บทความวิจัย

การเปรียบเทียบช่วงความเชื่อมั่นของพารามิเตอร์ สำหรับตัวแบบอัตตสหสัมพันธ์อันดับที่หนึ่ง เมื่อข้อมูลมีค่าผิดปกติและค่าสูญหาย

วราฤทธิ์ พานิชกิจโกศลกุล*

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของช่วงความเชื่อมั่นของพารามิเตอร์ สำหรับตัวแบบอัตตสหสัมพันธ์อันดับที่หนึ่ง (AR(1)) เมื่อข้อมูลมีค่าผิดปกติและค่าสูญหาย ช่วงความเชื่อมั่นที่พิจารณามี 3 วิธี คือ ช่วงความเชื่อมั่นโดยใช้วิธีกำลังสองน้อยที่สุดแบบค่าเฉลี่ยเวียนเกิด (RM) วิธีกำลังสองน้อยที่สุดแบบมัธยฐานเวียนเกิด (RMD) และวิธีกำลังสองน้อยที่สุดแบบมัธยฐานเวียนเกิดปรับปรุง (IRMD) ร้อยละของค่าผิดปกติ เท่ากับ 10 ขนาดของค่าผิดปกติ เท่ากับ $3\sigma_a$ และ $5\sigma_a$ ร้อยละของค่าสูญหาย เท่ากับ 10 กำหนดขนาดตัวอย่าง 4 ระดับคือ 25, 50, 100 และ 250 และคำนวณช่วงความเชื่อมั่น ที่ระดับความเชื่อมั่น 95% ในการวิจัยครั้งนี้ใช้วิธีการจำลองแบบมอนติคาร์โล และทำการทดลองซ้ำๆ กัน 10,000 ครั้งในแต่ละสถานการณ์ สำหรับเกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพของช่วงความเชื่อมั่นพิจารณาจากค่าความน่าจะเป็นครอบคลุม (Coverage Probability) และค่าเฉลี่ยความกว้างช่วง (Expected Length) ผลการวิจัยสรุปได้ดังนี้ ช่วงความเชื่อมั่นโดยใช้วิธี IRMD ให้ค่าความน่าจะเป็นครอบคลุมมากกว่าช่วงความเชื่อมั่นโดยใช้วิธี RM และวิธี RMD ในเกือบทุกกรณีที่ศึกษา ยกเว้นกรณีที่ค่าพารามิเตอร์ ρ เท่ากับ 0.1 นอกจากนี้ ค่าความน่าจะเป็นครอบคลุมของช่วงความเชื่อมั่นทั้ง 3 วิธี มีค่าต่ำกว่า 0.95 ในเกือบทุกกรณีที่ศึกษา

คำสำคัญ: ช่วงความเชื่อมั่น ตัวแบบอัตตสหสัมพันธ์อันดับที่หนึ่ง ค่าผิดปกติ ค่าสูญหาย

ภาควิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

*ผู้นิพนธ์ประสานงาน, e-mail: wararit@mathstat.sci.tu.ac.th

A Comparison of Confidence Intervals for Parameter in First-Order Autoregressive Model with Outliers and Missing Values

Wararit Panichkitkosolkul*

ABSTRACT

The objective of this research is to compare the efficiency of confidence intervals for parameter in first-order autoregressive model (AR(1)) with outliers and missing values. Three confidence intervals are estimated by using the recursive mean OLS method (RM), the recursive median OLS method (RMD) and the improved recursive median OLS method (IRMD). The percentage of outliers is 10% and the magnitude of outliers is equal to $3\sigma_a$ and $5\sigma_a$. The percentage of missing values is 10%. The sample sizes are 25, 50, 100, 250 and the 95% confidence intervals are constructed in this paper. This research uses the Monte Carlo simulation method. The experiment is repeated 10,000 times for each condition. The criteria for evaluating the efficiencies of confidence intervals are the coverage probability and expected length of the prediction intervals. The results of the research are as follows: The confidence interval for parameter ρ by IRMD method provides more coverage probability than that by RM method and RMD method in almost all cases except the parameter ρ is equal to 0.1. Moreover, the coverage probabilities of all confidence intervals are less than 0.95 in almost all cases.

Keywords: confidence interval, AR(1) model, outliers, missing values

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

*Corresponding author, email: wararit@mathstat.sci.tu.ac.th

บทนำ

การพยากรณ์อนุกรมเวลา (time series forecasting) เป็นวิธีการพยากรณ์วิธีหนึ่งที่นิยมใช้กันมาก วิธีนี้จะใช้ข้อมูลในอดีต โดยจะศึกษาถึงลักษณะการเปลี่ยนแปลงของข้อมูลเมื่อเวลาเปลี่ยนไปว่ามีลักษณะเช่นไร และทำการกำหนดรูปแบบของการแปรเปลี่ยนที่เหมาะสมที่สุดสำหรับข้อมูลนั้น โดยจะอยู่ในรูปของความสัมพันธ์กับเวลา การพยากรณ์อนุกรมเวลามีอยู่ด้วยกันหลายวิธี เช่น เทคนิคการทำให้เรียบ (smoothing techniques) การกรองแบบปรับได้ (adaptive filtering) วิธีอนุกรมเวลาแบบคลาสสิค (classical time series method) และวิธีอนุกรมเวลาแบบบ็อกซ์-เจนกินส์ (Box-Jenkins method) เป็นต้น

นอกเหนือจากการเลือกใช้วิธีการพยากรณ์ที่เหมาะสม และการกำหนดตัวแบบอนุกรมเวลาแล้ว วิธีการประมาณค่าพารามิเตอร์ของตัวแบบอนุกรมเวลาก็เป็นอีกปัจจัยหนึ่งที่ส่งผลถึงความแม่นยำและเชื่อถือได้ของค่าพยากรณ์ [1] การประมาณค่าพารามิเตอร์แบบจุด (point estimation) ของตัวแบบอนุกรมเวลามีอยู่หลายวิธี เช่น วิธีกำลังสองน้อยที่สุด (Ordinary Least Squares Method: OLS) วิธีการประมาณความควรจะเป็นสูงสุด (Maximum Likelihood Estimation Method: MLE) เป็นต้น โดยแต่ละวิธีการจะมีหลักการในการหาตัวประมาณค่าแตกต่างกันไป อีกทั้งยังให้ผลลัพธ์ที่แตกต่างกันออกไปด้วย ซึ่งวิธีการประมาณค่าพารามิเตอร์แบบจุดนั้นได้มีผู้ศึกษาจากอดีตถึงปัจจุบัน ภายใต้ข้อสมมติเบื้องต้นที่แตกต่างกันออกไป อาทิ โซ และชิน [2] ได้พัฒนาวิธีกำลังสองน้อยที่สุดแบบค่าเฉลี่ยเวียนเกิด (Recursive Mean OLS Method) สำหรับตัวแบบอัตตสหสัมพันธ์อันดับที่หนึ่ง (First-Order Autoregressive Model: AR(1)) วาวกัส [3] ได้พัฒนาวิธีการประมาณจีเอ็มเอ็ม (Generalized Method of Moment Method: GMM) สำหรับตัวแบบ AR(1) สอาด [4] ได้ปรับปรุงวิธีการประมาณจีเอ็มเอ็มของวาวกัส โดยใช้แนวคิดของโซ และชิน [2] วราฤทธิ์ และคณะ [5] ได้พัฒนาช่วงความเชื่อมั่นของพารามิเตอร์ของตัวแบบ AR(1) เมื่อข้อมูลมีค่าผิดปกติ นอกจากนี้ วราฤทธิ์ [6] เสนอวิธีกำลังสองน้อยที่สุดแบบมัธยฐานเวียนเกิดปรับปรุงสำหรับประมาณค่าพารามิเตอร์ของตัวแบบ AR(1) เมื่อข้อมูลมีค่าสูญหาย นอกจากการประมาณค่าแบบจุดแล้วยังมีการประมาณช่วงความเชื่อมั่น (confidence interval) อีกด้วย ซึ่งในการประมาณช่วงความเชื่อมั่นจะอาศัยข้อมูลในอดีตแต่ผู้วิเคราะห์อาจจะประสบปัญหาเกี่ยวกับข้อมูล เช่น ข้อมูลบางช่วงมีค่าสูงกว่าหรือต่ำกว่าปกติ เป็นต้น ซึ่งลักษณะข้อมูลที่มีค่าผิดปกตินี้อาจได้รับผลกระทบจากเหตุการณ์ภายนอก เช่น การเปลี่ยนแปลงนโยบายทางเศรษฐกิจ การลดค่าของเงิน การเกิดภาวะสงคราม การนัดหยุดงาน หรือการเกิดภัยธรรมชาติอย่างรุนแรง เป็นต้น เรียกข้อมูลที่ได้รับผลกระทบต่างๆ นี้ว่า ค่าผิดปกติ (outliers) หรือข้อมูลบางค่าสูญหาย (missing values) ทำให้ผลการวิเคราะห์และการพยากรณ์คลาดเคลื่อน หรืออาจนำไปสู่การตัดสินใจหรือวางแผนที่ผิดพลาดได้ ดังนั้น วัตถุประสงค์ของการวิจัยครั้งนี้ คือ การเปรียบเทียบช่วงความเชื่อมั่นของพารามิเตอร์สำหรับตัวแบบ AR(1) เมื่อข้อมูลมีค่าผิดปกติและค่าสูญหาย โดยประยุกต์วิธีการประมาณค่าช่วงความเชื่อมั่นซึ่งเสนอโดยวราฤทธิ์ และคณะ [5] และวราฤทธิ์ [6] และเปรียบเทียบกับวิธีกำลังสองน้อยที่สุดแบบค่าเฉลี่ยเวียนเกิดซึ่งเสนอโดย โซ และชิน [2]

ขอบเขตของการวิจัย

ขอบเขตของการวิจัย มีดังนี้

1. อนุกรมเวลา $\{Y_t\}$ เป็นอนุกรมเวลาตัวแบบ AR(1) เขียนตัวแบบได้ดังนี้ [7]

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

ในที่นี้กำหนดค่าเฉลี่ยของอนุกรมเวลา (μ) เท่ากับ 0

2. กำหนดค่าพารามิเตอร์ (ρ) ของตัวแบบ 9 ระดับ คือ 0.1, 0.3, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95 และ 0.99

3. การแจกแจงของความคลาดเคลื่อนสุ่ม a_t มีการแจกแจงแบบปกติ (normal distribution) โดยมีฟังก์ชันความหนาแน่นของความน่าจะเป็น

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

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

4. ลักษณะของค่าผิดปกติที่ศึกษา คือ Additive Outliers (AO) ซึ่งเป็นค่าผิดปกติที่ส่งผลกระทบต่อค่าสังเกต ณ เวลา t = T เท่านั้น ตัวแบบของค่าผิดปกติแบบ AO เขียนตัวแบบได้ดังนี้

$$Y_t = \begin{cases} Z_t, & t \neq T \\ Z_t + \delta, & t = T \end{cases}$$

เมื่อ Y_t คือ อนุกรมเวลาตัวแบบ AR(1) เมื่อเกิดค่าผิดปกติแบบ AO

Z_t คือ อนุกรมเวลาตัวแบบ AR(1)

δ คือ ขนาดของค่าผิดปกติ

T คือ เวลาที่เกิดค่าผิดปกติ

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

6. ร้อยละของค่าผิดปกติ (p_o) เท่ากับ 10 ร้อยละของค่าสูญหาย (p_m) เท่ากับ 10 และตำแหน่งของค่าผิดปกติและค่าสูญหายเป็นไปอย่างสุ่ม ซึ่งอยู่ระหว่างเวลาที่ t = 2 ถึง $t = n-1$

7. ขนาดของค่าผิดปกติ (δ) เท่ากับ $3\sigma_a$ และ $5\sigma_a$

8. ระดับความเชื่อมั่น $(1-\alpha)100\%$ เท่ากับ 95%

9. เกณฑ์ที่ใช้เปรียบเทียบประสิทธิภาพของช่วงความเชื่อมั่นพิจารณาจากค่าความน่าจะเป็นครอบคลุม และค่าเฉลี่ยความกว้างช่วง

10. โปรแกรมที่ใช้ในการวิจัยครั้งนี้ทั้งหมดเขียนด้วยโปรแกรม R เวอร์ชัน 2.8.1 ซึ่งทำการทดลองซ้ำ 10,000 ครั้งในแต่ละสถานการณ์ของการทดลอง

วิธีดำเนินการวิจัย

ในการวิจัยครั้งนี้มีวิธีดำเนินการวิจัยดังนี้

1. จำลองค่าความคลาดเคลื่อนสุ่ม a_t

การสร้างการแจกแจงของความคลาดเคลื่อนสุ่มที่ต้องการศึกษา มีการแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1

2 จำลองข้อมูลอนุกรมเวลาตัวแบบ AR(1) ที่มีค่าผิดปกติและค่าสูญหาย

สร้าง Z_0 ให้มีการแจกแจงปกติ ซึ่งมีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ $\frac{\sigma_a^2}{1-\rho^2} = \frac{1}{1-\rho^2}$ และสร้าง a_t ; $t = 1,2,...,n$ ให้มีการแจกแจงปกติ ซึ่งมีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 จากนั้นสร้าง Z_t ; $t = 1,2,...,n$ ให้มีรูปแบบความสัมพันธ์คือ

$$Z_t = \rho Z_{t-1} + a_t$$

เมื่อสร้างข้อมูลอนุกรมเวลาตัวแบบ AR(1) แล้ว ต่อไปจะทำการสร้างข้อมูลอนุกรมเวลาตัวแบบ AR(1) ที่มีค่าผิดปกติ ดังนี้

$$Y_t = \begin{cases} Z_t, & t \neq T \\ Z_t + \delta, & t = T \end{cases}$$

ต่อไปจะทำการสร้างข้อมูลอนุกรมเวลาตัวแบบ AR(1) ที่มีค่าสูญหาย โดยสุ่มเลือกตำแหน่งของค่าสูญหายระหว่างเวลาที่ $t = 2$ ถึง $t = n$-1 (ตำแหน่งของค่าสูญหายไม่ซ้ำกับตำแหน่งของค่าผิดปกติ) จากนั้นบันทึกค่า Y_t ในคาบเวลานั้นให้เป็นค่าสูญหาย

3. ประมาณค่าสูญหาย

สมมติว่า Y_K เป็นค่าสูญหาย ประมาณค่า Y_K ด้วย $E(Y_K | Y_1, Y_2, ..., Y_{K-1}) = Y_{K-1}$ [8] นั่นคือ ค่าประมาณของ Y_K คือ Y_{K-1}

4. คำนวณช่วงความเชื่อมั่นของพารามิเตอร์ของตัวแบบ AR(1) ทั้ง 3 วิธี

1) วิธีกำลังสองน้อยที่สุดแบบค่าเฉลี่ยเวียนเกิด โซ และชิน [2] ได้พัฒนาวิธีกำลังสองน้อยที่สุดแบบค่าเฉลี่ยเวียนเกิด โดยมีหลักการเช่นเดียวกับวิธีกำลังสองน้อยที่สุด คือการทำให้ผลรวมค่ากำลังสองของความคลาดเคลื่อนภายใต้พารามิเตอร์ของตัวแบบมีค่าต่ำที่สุด แต่วิธีกำลังสองน้อยที่สุดแบบค่าเฉลี่ยเวียนเกิดจะประมาณค่าเฉลี่ยของอนุกรมเวลา (μ) ด้วยค่าเฉลี่ยเวียนเกิด (recursive mean: $\bar{Y}_t$) แทนค่าเฉลี่ยเลขคณิต (arithmetic mean: $\bar{Y}$) ช่วงความเชื่อมั่นโดยใช้วิธีกำลังสองน้อยที่สุดแบบค่าเฉลี่ยเวียนเกิด คือ

$$CI_{RM} = [\ \hat{\rho}_{RM} - Z_{\alpha/2} \cdot SE(\hat{\rho}_{RM})\ ,\ \hat{\rho}_{RM} + Z_{\alpha/2} \cdot SE(\hat{\rho}_{RM})\ ]$$

โดยที่

$$\hat{\rho}_{RM} = \frac{\sum_{t=2}^{n}(Y_t - \overline{Y}_{t-1})(Y_{t-1} - \overline{Y}_{t-1})}{\sum_{t=2}^{n}(Y_{t-1} - \overline{Y}_{t-1})^2}, \qquad \overline{Y}_t = \frac{\sum_{i=1}^{t} Y_i}{t}$$

$$SE(\hat{\rho}_{RM}) = \frac{\hat{\sigma}_1}{\sqrt{\sum_{t=2}^{n}(Y_{t-1} - \overline{Y}_{t-1})^2}} \quad \text{และ} \quad \hat{\sigma}_1 = \sqrt{\frac{\sum_{t=2}^{n}(Y_t - \overline{Y}_t - \hat{\rho}_{RM}(Y_{t-1} - \overline{Y}_{t-1}))^2}{n-2}}$$

2) วิธีกำลังสองน้อยที่สุดแบบมัธยฐานเวียนเกิด วราฤทธิ์ และคณะ [5] ได้พัฒนาวิธีกำลังสองน้อยที่สุดแบบมัธยฐานเวียนเกิด โดยใช้แนวคิดของโซ และชิน [2] กล่าวคือ วิธีนี้มีหลักการเช่นเดียวกับวิธีกำลังสองน้อยที่สุด และประมาณค่าเฉลี่ยของอนุกรมเวลา (μ) ด้วยค่ามัธยฐานเวียนเกิด (recursive median: $\overline{Y}_t$) แทนค่าเฉลี่ยเลขคณิต ($\overline{Y}$) ช่วงความเชื่อมั่นโดยใช้วิธีกำลังสองน้อยที่สุดแบบมัธยฐานเวียนเกิด คือ

$$CI_{RMD} = [\ \hat{\rho}_{RMD} - Z_{\alpha/2} \cdot SE(\hat{\rho}_{RMD}),\ \hat{\rho}_{RMD} + Z_{\alpha/2} \cdot SE(\hat{\rho}_{RMD})\ ]$$

โดยที่

$$\hat{\rho}_{RMD} = \frac{\sum_{t=2}^{n}(Y_t - \tilde{Y}_t)(Y_{t-1} - \tilde{Y}_{t-1})}{\sum_{t=2}^{n}(Y_{t-1} - \tilde{Y}_{t-1})^2}, \quad \tilde{Y}_t = median\{Y_1, Y_2, ..., Y_t\}, \quad \overline{Y}_t = \frac{\sum_{i=1}^{t} Y_i}{t}$$

$$SE(\hat{\rho}_{RMD}) = \frac{\hat{\sigma}_2}{\sqrt{\sum_{t=2}^{n}(Y_{t-1} - \overline{Y}_{t-1})^2}} \quad \text{และ} \quad \hat{\sigma}_2 = \sqrt{\frac{\sum_{t=2}^{n}(Y_t - \overline{Y}_t - \hat{\rho}_{RMD}(Y_{t-1} - \overline{Y}_{t-1}))^2}{n-2}}$$

3) วิธีกำลังสองน้อยที่สุดแบบมัธยฐานเวียนเกิดปรับปรุง วราฤทธิ์ [6] ได้ปรับปรุงประสิทธิภาพของวิธีกำลังสองน้อยที่สุดแบบมัธยฐานเวียนเกิด โดยนำค่ามัธยฐานเวียนเกิด $\tilde{Y}_t$ มาหาค่าเฉลี่ยเวียนเกิดอีกครั้งหนึ่ง เพื่อลดอิทธิพลของข้อมูลผิดปกติ ช่วงความเชื่อมั่นโดยใช้วิธีกำลังสองน้อยที่สุดแบบมัธยฐานเวียนเกิดปรับปรุง คือ

$$CI_{IRMD} = [\ \hat{\rho}_{IRMD} - Z_{\alpha/2} \cdot SE(\hat{\rho}_{IRMD}),\ \hat{\rho}_{IRMD} + Z_{\alpha/2} \cdot SE(\hat{\rho}_{IRMD})\ ]$$

โดยที่

$$\hat{\rho}_{IRMD} = \frac{\sum_{t=2}^{n}(Y_t - \ddot{Y}_t)(Y_{t-1} - \ddot{Y}_{t-1})}{\sum_{t=2}^{n}(Y_{t-1} - \ddot{Y}_{t-1})^2}, \quad \ddot{Y}_t = \frac{\sum_{i=1}^{t} \tilde{Y}_i}{t}, \quad \tilde{Y}_t = median\{Y_1, Y_2, ..., Y_t\}, \quad \overline{Y}_t = \frac{\sum_{i=1}^{t} Y_i}{t}$$

$$SE(\hat{\rho}_{IRMD}) = \frac{\hat{\sigma}_3}{\sqrt{\sum_{t=2}^{n}(Y_{t-1} - \overline{Y}_{t-1})^2}} \quad \text{และ} \quad \hat{\sigma}_3 = \sqrt{\frac{\sum_{t=2}^{n}(Y_t - \overline{Y}_t - \hat{\rho}_{IRMD}(Y_{t-1} - \overline{Y}_{t-1}))^2}{n-2}}$$

5. คำนวณค่าความน่าจะเป็นครอบคลุม (Coverage Probability: CP) และค่าเฉลี่ยความกว้างช่วง (Expected Length: EL) ซึ่งคำนวณจาก

$$CP = \frac{\text{จำนวนครั้งทั้งหมดที่ช่วงความเชื่อมั่นครอบคลุมค่าพารามิเตอร์}}{M}$$

$$EL = 2x\left|Z_{a/2}\right| x\ SE(\hat{\rho}_k)$$

เมื่อ $SE(\hat{\rho}_k)$ แทน ค่าความคลาดเคลื่อนมาตรฐาน (standard error) ของตัวประมาณ ซึ่งคำนวณจากวิธีที่ k แทนที่วิธี RM, RMD และ T IRMD

M แทน จำนวนครั้งของการทำซ้ำ ในที่นี้เท่ากับ 10,000 ครั้ง

ผลการวิจัย

ในการนำเสนอผลการวิจัยเพื่อความสะดวกจะใช้สัญลักษณ์ต่างๆ แทนความหมาย ดังนี้

RM หมายถึง ช่วงความเชื่อมั่นโดยใช้วิธีกำลังสองน้อยที่สุดแบบค่าเฉลี่ยเวียนเกิด

RMD หมายถึง ช่วงความเชื่อมั่นโดยใช้วิธีกำลังสองน้อยที่สุดแบบมัธยฐานเวียนเกิด

IRMD หมายถึง ช่วงความเชื่อมั่นโดยใช้วิธีกำลังสองน้อยที่สุดแบบมัธยฐานเวียนเกิดปรับปรุง

ผลการเปรียบเทียบประสิทธิภาพของช่วงความเชื่อมั่นทั้ง 3 วิธี นำเสนอดังตารางที่ 1-2 ในที่นี้จะพิจารณาค่าความน่าจะเป็นครอบคลุม และค่าเฉลี่ยความกว้างช่วง ดังนี้

1. ค่าความน่าจะเป็นครอบคลุม

สำหรับทุกระดับขนาดตัวอย่าง (n = 25, 50, 100 และ 250) และทุกระดับขนาดของค่าผิดปกติ ($\delta = 3\sigma_a$ และ $5\sigma_a$) วิธี IRMD ให้ค่าความน่าจะเป็นครอบคลุมมากที่สุดในเกือบทุกระดับของค่าพารามิเตอร์ ρ ยกเว้นกรณีที่ค่า ρ เท่ากับ 0.1 รองลงมาคือ วิธี RMD และวิธี RM ตามลำดับ นอกจากนี้ วิธี RM และวิธี RMD ยังให้ค่าความน่าจะเป็นครอบคลุมใกล้เคียงกัน

เมื่อพิจารณาค่าความน่าจะเป็นในด้านขนาดตัวอย่าง พบว่า เมื่อระดับของขนาดตัวอย่างเพิ่มขึ้น จะมีผลทำให้ค่าความน่าจะเป็นครอบคลุมลดลงในเกือบทุกระดับของค่าพารามิเตอร์ ρ และทุกระดับของขนาดของค่าผิดปกติ

เมื่อพิจารณาค่าความน่าจะเป็นในด้านขนาดของค่าผิดปกติ พบว่า เมื่อระดับของขนาดของค่าผิดปกติเพิ่มขึ้น จะมีผลทำให้ค่าความน่าจะเป็นครอบคลุมเพิ่มขึ้นในเกือบทุกระดับของค่าพารามิเตอร์ ρ และทุกระดับของขนาดตัวอย่าง

เมื่อพิจารณาค่าความน่าจะเป็นในด้านค่าพารามิเตอร์ ρ พบว่า เมื่อระดับของค่าพารามิเตอร์ ρ เพิ่มขึ้น จะมีผลทำให้ค่าความน่าจะเป็นครอบคลุมลดลงในเกือบทุกระดับของขนาดของค่าผิดปกติ และทุกระดับของขนาดตัวอย่าง

2. ค่าเฉลี่ยความกว้างช่วง

ช่วงความเชื่อมั่นที่คำนวณจากวิธี RM, RMD และ IRMD ให้ค่าความน่าจะเป็นครอบคลุมต่ำกว่า 0.95 ในเกือบทุกกรณีที่ศึกษา วิธี IRMD ให้ค่าเฉลี่ยความกว้างช่วงมากที่สุด ในทุกๆ ระดับขนาดตัวอย่าง (n = 25, 50, 100 และ 250) ทุกระดับขนาดของค่าผิดปกติ ($\delta = 3\sigma_a$ และ $5\sigma_a$) และทุกระดับของค่าพารามิเตอร์ ρ (ρ = 0.1, 0.3, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95 และ 0.99) เนื่องจากวิธี IRMD ให้ค่าความน่าจะเป็นครอบคลุมมากที่สุด

ตารางที่ 1 ค่าความน่าจะเป็นครอบคลุม (Coverage Probability) และค่าเฉลี่ยความกว้างช่วง (Expected Length) ของช่วงความเชื่อมั่นของพารามิเตอร์ของตัวแบบ AR(1) เมื่อ $\delta = 3\sigma_a$

		Coverage Probability			Expected Length		
n	ρ	**RM**	**RMD**	**IRMD**	**RM**	**RMD**	**IRMD**
25	0.1	0.9464	0.9426	0.9170	0.8090	0.8060	0.8185
	0.3	0.9249	0.9317	0.9479	0.7976	0.793	0.8137
	0.5	0.8470	0.8635	0.9200	0.7748	0.7715	0.8054
	0.6	0.8000	0.8119	0.8963	0.7582	0.7543	0.8003
	0.7	0.7391	0.7505	0.8708	0.7342	0.7304	0.7903
	0.8	0.6961	0.7041	0.8529	0.6997	0.6959	0.7794
	0.9	0.6331	0.6390	0.8336	0.6536	0.6508	0.7631
	0.95	0.6074	0.6078	0.8379	0.6229	0.6208	0.7510
	0.99	0.5861	0.5821	0.8315	0.5964	0.5958	0.7418
50	0.1	0.9294	0.9138	0.857	0.5585	0.5560	0.5642
	0.3	0.9037	0.9189	0.9320	0.5499	0.5470	0.5616
	0.5	0.7615	0.7866	0.8769	0.5330	0.5298	0.5568
	0.6	0.6676	0.6975	0.8395	0.5195	0.5161	0.5525
	0.7	0.5874	0.6097	0.8061	0.4980	0.4950	0.5459
	0.8	0.5183	0.5426	0.7942	0.4665	0.4631	0.5355
	0.9	0.4992	0.5131	0.8216	0.415	0.4119	0.5174
	0.95	0.5001	0.5067	0.8411	0.3776	0.3748	0.5042
	0.99	0.4967	0.4998	0.8520	0.3418	0.3395	0.4883
100	0.1	0.9133	0.8847	0.8035	0.3912	0.3898	0.3948
	0.3	0.8834	0.9046	0.9218	0.3846	0.3829	0.3934
	0.5	0.6219	0.6631	0.8089	0.3716	0.3697	0.3906
	0.6	0.4764	0.5166	0.7312	0.3606	0.3586	0.3880
	0.7	0.3564	0.3894	0.6868	0.3440	0.3420	0.3836
	0.8	0.3176	0.3424	0.6932	0.3171	0.3151	0.3768
	0.9	0.3545	0.3745	0.7979	0.2697	0.2675	0.3625
	0.95	0.4121	0.4263	0.8648	0.2313	0.2290	0.3500
	0.99	0.4614	0.4737	0.9092	0.1896	0.1876	0.3327
250	0.1	0.8814	0.8181	0.7205	0.2461	0.2454	0.2485
	0.3	0.8031	0.8460	0.8867	0.2419	0.2410	0.2479
	0.5	0.2855	0.3442	0.5565	0.2331	0.2321	0.2468
	0.6	0.1242	0.1560	0.3936	0.2256	0.2247	0.2457
	0.7	0.0631	0.0790	0.3356	0.2140	0.2131	0.2438
	0.8	0.0471	0.0600	0.3805	0.1953	0.1945	0.2403
	0.9	0.0988	0.1168	0.6539	0.1605	0.1595	0.2326
	0.95	0.2177	0.2403	0.8511	0.1290	0.1279	0.2240
	0.99	0.4082	0.4303	0.9611	0.0882	0.0868	0.2082

ตารางที่ 2 ค่าความน่าจะเป็นครอบคลุม (Coverage Probability) และค่าเฉลี่ยความกว้างช่วง (Expected Length) ของช่วงความเชื่อมั่นของพารามิเตอร์ของตัวแบบ AR(1) เมื่อ $\delta = 5\sigma_a$

		Coverage Probability			Expected Length		
n	ρ	RM	RMD	IRMD	RM	RMD	IRMD
25	0.1	0.9425	0.9442	0.9392	0.8055	0.8030	0.8176
	0.3	0.8782	0.9025	0.9341	0.8018	0.7984	0.8158
	0.5	0.6829	0.7121	0.8050	0.7929	0.7895	0.8121
	0.6	0.5671	0.5944	0.7265	0.7860	0.7823	0.8085
	0.7	0.4743	0.4948	0.6684	0.7743	0.7706	0.8030
	0.8	0.3905	0.4052	0.6239	0.7560	0.7518	0.7946
	0.9	0.3331	0.3434	0.5997	0.7281	0.7236	0.7826
	0.95	0.3139	0.3169	0.5982	0.7072	0.7031	0.7713
	0.99	0.3070	0.3081	0.6134	0.6867	0.6832	0.7623
50	0.1	0.9277	0.9118	0.8856	0.5582	0.5550	0.5646
	0.3	0.8116	0.8611	0.8938	0.5548	0.5510	0.5634
	0.5	0.4904	0.5571	0.6817	0.5481	0.5441	0.5612
	0.6	0.3235	0.3778	0.5514	0.5421	0.5379	0.5590
	0.7	0.2084	0.2475	0.4620	0.5323	0.5280	0.5548
	0.8	0.1379	0.1658	0.4225	0.5156	0.5110	0.5474
	0.9	0.1419	0.1614	0.4825	0.4802	0.4754	0.5316
	0.95	0.1588	0.1775	0.5335	0.4520	0.4474	0.5200
	0.99	0.1944	0.2033	0.6005	0.4193	0.4150	0.5049
100	0.1	0.9170	0.877	0.8361	0.3916	0.3892	0.3952
	0.3	0.7146	0.8055	0.8477	0.3889	0.3862	0.3946
	0.5	0.2387	0.3262	0.4766	0.3834	0.3804	0.3932
	0.6	0.0935	0.1375	0.3036	0.3785	0.3754	0.3917
	0.7	0.0386	0.0539	0.2079	0.3704	0.3674	0.3894
	0.8	0.0193	0.0276	0.1898	0.3548	0.3520	0.3841
	0.9	0.0283	0.0383	0.3208	0.3214	0.3186	0.3720
	0.95	0.0715	0.0866	0.4936	0.2876	0.2845	0.3594
	0.99	0.1470	0.1605	0.6591	0.2452	0.2423	0.340
250	0.1	0.9106	0.8068	0.7514	0.2467	0.2453	0.2486
	0.3	0.4717	0.6564	0.7221	0.2448	0.2431	0.2484
	0.5	0.0178	0.0484	0.1271	0.2410	0.2392	0.2478
	0.6	0.0015	0.0050	0.0354	0.2376	0.2358	0.2473
	0.7	0.0001	0.0003	0.0104	0.2318	0.2300	0.2462
	0.8	0.0001	0.0002	0.0089	0.2209	0.2194	0.2437
	0.9	0.0000	0.0003	0.0573	0.1952	0.1938	0.2371
	0.95	0.0037	0.0066	0.2797	0.1663	0.1648	0.2282
	0.99	0.0928	0.1146	0.7391	0.1192	0.1174	0.2111

สรุปผลการวิจัย

จากการวิจัยได้ข้อสรุปดังนี้

ช่วงความเชื่อมั่นของพารามิเตอร์ (ρ) สำหรับตัวแบบอัตตสหสัมพันธ์อันดับที่หนึ่ง (AR(1)) เมื่อข้อมูลมีค่าผิดปกติและค่าสูญหายโดยใช้วิธีกำลังสองน้อยที่สุดแบบมัธยฐานเวียนเกิดปรับปรุง (IRMD) ให้ค่าความน่าจะเป็นครอบคลุมมากกว่าช่วงความเชื่อมั่นโดยใช้วิธีกำลังสองน้อยที่สุดแบบค่าเฉลี่ยเวียนเกิด (RM) และวิธีกำลังสองน้อยที่สุดแบบมัธยฐานเวียนเกิด (RMD) ในเกือบทุกกรณีที่ศึกษา ยกเว้นกรณีที่ค่าพารามิเตอร์ ρ เท่ากับ 0.1 นอกจากนี้ ค่าความน่าจะเป็นครอบคลุมของช่วงความเชื่อมั่นทั้ง 3 วิธีมีค่าต่ำกว่า 0.95 ในเกือบทุกกรณีที่ศึกษา

ข้อเสนอแนะ

ในการวิจัยครั้งนี้ศึกษาเฉพาะกรณีที่ค่าผิดปกติเป็นแบบ Additive Outliers (AO) เพื่อให้การศึกษาเกิดประโยชน์มากยิ่งขึ้นอาจจะศึกษากรณีที่ค่าผิดปกติเป็นแบบ Innovation Outliers (IO)

เอกสารอ้างอิง

1. วราฤทธิ์ พานิชกิจโกศลกุล. 2545. การเปรียบเทียบการประมาณค่าพารามิเตอร์ของตัวแบบอนุกรมเวลา. วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ. กรุงเทพฯ. บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
2. So, B. S., and Shin, D. W. 1999. Recursive Mean Adjustment in Time Series Inferences. *Statistics and Probability Letters* 43: 65-73.
3. 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.
4. 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.
5. วราฤทธิ์ พานิชกิจโกศลกุล ลักขณา เศาธยะนันท์ สุณี ทวีสกุลวัชระ และยุพิน กาญจนะศักดิ์ดา. 2551. ช่วงความเชื่อมั่นของพารามิเตอร์อัตตสัมพันธ์อันดับที่ 1 เมื่อข้อมูลมีค่าผิดปกติ. *วารสารวิทยาศาสตร์ประยุกต์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ* 7(2): 1-8.
6. วราฤทธิ์ พานิชกิจโกศลกุล. 2552. การพัฒนาวิธีประมาณค่าพารามิเตอร์ของตัวแบบอัตตสหสัมพันธ์อันดับที่หนึ่ง เมื่อข้อมูลมีค่าสูญหาย. *วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์* (อยู่ระหว่างตีพิมพ์).
7. Wei, W. W. S. 2006. Time Series Analysis: Univariate and Multivariate Methods. Boston. Addison-Wesley.
8. 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: 309-321.

ได้รับบทความวันที่ 27 เมษายน 2552
ยอมรับตีพิมพ์วันที่ 25 พฤษภาคม 2552