7	1.0886	(0.0050)	1.0886*	(0.0050)	1.0923	(0.0050)
	0.8	1.0912	(0.0049)	1.0918	(0.0048)	1.0906*	(0.0049)
	0.9	1.1072	(0.0052)	1.1086	(0.0052)	1.1017*	(0.0052)
100	0.1	1.0312*	(0.0046)	1.0328	(0.0047)	1.0431	(0.0047)
	0.2	1.0233*	(0.0046)	1.0241	(0.0046)	1.0342	(0.0046)
	0.3	1.0273*	(0.0050)	1.0282	(0.0050)	1.0374	(0.0051)
	0.4	1.0309*	(0.0040)	1.0314	(0.0040)	1.0400	(0.0040)
	0.5	1.0331*	(0.0048)	1.0334	(0.0048)	1.0407	(0.0048)
	0.6	1.0491*	(0.0047)	1.0493	(0.0047)	1.0546	(0.0047)
	0.7	1.0442*	(0.0042)	1.0446	(0.0042)	1.0488	(0.0043)
	0.8	1.0565	(0.0046)	1.0565*	(0.0046)	1.0587	(0.0045)
	0.9	1.0661	(0.0052)	1.0664	(0.0052)	1.0653*	(0.0051)
250	0.1	1.0197*	(0.0048)	1.0203	(0.0048)	1.0250	(0.0048)
	0.2	1.0119*	(0.0043)	1.0122	(0.0043)	1.0168	(0.0044)
	0.3	1.0223*	(0.0045)	1.0227	(0.0045)	1.0273	(0.0045)
	0.4	1.0259*	(0.0042)	1.0263	(0.0042)	1.0302	(0.0042)
	0.5	1.0143*	(0.0049)	1.0145	(0.0049)	1.0187	(0.0049)
	0.6	1.0263*	(0.0044	1.0265	(0.0044)	1.0301	(0.0044)
	0.7	1.0513*	(0.0054)	1.0515	(0.0054)	1.0550	(0.0054)
	0.8	1.0380*	(0.0044)	1.0381	(0.0044)	1.0403	(0.0044)
	0.9	1.0460*	(0.0052)	1.0461	(0.0052)	1.0474	(0.0052)

* หมายถึง วิธีการพยากรณ์ข้อมูลล่วงหน้าหนึ่งคาบเวลาที่ให้ค่า PMSE ต่ำที่สุดจาก 3 วิธี สำหรับแต่ละค่า ρ

ตารางที่ 2 ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าพยากรณ์ล่วงหน้าหนึ่งคาบเวลา (PMSE) ของตัวแบบ AR(1) เมื่อร้อยละของค่าสูญหาย (p) เท่ากับร้อยละ 10

n	ρ	PMSE					
		RM	(se)	RMD	(se)	IRMD	(se)
25	0.1	1.1038*	(0.0047)	1.1094	(0.0047)	1.1401	(0.0047)
	0.2	1.1081*	(0.0045)	1.1099	(0.0044)	1.1379	(0.0046)
	0.3	1.1255*	(0.0055)	1.1262	(0.0055)	1.1487	(0.0057)
	0.4	1.1199*	(0.0049)	1.1205	(0.0049)	1.1382	(0.0050)
	0.5	1.1415*	(0.0050)	1.1423	(0.0050)	1.1538	(0.0050)
	0.6	1.1490*	(0.0051)	1.1492	(0.0051)	1.1563	(0.0050)
	0.7	1.1799*	(0.0052)	1.1800	(0.0051)	1.1818	(0.0053)
	0.8	1.1979	(0.0061)	1.1993	(0.0061)	1.1936*	(0.0060)
	0.9	1.2168	(0.0060)	1.2211	(0.0060)	1.2050*	(0.0058)
50	0.1	1.0493*	(0.0044)	1.0523	(0.0045)	1.0732	(0.0046)
	0.2	1.0581*	(0.0049)	1.0605	(0.0049)	1.0794	(0.0050)
	0.3	1.0663*	(0.0045)	1.0682	(0.0046)	1.0856	(0.0046)
	0.4	1.0706*	(0.0048)	1.0716	(0.0048)	1.0867	(0.0049)
	0.5	1.0798*	(0.0046)	1.0802	(0.0046)	1.0932	(0.0047)
	0.6	1.0934*	(0.0044)	1.0936	(0.0045)	1.1041	(0.0044)
	0.7	1.1097*	(0.0052)	1.1104	(0.0052)	1.1181	(0.0053)
	0.8	1.1319*	(0.0054)	1.1326	(0.0054)	1.1336	(0.0054)
	0.9	1.1481	(0.0057)	1.1489	(0.0057)	1.1438*	(0.0057)
100	0.1	1.0336*	(0.0047)	1.0364	(0.0047)	1.0512	(0.0048)
	0.2	1.0295*	(0.0046)	1.0315	(0.0047)	1.0444	(0.0047)
	0.3	1.0390*	(0.0042)	1.0405	(0.0042)	1.0521	(0.0042)
	0.4	1.0452*	(0.0042)	1.0464	(0.0042)	1.0577	(0.0042)
	0.5	1.0608*	(0.0045)	1.0616	(0.0045)	1.0714	(0.0046)
	0.6	1.0733*	(0.0048)	1.0739	(0.0048)	1.0823	(0.0049)
	0.7	1.0816*	(0.0051)	1.0818	(0.0051)	1.0883	(0.0051)
	0.8	1.0961*	(0.0056)	1.0966	(0.0056)	1.1006	(0.0056)
	0.9	1.1122	(0.0047)	1.1124	(0.0047)	1.1119*	(0.0047)
250	0.1	1.0195*	(0.0043)	1.0210	(0.0043)	1.0289	(0.0043)
	0.2	1.0258*	(0.0044)	1.0268	(0.0044)	1.0339	(0.0044)
	0.3	1.0271*	(0.0045)	1.0279	(0.0044)	1.0339	(0.0045)
	0.4	1.0307*	(0.0046)	1.0315	(0.0046)	1.0375	(0.0046)
	0.5	1.0493*	(0.0050)	1.0498	(0.0050)	1.0558	(0.0051)
	0.6	1.0566*	(0.0051)	1.0569	(0.0051)	1.0618	(0.0051)
	0.7	1.0642*	(0.0042)	1.0645	(0.0042)	1.0691	(0.0042)
	0.8	1.0877*	(0.0051)	1.0878	(0.0051)	1.0918	(0.0052)
	0.9	1.1050*	(0.0061)	1.1051	(0.0061)	1.1065	(0.0061)

* หมายถึง วิธีการพยากรณ์ข้อมูลล่วงหน้าหนึ่งคาบเวลาที่ให้ค่า PMSE ต่ำที่สุดจาก 3 วิธี สำหรับแต่ละค่า ρ

รายการอ้างอิง

วราฤทธิ์ พานิชกิจโกศลกุล. 2545. "การเปรียบเทียบการประมาณค่าพารามิเตอร์ของตัวแบบอนุกรมเวลา." งานนิพนธ์ปริญญาสถิติศาสตรมหาบัณฑิต สาขาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

วราฤทธิ์ พานิชกิจโกศลกุล. 2552. "การพัฒนาวิธีประมาณค่าพารามิเตอร์ของตัวแบบอัตตสหสัมพันธ์อันดับที่หนึ่งเมื่อข้อมูลมีค่าสูญหาย." **วารสารวิทยาศาสตร์และเทคโนโลยี.** 17, 4: 29-38.

Niwitpong, S. 2004. "Improved GMM Estimator of AR(1) Process Near Unit Root." **Journal of Applied Science-King Mongkut's Institute of Technology North Bangkok.** 3, 1: 21-28.

Pankratz, A. 1983. **Forecasting with Univariate Box-Jenkins Models.** New York: John Wiley & Sons.

Shin, D.W. and Sarkar, S. 1996. "Testing for a Unit Root in an AR(1) Time Series Using Irregularly Observed Data." **Journal of Time Series Analysis.** 17, 3: 309-321.

So, B.S. and Shin, D.W. 1999. "Recursive Mean Adjustment in Time Series Inferences." **Statistics and Probability Letters.** 43, 1: 65-73.

Vougas, D.V. 2000. "A Comparison of LS/ML and GMM Estimation in a Sample AR(1) Model." **Communications in Statistics: Simulation and Computation.** 29, 1: 239-258.

Wei, W.W.S. 2006. **Time Series Analysis: Univariate and Multivariate Methods.** Boston: Addison-Wesley.

>> วราฤทธิ์ พานิชกิจโกศลกุล

จบการศึกษาหลักสูตรสถิติศาสตรมหาบัณฑิต สาขาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย หลักสูตรวิทยาศาสตรบัณฑิต สาขาสถิติประยุกต์ คณะวิทยาศาสตร์ประยุกต์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ หลักสูตรบริหารธุรกิจบัณฑิต สาขาการตลาด มหาวิทยาลัยสุโขทัยธรรมาธิราช หลักสูตรเทคโนโลยีบัณฑิต สาขาเทคโนโลยีสารสนเทศธุรกิจ มหาวิทยาลัยสุโขทัยธรรมาธิราช และหลักสูตรเศรษฐศาสตรบัณฑิต สาขาเศรษฐศาสตร์ธุรกิจ มหาวิทยาลัยสุโขทัยธรรมาธิราช

ปัจจุบันทำงานในตำแหน่งผู้ช่วยศาสตราจารย์ ภาควิชาคณิตศาสตร์และสถิติคณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

ผลงานทางวิชาการ เช่น New Estimator for an Unknown Mean Gaussian AR (1) Process with Additive Outliers, Asymptotic Confidence Intervals for the Coefficient of Variation of a Poisson Distribution: A Simulation Study เป็นต้น