

การประเมินประสิทธิภาพของช่วงพยากรณ์ด้วยการถดถอยควอนไทล์

กิตยากร อิศรางกูร ณ อยุธยา¹ และ ชนาพันธุ์ ชนาเนตร^{2*}

¹สำนักวิชาศึกษาทั่วไป สถาบันเทคโนโลยีจิตรลดา กรุงเทพมหานคร 10300 ประเทศไทย

²ภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์ประยุกต์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ กรุงเทพมหานคร 10800 ประเทศไทย

* ผู้นิพนธ์ประสานงาน อีเมล: chanaphun.c@sci.kmutnb.ac.th DOI:10.14416/JASET.KMUTNB.2026.02.005

รับเมื่อ 30 มิถุนายน 2569 แก้ไขเมื่อ 11 สิงหาคม 2569 ตอรับเมื่อ 12 สิงหาคม 2569

บทคัดย่อ

การวิจัยนี้เปรียบเทียบประสิทธิภาพของช่วงพยากรณ์จากวิธีกำลังสองน้อยสุดสามัญ (Ordinary Least Squares: OLS) และวิธีถดถอยควอนไทล์ (Quantile Regression: QR) ในการถดถอยเชิงเส้นอย่างง่าย ด้วยการจำลองมอนติคาร์โลภายใต้การแจกแจงของความคลาดเคลื่อน 5 รูปแบบ คือ การแจกแจงปกติซึ่งสอดคล้องกับข้อสมมติ และกรณีที่ละเมิดข้อสมมติ ได้แก่ การแจกแจงเบ้ ทางหนา ทางหนาชั้นรุนแรง และความแปรปรวนไม่คงที่ ประเมินประสิทธิภาพด้วยค่าความน่าจะเป็นครอบคลุม (Coverage Probability: CP) ความกว้างเฉลี่ย (Average Width: AW) และค่าคะแนนช่วงเฉลี่ย (Average Interval Score: AIS) ผลการวิจัยพบว่า ภายใต้การแจกแจงปกติ วิธี OLS ให้ความน่าจะเป็นครอบคลุมใกล้เคียงระดับที่กำหนดและให้ค่า AIS ต่ำกว่าในทุกเงื่อนไข ขณะที่วิธี QR ให้ค่าความน่าจะเป็นครอบคลุมต่ำกว่าเป้าหมายเมื่อตัวอย่างเล็กแล้วเข้าสู่ระดับเป้าหมายเมื่อตัวอย่างใหญ่ขึ้น สำหรับกรณีละเมิดข้อสมมติ วิธี QR ให้ประสิทธิภาพเหนือกว่าเมื่อตัวอย่างใหญ่เพียงพอ โดยในการแจกแจงเบ้และทางหนาชั้นรุนแรงให้ค่า AIS ต่ำกว่าอย่างชัดเจน และในกรณีความแปรปรวนไม่คงที่ ค่าความน่าจะเป็นครอบคลุมวิธี OLS เบี่ยงเบนจากระดับเป้าหมายไปมากตามตำแหน่งค่าพยากรณ์ (x_0) ขณะที่วิธี QR ปรับตัวได้ดีกว่า ผลการวิจัยชี้ว่าการเลือกวิธีควรพิจารณาทั้งลักษณะการแจกแจงของความคลาดเคลื่อนและขนาดตัวอย่าง โดยวิธี QR เหมาะสมเมื่อข้อสมมติของวิธี OLS ถูกละเมิดและมีขนาดตัวอย่างเพียงพอ

คำสำคัญ: การถดถอยควอนไทล์, วิธีกำลังสองน้อยสุดสามัญ, ช่วงพยากรณ์, การจำลองมอนติคาร์โล, ค่าคะแนนช่วง

Evaluating the Performance of Quantile Regression-Based Prediction Intervals

Kittayakarn Isarangura Na Ayuthya¹ and Chanaphun Chananet ^{2*}

¹Office of General Education, Chitralada Technology Institute, Bangkok 10300, Thailand

²Department of Applied Statistics, Faculty of Applied Science, King's Mongkut University of Technology North Bangkok, Bangkok 10800, Thailand

* Corresponding Author, E-mail: chanaphun.c@sci.kmutnb.ac.th DOI: 10.14416/JASET.KMUTNB.2026.02.005

Received 30 June 2026; **Revised** 11 August 2026; **Accepted** 12 August 2026

ABSTRACT

This study compares the performance of prediction intervals from Ordinary Least Squares (OLS) and Quantile Regression (QR) in simple linear regression. Using Monte Carlo simulations, five error distributions were examined: the normal distribution (which satisfies the model assumptions) and four that violate the assumptions (skewed, heavy-tailed, extremely heavy-tailed, and heteroscedastic). Performance was evaluated using Coverage Probability (CP), Average Width (AW), and Average Interval Score (AIS).

The results showed that under the normal distribution, OLS achieved coverage close to the nominal level and provided lower AIS values across all conditions. In contrast, QR undercovered at small sample sizes but converged to the nominal level as the sample size increased. When assumptions were violated, QR outperformed OLS provided that the sample size was sufficiently large. Specifically, for skewed and extremely heavy-tailed errors, QR yielded a clearly lower AIS. Under heteroscedasticity, the coverage probability of OLS deviated significantly from the target level depending on the prediction point (x_0), whereas QR adapted more effectively.

The findings suggest that selecting the appropriate method should take into account both the error distribution and the sample size. QR is a suitable alternative when OLS assumptions are violated, and the sample size is adequate.

KEYWORDS: Quantile regression, Ordinary least squares, Prediction interval, Monte Carlo simulation, Interval score

1. บทนำ

การวิเคราะห์การถดถอยเชิงเส้น (Linear Regression Analysis) เป็นเทคนิคทางสถิติที่ใช้กันอย่างแพร่หลายในการอธิบายความสัมพันธ์ระหว่างตัวแปรอิสระ (X) และตัวแปรตาม (Y) โดยมีวิธีกำลังสองน้อยสุดสามัญ (Ordinary Least Squares: OLS) เป็นวิธีประมาณค่าพารามิเตอร์ที่นิยมใช้กัน ซึ่งให้ตัวประมาณค่าที่มีคุณสมบัติที่ดีภายใต้ข้อสมมติเกี่ยวกับความคลาดเคลื่อน อย่างไรก็ตาม เมื่อข้อสมมติเหล่านี้ถูกละเมิด ความน่าเชื่อถือของการอนุมานทางสถิติที่ได้จากวิธี OLS ย่อมลดลง (Montgomery และคณะ, 2012; Greene, 2018)

ในการพยากรณ์ค่าตัวแปรตาม Y เมื่อทราบตำแหน่งของค่า X ช่วงพยากรณ์ (Prediction Interval: PI) เป็นเครื่องมือสำคัญที่ระบุกรอบค่าซึ่งคาดว่าจะครอบคลุมค่าสังเกตใหม่ภายใต้ระดับความเชื่อมั่นที่กำหนด แตกต่างจากช่วงความเชื่อมั่น (Confidence Interval: CI) ที่ครอบคลุมเฉพาะค่าเฉลี่ยแบบมีเงื่อนไข อย่างไรก็ตาม ช่วงพยากรณ์ที่สร้างจากวิธี OLS ตั้งอยู่บนข้อสมมติว่าความคลาดเคลื่อนมีการแจกแจงปกติและความแปรปรวนคงที่ ในทางปฏิบัติข้อมูลจริงมักมีลักษณะที่ขัดกับข้อสมมติดังกล่าว เช่น การแจกแจงเบ้ (skewness) การแจกแจงหางหนา (heavy-tailed distribution) หรือความแปรปรวนไม่คงที่ (heteroscedasticity) ซึ่งทำให้ช่วงพยากรณ์จากวิธี OLS อาจมีค่าความน่าจะเป็นครอบคลุม (coverage probability) เบี่ยงเบนไปจากระดับความเชื่อมั่นที่กำหนด

ด้วยเหตุนี้จึงมีการพิจารณาวิธีที่ยืดหยุ่นและทนทานต่อการละเมิดข้อสมมติได้มากขึ้น หนึ่งในวิธีนั้นคือวิธีถดถอยควอนไทล์ (Quantile Regression: QR) ที่เสนอโดย Koenker และ Bassett (1978) ซึ่งสามารถประมาณค่าควอนไทล์ ณ ตำแหน่งต่าง ๆ ของการแจกแจงแบบมีเงื่อนไขของตัวแปรตามได้ จึงมีความแข็งแกร่ง (robust) ต่อค่าผิดปกติและสามารถรองรับความแปรปรวนที่ไม่คงที่ได้โดยไม่ต้องอาศัยข้อสมมติที่เข้มงวดเช่นเดียวกับวิธี OLS (Koenker, 2005) คุณสมบัตินี้ทำให้วิธี QR ถูกนำไปประยุกต์ใช้อย่างกว้างขวางในหลากหลายสาขา เช่น เศรษฐศาสตร์ (Koenker และ Hallock, 2001) ชีวสถิติและการแพทย์ (Wei และคณะ, 2006) นิเวศวิทยา (Cade และ Noon, 2003) ตลอดจนวิทยาการข้อมูลและการเรียนรู้ของเครื่อง (Meinshausen, 2006)

อย่างไรก็ตาม ประสิทธิภาพของช่วงพยากรณ์ไม่ว่าจะสร้างจากวิธีใดนั้นมิได้ขึ้นอยู่กับวิธีการประมาณค่าเพียงอย่างเดียว แต่ยังได้รับอิทธิพลจากปัจจัยเชิงข้อมูลหลายประการด้วยกัน ประการแรก ขนาดตัวอย่าง (sample size) มีผลโดยตรงต่อความน่าจะเป็นที่ครอบคลุมและความกว้างของช่วงพยากรณ์ Yang และ Mo (1990) พบว่าสำหรับตัวอย่างขนาดเล็กความน่าจะเป็นที่ครอบคลุมเชิงประจักษ์มักเบี่ยงเบนจากระดับความครอบคลุมที่กำหนด (nominal coverage level) และจะเริ่มลู่เข้าหากันเมื่อขนาดตัวอย่างมีค่ามากกว่า 50 ประการที่สอง ระดับความเชื่อมั่น (confidence level) มีความสัมพันธ์โดยตรงกับความกว้างของช่วงพยากรณ์ โดยระดับความเชื่อมั่นที่สูงขึ้นย่อมนำไปสู่ช่วงที่กว้างขึ้น ซึ่งส่งผลต่อความคมชัดในการพยากรณ์ (Olive, 2007) ประการที่สาม รูปแบบของการแจกแจงความคลาดเคลื่อน มีผลกระทบอย่างมีนัยสำคัญต่อความเหมาะสมของช่วงพยากรณ์ โดยเฉพาะเมื่อข้อมูลมีลักษณะเบ้หรือหางหนัก ซึ่งวิธี OLS ที่อาศัยข้อสมมติความเป็นปกติจะสูญเสียความน่าเชื่อถือ ในขณะที่วิธีที่ทนทานกว่าอาจให้ผลที่ดีกว่า (Mak, 2000)

ปัจจัยอีกประการที่สำคัญแต่ถูกมองข้ามในการศึกษาเชิงการจำลอง คือ การแจกแจงของตัวแปรอิสระและตำแหน่งของจุดพยากรณ์ (prediction location) Schmoeyer (1992) และ Olive (2007) ชี้ให้เห็นว่าความแปรปรวนของการพยากรณ์เพิ่มขึ้นตามระยะห่างของ X_0 จากค่าเฉลี่ย $\bar{X}$ ของข้อมูล ส่งผลให้ความกว้างของช่วงพยากรณ์แปรผันตามตำแหน่งพยากรณ์อย่างเป็นระบบ และยิ่งในกรณีของการพยากรณ์นอกช่วงข้อมูล (extrapolation) ทั้ง OLS และ QR ต่างมีข้อจำกัดที่เพิ่มขึ้นอย่างมีนัยสำคัญ (Stine, 1985) ดังนั้นการศึกษาเปรียบเทียบประสิทธิภาพของช่วงพยากรณ์จาก OLS และ QR ที่ครอบคลุมปัจจัยเหล่านี้อย่างเป็นระบบภายใต้สถานการณ์จำลองที่หลากหลายจึงมีความจำเป็นเพื่อให้ได้แนวทางปฏิบัติที่เชื่อถือได้สำหรับการเลือกวิธีที่เหมาะสมในบริบทข้อมูลจริง

จากที่กล่าวมา การเปรียบเทียบประสิทธิภาพของช่วงพยากรณ์ระหว่างวิธี OLS และวิธี QR จึงเป็นประเด็นที่มีความสำคัญ งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบประสิทธิภาพของช่วงพยากรณ์ที่สร้างจากวิธี OLS และวิธี QR ภายใต้สถานการณ์จำลองที่มีลักษณะข้อมูลแตกต่างกัน เพื่อสะท้อนแนวทางการเลือกใช้วิธีที่เหมาะสม

สำหรับการวิเคราะห์ในบริบทต่าง ๆ ที่ต้องการความน่าเชื่อถือของการพยากรณ์

2. ทฤษฎีที่เกี่ยวข้อง

2.1 ตัวแบบการถดถอยเชิงเส้นอย่างง่าย

การถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression) เป็นวิธีการทางสถิติที่ใช้ในการศึกษาความสัมพันธ์ในรูปแบบเชิงเส้นระหว่างตัวแปรตาม (Dependent Variable) และตัวแปรอิสระ (Independent Variable) โดยสามารถเขียนในรูปแบบตัวแบบความสัมพันธ์ได้ดังนี้

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, i = 1, 2, \dots, n \quad (1)$$

โดยที่ y_i คือ ค่าสังเกตของตัวแปรตามลำดับที่ i

x_i คือ ค่าสังเกตของตัวแปรอิสระลำดับที่ i

β_0 คือ พารามิเตอร์จุดตัดแกน Y

β_1 คือ พารามิเตอร์ความชันของเส้นถดถอย

ε_i คือ ค่าคลาดเคลื่อนสุ่มลำดับที่ i

n คือ จำนวนค่าสังเกต

ในตัวอย่าง (1) เป็นตัวแบบรวมของทั้งวิธีกำลังสองน้อยสุดสามัญ (Ordinary Least Square: OLS) และวิธีถดถอยควอนไทล์ (Quantile Regression: QR) ความแตกต่างระหว่างวิธี OLS และวิธี QR อยู่ที่ฟังก์ชันความสูญเสีย (Loss Function) ที่ใช้ในการประมาณค่าพารามิเตอร์และข้อสมมติของตัวแบบ ทฤษฎีบทเกาส์-มาร์คอฟ (Gauss-Markov Theorem) กำหนดข้อสมมติที่จำเป็นของความคลาดเคลื่อน (ε_i) ประกอบด้วย 1) ค่าคาดหวังแบบมีเงื่อนไขของความคลาดเคลื่อนมีค่าเป็นศูนย์ นั่นคือ $E(\varepsilon_i | X_i) = 0$ 2) ความแปรปรวนของความคลาดเคลื่อนมีค่าคงที่ นั่นคือ $Var(\varepsilon_i) = \sigma^2$ และ 3) ค่าคลาดเคลื่อนแต่ละค่าต้องเป็นอิสระกัน นั่นคือ $Cov(\varepsilon_i, \varepsilon_j) = 0$ เมื่อ $i \neq j$ (Montgomery และคณะ, 2012) หากข้อสมมติทั้งสามข้อนี้เป็นจริงแล้ว ตัวประมาณค่าพารามิเตอร์ที่ได้จากวิธี OLS จะมีคุณสมบัติเป็นตัวประมาณเชิงเส้นที่ไม่มีความเอนเอียงและมีประสิทธิภาพสูงสุด (Best Linear Unbiased Estimator: BLUE) โดยคุณสมบัติดังกล่าวไม่ได้กล่าวถึงรูปแบบการแจกแจงของความคลาดเคลื่อนแต่อย่างใด ส่วนช่วงพยากรณ์วิธี OLS นอกจากจะอาศัยข้อสมมติของเกาส์-มาร์คอฟแล้วยังเพิ่มข้อสมมติว่าความคลาดเคลื่อนต้องมีการแจกแจงปกติเข้าไปด้วย งานวิจัยนี้จึงมีความสนใจศึกษากรณีที่รูปแบบการ

แจกแจงของความคลาดเคลื่อนและความแปรปรวนคงที่นั้นถูกละเมิด

2.2 วิธีกำลังสองน้อยที่สุดสามัญและช่วงพยากรณ์

ในการประมาณค่าพารามิเตอร์ของตัวแบบถดถอยเชิงเส้นอย่างง่ายด้วยวิธีกำลังสองน้อยสุดสามัญ (Ordinary Least Square: OLS) ตัวแบบอยู่ในรูปสมการ (1) โดยค่าพยากรณ์ของตัวแบบคือ $\hat{y}_i = \beta_0 + \beta_1 x_i$ จะได้ค่าความคลาดเคลื่อนของค่าสังเกตลำดับที่ i คือ $e_i = y_i - \hat{y}_i$ วิธี OLS จะเลือกค่า β_0 และ β_1 ที่ทำให้ผลรวมกำลังสองของความคลาดเคลื่อนมีค่าน้อยที่สุด โดยกำหนดให้ฟังก์ชันความสูญเสีย (Loss Function) ของวิธี OLS สำหรับค่าสังเกต n ตัว คือ

$$L(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \quad (2)$$

ดังนั้นตัวประมาณค่าวิธี OLS คือ $\hat{\beta}_0$ และ $\hat{\beta}_1$ ที่ทำให้ฟังก์ชันความสูญเสียมีค่าต่ำสุด นั่นคือ

$$(\hat{\beta}_0, \hat{\beta}_1) = \arg \min_{\beta_0, \beta_1} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \quad (3)$$

ผลลัพธ์ประการแรกที่ได้จากการลดฟังก์ชันความสูญเสียคือค่าพยากรณ์แบบจุดซึ่งระบุตำแหน่งศูนย์กลางของการแจกแจงแบบมีเงื่อนไขของ Y ณ ตำแหน่ง x_0 ดังแสดงในสมการ (4)

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0 \quad (4)$$

ผลลัพธ์ประการที่สองคือค่าของฟังก์ชันสูญเสีย ณ จุดต่ำสุดซึ่งนิยามเป็นผลรวมกำลังสองของส่วนเหลือ (Sum Squares Error: SSE)

$$SSE = \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \quad (5)$$

ค่า SSE ในสมการ (5) นำไปสู่ตัวประมาณค่าความแปรปรวนของความคลาดเคลื่อนแบบไม่เอนเอียง (S^2) ดังแสดงในสมการ (6)

$$S^2 = \frac{SSE}{n-2} \quad (6)$$

จากการลดฟังก์ชันความสูญเสีย สามารถระบุตำแหน่งศูนย์กลางของการแจกแจงแบบมีเงื่อนไขของ Y ณ ตำแหน่ง x_0 หรือ $E(Y | X = x_0)$ ได้จากสมการ (4) และเมื่อนำค่า SSE ในสมการ (5) ไปประมาณค่าความแปรปรวนของความคลาดเคลื่อนโดยใช้ S^2 ในสมการ (6) จะได้มาตราส่วนที่ใช้กำหนดความกว้างของช่วงพยากรณ์ อย่างไรก็ตามตำแหน่งศูนย์กลางและมาตราส่วนยังไม่เพียงพอต่อการ

กำหนดรูปร่างของช่วงซึ่งจำเป็นต้องอาศัยข้อสมมติเกี่ยวกับการแจกแจงของความคลาดเคลื่อนเพิ่มเติม

ภายใต้ข้อสมมติของความคลาดเคลื่อนมีการแจกแจงปกติ มีค่าเฉลี่ยเท่ากับศูนย์และมีความแปรปรวนคงที่เท่ากับ σ^2 พิจารณาค่าสังเกตใหม่ Y_0 ณ ตำแหน่ง x_0 ที่เป็นอิสระจากชุดข้อมูลที่ใช้ประมาณค่า ความแปรปรวนของความคลาดเคลื่อนในการพยากรณ์ $Y_0 - \hat{y}_0$ ซึ่งประกอบด้วยความผันแปรของค่าสังเกตใหม่และความผันแปรของตัวประมาณค่าพารามิเตอร์ สามารถแสดงได้ในสมการ (7)

$$\text{Var}(Y_0 - \hat{y}_0) = \sigma^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right] \quad (7)$$

โดยที่ Y_0 คือ ค่าสังเกตใหม่ของตัวแปรตาม ณ ตำแหน่ง x_0 σ^2 คือ ความแปรปรวนของความคลาดเคลื่อน

$\bar{x}$ คือ ค่าเฉลี่ยของตัวแปรอิสระ ซึ่ง $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$

เมื่อประมาณค่า σ^2 ด้วยตัวประมาณค่า S^2 จากสมการ (6) อัตราส่วนระหว่างความคลาดเคลื่อนในการพยากรณ์กับค่าความคลาดเคลื่อนมาตรฐานโดยประมาณจะมีการแจกแจงแบบ t ที่มีองศาความเป็นอิสระ $n-2$ นำไปสู่ช่วงพยากรณ์ (Prediction Interval: PI) วิธี OLS ที่ระดับความเชื่อมั่น $1-\alpha$ สำหรับค่าสังเกตใหม่ ณ ตำแหน่ง x_0 ดังสมการ (8)

$$PI_{OLS}(x_0) = \hat{y}_0 \pm t_{1-\alpha/2, n-2} S \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \quad (8)$$

โดยที่ $t_{1-\alpha/2, n-2}$ คือ ค่าควอนไทล์ที่ $1-\alpha/2$ ของการแจกแจง t ที่องศาอิสระ $n-2$

S คือ ค่าคลาดเคลื่อนมาตรฐานของสมการถดถอย

จะเห็นว่าช่วงพยากรณ์ในสมการ (8) ตั้งอยู่บนข้อสมมติว่าความแปรปรวนของความคลาดเคลื่อนมีค่าคงที่เท่ากับ σ^2 ที่ประมาณจากค่า x_i เมื่อ $i=1, 2, \dots, n$ ส่งผลให้กว้างของช่วงพยากรณ์ไม่ปรับเปลี่ยนตามตำแหน่ง x_0 ใดๆ ดังนั้นเมื่อข้อสมมติดังกล่าวถูกละเมิด เช่น ในกรณีที่ข้อมูลมีความแปรปรวนไม่คงที่ ช่วงพยากรณ์ที่ได้จากวิธี OLS อาจมีค่าความน่าจะเป็นครอบคลุมเบี่ยงเบนไปจากระดับความ

เชื่อมั่นที่กำหนด ซึ่งเป็นข้อจำกัดสำคัญที่นำไปสู่การพิจารณาวิธีถดถอยควอนไทล์ในหัวข้อถัดไป

2.3 วิธีถดถอยควอนไทล์และช่วงพยากรณ์

วิธีถดถอยควอนไทล์ (Quantile Regression: QR) นำเสนอโดย Koenker และ Bassett (1978) โดยการเปลี่ยนรูปแบบของฟังก์ชันความสูญเสียจากวิธี OLS ที่มุ่งเน้นการประมาณค่าเฉลี่ยแบบมีเงื่อนไขของตัวแปรตาม Y ไปเป็นฟังก์ชันความสูญเสียแบบไม่สมมาตร (asymmetric loss function) ซึ่งมุ่งเน้นการประมาณค่าควอนไทล์แบบมีเงื่อนไข ณ ตำแหน่งใด ๆ ของการแจกแจงของตัวแปรตาม Y

สำหรับระดับควอนไทล์ τ โดยที่ $\tau \in (0, 1)$ ตัวแบบการถดถอยควอนไทล์เชิงเส้นอย่างง่าย คือ

$$Q_\tau(\tau | X = x) = \beta_0(\tau) + \beta_1(\tau)x \quad (9)$$

โดยที่ $\beta_0(\tau)$ และ $\beta_1(\tau)$ คือ พารามิเตอร์จุดตัดแกน (intercept) และพารามิเตอร์ความชัน (slope) ตามลำดับ พารามิเตอร์ทั้งสองค่านี้เป็นฟังก์ชันที่ขึ้นกับระดับของค่าควอนไทล์ τ

สำหรับฟังก์ชันสูญเสียของวิธีถดถอยควอนไทล์ หรือเรียกอีกชื่อว่า ฟังก์ชันตรวจสอบ (check function) โดยกำหนดเงื่อนไขในการถ่วงน้ำหนักความคลาดเคลื่อน u ใด ๆ ไว้ดังนี้

$$\rho_\tau(u) = \begin{cases} \tau u, & u \geq 0 \\ (\tau - 1)u, & u < 0 \end{cases} \quad (10)$$

ฟังก์ชันนี้ให้น้ำหนักกับความคลาดเคลื่อนทางบวกด้วย τ และความคลาดเคลื่อนทางลบด้วย $(1-\tau)$

ตัวประมาณค่าการถดถอยควอนไทล์ได้จากการลดผลรวมของฟังก์ชันความสูญเสียสำหรับค่าสังเกต n ตัว โดยฟังก์ชันความสูญเสียวิธี QR ที่ระดับ τ คือ

$$L_\tau(\beta_0, \beta_1) = \sum_{i=1}^n \rho_\tau(y_i - \beta_0 - \beta_1 x_i) \quad (11)$$

ดังนั้นตัวประมาณค่า $\hat{\beta}_0(\tau)$ และ $\hat{\beta}_1(\tau)$ ที่ทำให้ฟังก์ชันความสูญเสียมีค่าต่ำสุด คือ

$$(\hat{\beta}_0(\tau), \hat{\beta}_1(\tau)) = \arg \min_{\beta_0, \beta_1} \sum_{i=1}^n \rho_\tau(y_i - \beta_0 - \beta_1 x_i) \quad (12)$$

เนื่องจากฟังก์ชันดังกล่าวไม่สามารถหาผลเฉลยแบบปิด (Closed-form Solution) ได้ การประมาณค่า $\hat{\beta}_0(\tau)$ และ $\hat{\beta}_1(\tau)$ จึงต้องอาศัยการหาค่าเหมาะสมที่สุดด้วยวิธี

กำหนดการเชิงเส้น (Linear Programming) (Koenker, 2005)

สำหรับการสร้างช่วงพยากรณ์โดยวิธี QR สามารถหาจากฟังก์ชันความสูญเสียได้โดยตรงโดยไม่ต้องอาศัยข้อสมมติเกี่ยวกับการแจกแจงของความคลาดเคลื่อนเพิ่มเติมเหมือนในกรณีของวิธี OLS ดังนั้นช่วงพยากรณ์ (Prediction Interval: PI) วิธี QR ที่ระดับความเชื่อมั่น $1-\alpha$ สำหรับค่าสังเกตใหม่ ณ ตำแหน่ง x_0 คือ

$$PI_{QR}(x_0) = [\hat{\beta}_0(\tau_L) + \hat{\beta}_1(\tau_L)x_0, \hat{\beta}_0(\tau_U) + \hat{\beta}_1(\tau_U)x_0] \quad (13)$$

โดยที่ τ_L และ τ_U คือ ระดับควอนไทล์ที่ $\alpha/2$ และ $1-\alpha/2$ ตามลำดับ

3. วิธีดำเนินการวิจัย

3.1.1 ข้อมูล

งานวิจัยนี้ใช้ข้อมูลจากการจำลองมอนติคาร์โล (Monte Carlo Simulation) เพื่อเปรียบเทียบประสิทธิภาพของช่วงพยากรณ์วิธี OLS และวิธี QR ภายใต้ตัวแบบความสัมพันธ์เชิงเส้นอย่างง่าย ดังแสดงในสมการ (1) กำหนด n แทนจำนวนตัวอย่างสุ่มให้ตัวแปรอิสระ x_i เมื่อ $i=1,2,\dots,n$ สุ่มจากการแจกแจงเอกรูป (Uniform Distribution) บนช่วง $(-1, 1)$ พารามิเตอร์ β_0 และ β_1 มีค่าเป็น 5 และ 10 ตามลำดับ และการแจกแจงของความคลาดเคลื่อนที่แตกต่างกัน

3.1.2 การแจกแจงของความคลาดเคลื่อน

เพื่อศึกษาความแกร่ง (Robustness) ของแต่ละวิธีในการสร้างช่วงพยากรณ์ ภายใต้เงื่อนไขที่หลากหลายของความคลาดเคลื่อน งานวิจัยนี้กำหนดรูปแบบการแจกแจงของความคลาดเคลื่อนไว้ 5 รูปแบบ ดังนี้

(1) ความคลาดเคลื่อนเป็นไปตามข้อสมมติของวิธี OLS นั่นคือ $\varepsilon_i \sim N(0, \sigma^2)$ โดย $\sigma=5$

(2) ความคลาดเคลื่อนมีการแจกแจงแบบ ใช้การแจกแจงเลขชี้กำลัง (Exponential distribution) โดยปรับค่าเฉลี่ยให้มีค่าเป็น 0 นั่นคือ $\varepsilon_i \sim \text{Expo}(1/\lambda=5)-5$

(3) ความคลาดเคลื่อนมีการแจกแจงหางหนา (heavy-tailed distribution) โดยใช้การแจกแจงที่ทอัสคาลิสระ 5 นั่นคือ $\varepsilon_i \sim t(v=5)$

(4) ความคลาดเคลื่อนมีการแจกแจงหางหนารุนแรงแบบไม่นิยามค่าเฉลี่ยและความแปรปรวน (very heavy-

tailed with undefined moments) โดยใช้การแจกแจงโคชี นั่นคือ $\varepsilon_i \sim \text{Cauchy}(0,1)$

(5) กรณีความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ (heteroscedasticity) โดยกำหนดให้ค่าส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนมีความสัมพันธ์กับค่าของตัวแปรอิสระ ดังแสดงในสมการ (14) (He, 1997; Koenker, 2005)

$$\varepsilon_i \sim N\left(0, [\sigma_0(1+\gamma|x_i|)]^2\right) \quad (14)$$

โดยที่ σ_0 คือ ค่าส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อน กำหนดให้มีค่าเป็น 3

γ คือ พารามิเตอร์ที่ควบคุมระดับความรุนแรงของความแปรปรวนไม่คงที่ (heteroscedasticity parameter) ถ้า $\gamma=0$ จะลดรูปเป็นความแปรปรวนคงที่ (homoscedasticity) ในงานวิจัยนี้กำหนดให้ค่า $\gamma=1.5$

3.1.3 ค่าในการจำลอง

(1) กำหนดขนาดตัวอย่าง (n) = 10, 30, 50, 100, 1,000 และ 5,000

(2) ประเมินประสิทธิภาพของช่วงพยากรณ์ ณ ตำแหน่งค่าพยากรณ์ $x_0 = 0, 0.5, 0.8$ (ภายในพิสัยของข้อมูล) และ $x_0 = 1.2$ และ 1.5 (ภายนอกพิสัยของข้อมูล)

(3) กำหนดจำนวนรอบการจำลอง 20,000 รอบ ในแต่ละสถานการณ์

(4) พิจารณาสองช่วงพยากรณ์ที่ระดับความเชื่อมั่น 95%

3.1.4 การประมาณค่าและการสร้างช่วงพยากรณ์

ในการจำลองแต่ละรอบ ช่วงพยากรณ์วิธี OLS ที่คำนวณตามสมการ (8) และช่วงพยากรณ์วิธี QR ที่คำนวณตามสมการ (13) ที่ระดับ $\tau_L=0.025$ และ $\tau_U=0.975$ ดำเนินการด้วยโปรแกรม R (R Core Team, 2025)

3.1.5 เกณฑ์ประเมินประสิทธิภาพของช่วงพยากรณ์

เพื่อเปรียบเทียบประสิทธิภาพของช่วงพยากรณ์ที่ได้จากวิธี OLS และวิธี QR งานวิจัยนี้ใช้เกณฑ์การประเมิน 3 เกณฑ์ ได้แก่ ค่าความน่าจะเป็นครอบคลุม (coverage probability) ความกว้างเฉลี่ย (average width) และค่าคะแนนของช่วง (interval score) กำหนดให้ M คือ จำนวนรอบในการจำลอง และ ให้ $[L_m, U_m]$ แทนช่วงพยากรณ์ในรอบที่ m ที่ได้ และ $y_{0,m}$ แทนค่าสังเกตใหม่ในรอบที่ m ที่ใช้ในการประเมิน

ตารางที่ 1 ค่าความน่าจะเป็นครอบคลุม (CP) ค่าความกว้างช่วง (AW) และค่าคะแนนช่วงเฉลี่ย (AIS) ของช่วงพยากรณ์วิธี OLS และวิธี QR กรณี

$$\varepsilon_i \sim N(0, 5^2)$$

n	x ₀	CP_OLS	AW_OLS	AIS_OLS	CP_QR	AW_QR	AIS_QR
10	0.0	0.9498	23.73	28.15 *	0.7202	12.24	44.75
	0.5	0.9492	24.69	29.35 *	0.6718	12.26	54.96
	0.8	0.9514	26.13	31.08 *	0.6105	12.27	73.02
	1.2	0.9520	28.84	33.95 *	0.5336	12.28	107.03
	1.5	0.9534	31.34	36.72 *	0.4930	12.30	140.87
30	0.0	0.9496	20.65	24.67 *	0.8990	17.84	26.92
	0.5	0.9529	20.92	24.66 *	0.8804	17.84	28.95
	0.8	0.9520	21.32	25.12 *	0.8482	17.83	33.81
	1.2	0.9496	22.13	26.27 *	0.7840	17.83	45.52
	1.5	0.9500	22.92	27.30 *	0.7300	17.83	60.37
50	0.0	0.9501	20.18	24.07 *	0.9290	19.15	25.20
	0.5	0.9498	20.34	24.26 *	0.9188	19.17	26.48
	0.8	0.9506	20.57	24.61 *	0.9010	19.18	28.82
	1.2	0.9461	21.05	25.40 *	0.8569	19.20	34.96
	1.5	0.9476	21.52	25.84 *	0.8200	19.21	42.21
1,00	0.0	0.9480	19.88	23.85 *	0.9378	19.32	24.33
	0.5	0.9501	19.96	23.42 *	0.9322	19.31	24.46
	0.8	0.9488	20.08	24.17 *	0.9204	19.31	26.36
	1.2	0.9531	20.31	23.98 *	0.9060	19.30	28.14
	1.5	0.9478	20.55	24.74 *	0.8830	19.30	31.67
1000	0.0	0.9502	19.63	23.17 *	0.9489	19.58	23.24
	0.5	0.9494	19.64	23.42 *	0.9476	19.58	23.52
	0.8	0.9508	19.65	23.42 *	0.9488	19.58	23.56
	1.2	0.9494	19.67	23.57 *	0.9434	19.58	23.95
	1.5	0.9500	19.70	23.62 *	0.9436	19.58	24.10
5,000	0.0	0.9526	19.60	23.02 *	0.9524	19.59	23.02
	0.5	0.9512	19.61	23.24 *	0.9502	19.59	23.26
	0.8	0.9491	19.61	23.33 *	0.9483	19.59	23.35
	1.2	0.9479	19.61	23.43 *	0.9476	19.59	23.49
	1.5	0.9504	19.62	23.21 *	0.9492	19.59	23.36

หมายเหตุ ตัวเลขหนา หมายถึง ค่า CP ที่ไม่ต่ำกว่า 0.95 อย่างมีนัยสำคัญ (CP ≥ 0.9475 ซึ่งเป็นขอบล่างทางเดียวที่ระดับนัยสำคัญ 0.05 จากการจำลอง 20,000 รอบ)

* หมายถึง ค่า AIS ที่มีค่าต่ำกว่า (สำหรับกรณีที่ไม่มี CP ค่าได้ถึงเกณฑ์ ค่า AIS ที่ต่ำกว่าบางชี้เพียงวิธีที่ด้อยกว่าอีกวิธีเท่านั้น)

ค่าความน่าจะเป็นครอบคลุม

ค่าความน่าจะเป็นครอบคลุม (Coverage Probability: CP) คือ ค่าสัดส่วนของค่าสังเกตใหม่ที่ตกอยู่ในช่วงพยากรณ์สามารถคำนวณได้จากสมการ (15)

$$CP = \frac{1}{M} \sum_{m=1}^M I(L_m \leq y_{0,m} \leq U_m) \quad (15)$$

โดยที่ $I(\cdot)$ เป็นฟังก์ชันบ่งชี้ วิธีที่ให้ช่วงพยากรณ์ที่ดีควรมีค่า CP อย่างน้อย $1-\alpha$ ตามที่ได้กำหนดไว้ สำหรับในงานวิจัยนี้กำหนดจำนวนรอบจำลอง $M = 20,000$ รอบ ความคลาดเคลื่อนมาตรฐานเชิงมอนติคาร์โลของ CP ที่ระดับ $1-\alpha = 0.95$ เท่ากับ $\sqrt{0.95(0.05)/20000}$

≈ 0.00154 ดังนั้นขอบล่างแบบทางเดียวที่ระดับนัยสำคัญ 0.05 คือ $0.95 - 1.645 \times 0.00154 \approx 0.9475$

ความกว้างเฉลี่ย

ความกว้างเฉลี่ย (Average Width: AW) คือ ค่าเฉลี่ยความกว้างของช่วงพยากรณ์ทุกรอบของการจำลอง คำนวณได้ดังสมการ (16)

$$AW = \frac{1}{M} \sum_{m=1}^M (U_m - L_m) \quad (16)$$

ภายใต้ระดับความครอบคลุมที่มีค่าเกินเกณฑ์ 0.9475 วิธีที่ให้ค่า AW น้อยกว่าย่อมมีประสิทธิภาพมากกว่า เนื่องจากให้ช่วงพยากรณ์ที่แคบกว่าในขณะที่ยังครอบคลุมค่าจริงได้

ตารางที่ 2 ค่าความน่าจะเป็นครอบคลุม (CP) ค่าความกว้างช่วง (AW) และค่าคะแนนช่วงเฉลี่ย (AIS) ของช่วงพยากรณ์วิธี OLS และวิธี QR กรณี

$$\varepsilon_i \sim \text{Expo}(1/\lambda = 5) - 5$$

n	x ₀	CP_OLS	AW_OLS	AIS_OLS	CP_QR	AW_QR	AIS_QR
10	0.0	0.9312	22.47	36.14 *	0.7006	11.07	43.21
	0.5	0.9326	23.39	36.26 *	0.6698	11.09	51.53
	0.8	0.9332	24.75	37.92 *	0.6191	11.10	67.87
	1.2	0.9346	27.32	40.64 *	0.5383	11.11	103.73
	1.5	0.9362	29.68	40.95 *	0.4964	11.12	140.93
30	0.0	0.9401	20.16	32.19	0.8879	16.85	27.16 *
	0.5	0.9425	20.42	31.78	0.8787	16.85	28.84 *
	0.8	0.9431	20.81	32.45 *	0.8530	16.85	33.44
	1.2	0.9432	21.60	32.55 *	0.7832	16.85	49.89
	1.5	0.9422	22.37	33.49 *	0.7333	16.85	72.15
50	0.0	0.9460	19.91	31.00	0.9242	18.32	25.43 *
	0.5	0.9468	20.06	30.46	0.9171	18.30	25.81 *
	0.8	0.9460	20.30	30.54	0.9036	18.29	28.68 *
	1.2	0.9470	20.77	31.20 *	0.8603	18.28	37.35
	1.5	0.9464	21.24	31.67 *	0.8120	18.26	50.23
100	0.0	0.9461	19.75	30.38	0.9368	18.23	24.25 *
	0.5	0.9446	19.82	30.97	0.9292	18.22	25.34 *
	0.8	0.9438	19.94	31.56	0.9224	18.22	27.15 *
	1.2	0.9475	20.17	30.76	0.9058	18.22	29.66 *
	1.5	0.9486	20.41	31.15 *	0.8840	18.22	35.33
1,000	0.0	0.9482	19.60	30.18	0.9498	18.31	23.81 *
	0.5	0.9496	19.61	29.16	0.9513	18.30	22.95 *
	0.8	0.9484	19.62	29.39	0.9482	18.29	23.15 *
	1.2	0.9460	19.65	30.61	0.9442	18.28	24.38 *
	1.5	0.9470	19.67	30.31	0.9436	18.28	24.38 *
5,000	0.0	0.9491	19.60	29.90	0.9502	18.31	23.35 *
	0.5	0.9488	19.60	29.53	0.9499	18.32	23.07 *
	0.8	0.9512	19.60	29.44	0.9497	18.32	23.20 *
	1.2	0.9483	19.61	29.69	0.9508	18.32	23.34 *
	1.5	0.9475	19.61	30.20	0.9482	18.32	23.73 *

หมายเหตุ ตัวเลขหนา หมายถึง ค่า CP ที่ไม่ต่ำกว่า 0.95 อย่างมีนัยสำคัญ (CP ≥ 0.9475 ซึ่งเป็นขอบล่างทางเดียวที่ระดับนัยสำคัญ 0.05 จากการจำลอง 20,000 รอบ)

* หมายถึง ค่า AIS ที่มีค่าต่ำกว่า (สำหรับกรณีที่ไม่มี CP ค่าใดถึงเกณฑ์ ค่า AIS ที่ต่ำกว่าบ่งชี้เพียงวิธีที่ด้อยกว่าอีกวิธีเท่านั้น)

สัดส่วนเดียวกัน อย่างไรก็ตาม ค่า AW และ CP มีความสัมพันธ์ที่ต้องแลกเปลี่ยนกัน (trade-off) กล่าวคือ การลดความกว้างของช่วงย่อมทำให้ค่าความเบี่ยงเบนของความครอบคลุมสูงขึ้น ในขณะที่การขยายช่วงเพื่อให้ค่า CP ใกล้เคียงระดับความเชื่อมั่นที่กำหนดย่อมทำให้ค่า AW สูงขึ้น จึงไม่สามารถพิจารณาเกณฑ์ใดเกณฑ์หนึ่งโดยลำพังได้ และจำเป็นต้องอาศัยเกณฑ์ที่ประเมินทั้งสองมิติพร้อมกัน Gneiting และ Raftery (2007) ดังนั้นเสนอในหัวข้อถัดไป

ค่าคะแนนช่วง (Interval Score: IS) เป็นเกณฑ์ที่รวมการประเมินความกว้างของช่วงและความสามารถในการ

ครอบคลุมค่าจริงไว้ในค่าเดียว จัดเป็นกฎการให้คะแนนที่เหมาะสมสำหรับช่วงพยากรณ์ในรอบที่ m ค่า IS สามารถคำนวณได้ดังสมการ (17)

$$IS_{\alpha,m} = (U_m - L_m) + \frac{2}{\alpha} (L_m - y_{0,m}) I(y_{0,m} < L_m) + \frac{2}{\alpha} (y_{0,m} - U_m) I(y_{0,m} > U_m) \quad (17)$$

ค่าคะแนนช่วงเฉลี่ย (Average Interval Score: AIS) สำหรับทุกรอบการจำลอง คำนวณได้ดังสมการ (18)

ตารางที่ 3 ค่าความน่าจะเป็นครอบคลุม (CP) ค่าความกว้างช่วง (AW) และค่าคะแนนช่วงเฉลี่ย (AIS) ของช่วงพยากรณ์วิธี OLS และวิธี QR กรณี

$$\varepsilon_i \sim t(\nu=5)$$

n	x_0	CP_OLS	AW_OLS	AIS_OLS	CP_QR	AW_QR	AIS_QR
10	0.0	0.9349	22.93	31.93 *	0.7278	11.91	48.17
	0.5	0.9364	23.86	32.23 *	0.6696	11.88	57.57
	0.8	0.9391	25.25	34.05 *	0.5948	11.87	76.43
	1.2	0.9400	27.87	36.21 *	0.4978	11.85	117.00
	1.5	0.9432	30.28	39.27 *	0.4531	11.84	157.34
30	0.0	0.9422	20.35	28.52 *	0.9059	18.89	31.19
	0.5	0.9419	20.61	28.83 *	0.8848	18.88	33.95
	0.8	0.9419	21.01	29.36 *	0.8408	18.88	40.17
	1.2	0.9436	21.81	29.66 *	0.7454	18.87	59.52
	1.5	0.9422	22.59	30.42 *	0.6670	18.87	85.12
50	0.0	0.9408	20.03	28.63 *	0.9316	20.67	30.41
	0.5	0.9447	20.18	28.56 *	0.9212	20.66	31.59
	0.8	0.9428	20.41	28.52 *	0.8934	20.66	34.60
	1.2	0.9436	20.89	28.80 *	0.8299	20.65	45.72
	1.5	0.9422	21.35	29.97 *	0.7554	20.64	61.88
100	0.0	0.9462	19.82	27.37 *	0.9411	20.17	28.20
	0.5	0.9473	19.89	27.45 *	0.9354	20.17	29.08
	0.8	0.9456	20.01	27.47 *	0.9233	20.17	30.23
	1.2	0.9419	20.25	27.85 *	0.8855	20.17	35.23
	1.5	0.9450	20.48	28.02 *	0.8486	20.17	40.64
1,000	0.0	0.9448	19.62	27.53 *	0.9462	19.95	27.54
	0.5	0.9455	19.63	27.54 *	0.9468	19.95	27.70
	0.8	0.9466	19.64	27.48 *	0.9467	19.95	27.78
	1.2	0.9454	19.66	27.69 *	0.9439	19.95	28.18
	1.5	0.9464	19.69	27.79 *	0.9404	19.95	28.78
5,000	0.0	0.9464	19.60	27.56 *	0.9491	19.92	27.57
	0.5	0.9474	19.60	27.52 *	0.9502	19.92	27.57
	0.8	0.9468	19.60	27.24 *	0.9486	19.92	27.30
	1.2	0.9466	19.61	27.60 *	0.9488	19.92	27.72
	1.5	0.9468	19.61	27.30 *	0.9472	19.92	27.46

หมายเหตุ ตัวเลขหนา หมายถึง ค่า CP ที่ไม่ต่ำกว่า 0.95 อย่างมีนัยสำคัญ ($CP \geq 0.9475$ ซึ่งเป็นขอบล่างทางเดียวที่ระดับนัยสำคัญ 0.05 จากการจำลอง 20,000 รอบ)

* หมายถึง ค่า AIS ที่มีค่าต่ำกว่า (สำหรับกรณีที่ไม่ใช่ CP ค่าใดถึงเกณฑ์ ค่า AIS ที่ต่ำกว่าบ่งชี้เพียงวิธีที่ด้อยกว่าอีกวิธีเท่านั้น)

$$AIS = \frac{1}{M} \sum_{m=1}^M IS_{\alpha, m} \quad (18)$$

ค่าคะแนนช่วงเฉลี่ยที่น้อยกว่าบ่งชี้ถึงช่วงพยากรณ์ที่มีคุณภาพดีกว่า เนื่องจากสะท้อนความสมดุลระหว่างความกว้างที่เหมาะสมและความสามารถในการครอบคลุมค่าจริง งานวิจัยนี้พิจารณาเกณฑ์แบบเข้มที่กำหนดให้ค่า CP ในแต่ละวิธีต้องครอบคลุมไม่ต่ำกว่าระดับความเชื่อมั่นที่กำหนด ($CP \geq 0.9470$) อย่างมีนัยสำคัญก่อนจึงเปรียบเทียบความกว้าง อย่างไรก็ตาม ภายใต้ขนาดตัวอย่างที่แตกต่างกัน โดยเฉพาะขนาดตัวอย่างเล็ก ($n = 10$) และการแจกแจงทาง

หนา ไม่มีวิธีผ่านเกณฑ์ความครอบคลุมตามที่กำหนด ทำให้เกณฑ์ดังกล่าวไม่สามารถระบุวิธีที่ดีกว่าได้ในหลายเงื่อนไข ดังนั้นงานวิจัยนี้จึงใช้ค่าคะแนนช่วงเฉลี่ย (AIS) ซึ่งเป็นการให้น้ำหนักคะแนนที่เหมาะสมและประสานทั้งความครอบคลุมและความกว้างไว้ในค่าเดียว เป็นเกณฑ์หลักในการเปรียบเทียบประสิทธิภาพ

4. ผลการวิจัยและการอภิปรายผล

งานวิจัยนี้นำเสนอผลการเปรียบเทียบประสิทธิภาพของช่วงพยากรณ์ระหว่างวิธี OLS และวิธี QR จำแนกตามการแจกแจงของความคลาดเคลื่อน 5 รูปแบบ ดังแสดงในตารางที่ 1

ตารางที่ 4 ค่าความน่าจะเป็นครอบคลุม (CP) ค่าความกว้างช่วง (AW) และค่าคะแนนช่วงเฉลี่ย (AIS) ของช่วงพยากรณ์วิธี OLS และวิธี QR กรณี

$$\varepsilon_i \sim \text{Cauchy}(0,1)$$

n	x_0	CP_OLS	AW_OLS	AIS_OLS	CP_QR	AW_QR	AIS_QR
10	0.0	0.9096	78.05	318.61	0.7574	35.33	301.07 *
	0.5	0.9136	81.11	221.25	0.6942	34.69	215.58 *
	0.8	0.9147	85.75	575.93 *	0.5642	34.31	633.61
	1.2	0.9190	94.55	335.38 *	0.3942	33.80	664.41
	1.5	0.9158	102.69	240.57 *	0.3537	33.42	817.06
30	0.0	0.9445	279.50	485.11	0.9210	270.68	481.36 *
	0.5	0.9408	282.64	451.05 *	0.9028	324.07	502.68
	0.8	0.9455	287.91	444.02 *	0.8492	356.09	532.41
	1.2	0.9460	298.61	421.80 *	0.4931	398.80	1439.82
	1.5	0.9436	309.15	442.07 *	0.4228	430.83	2570.81
50	0.0	0.9549	294.34	686.21 *	0.9391	311.56	709.59
	0.5	0.9540	296.43	436.88	0.9299	258.89	407.06 *
	0.8	0.9541	299.74	415.48	0.9054	227.28	356.58 *
	1.2	0.9548	306.45	523.58 *	0.5604	185.14	2256.00
	1.5	0.9540	313.11	781.16 *	0.4731	153.53	4904.53
100	0.0	0.9674	310.59	425.87	0.9449	40.36	175.34 *
	0.5	0.9686	311.79	494.28	0.9372	40.44	243.36 *
	0.8	0.9691	313.61	2216.89	0.9225	40.49	1968.05 *
	1.2	0.9680	317.27	474.32	0.6840	40.55	296.84 *
	1.5	0.9684	320.92	428.18	0.5696	40.60	374.41 *
1,000	0.0	0.9892	829.98	1013.82	0.9511	26.46	268.12 *
	0.5	0.9902	830.30	1042.40	0.9484	26.45	284.46 *
	0.8	0.9902	830.78	920.52	0.9462	26.45	168.25 *
	1.2	0.9906	831.78	989.99	0.9378	26.45	236.15 *
	1.5	0.9901	832.79	1846.44	0.9107	26.44	1101.06 *
5,000	0.0	0.9949	3137.67	3215.11	0.9495	25.61	188.84 *
	0.5	0.9956	3137.90	3391.18	0.9502	25.62	354.56 *
	0.8	0.9950	3138.27	3338.91	0.9474	25.62	295.41 *
	1.2	0.9954	3139.02	4027.28	0.9486	25.63	989.07 *
	1.5	0.9960	3139.78	3207.18	0.9463	25.64	160.02 *

หมายเหตุ ตัวเลขหนา หมายถึง ค่า CP ที่ไม่ต่ำกว่า 0.95 อย่างมีนัยสำคัญ (CP $\geq$ 0.9475 ซึ่งเป็นขอบล่างทางเดียวที่ระดับนัยสำคัญ 0.05 จากการจำลอง 20,000 รอบ)

* หมายถึง ค่า AIS ที่มีค่าต่ำกว่า (สำหรับกรณีที่ไม่มี CP ค่าใดถึงเกณฑ์ ค่า AIS ที่ต่ำกว่าบ่งชี้เพียงวิธีที่ดีน้อยกว่าอีกวิธีเท่านั้น)

ถึง 5 ตารางที่ 1 แสดงผลการเปรียบเทียบประสิทธิภาพของช่วงพยากรณ์วิธี OLS และวิธี QR เมื่อความคลาดเคลื่อนมีการแจกแจงปกติ ซึ่งสอดคล้องกับข้อสมมติของวิธี OLS พบว่าวิธี OLS ให้ค่า CP ไม่ต่ำกว่า 0.95 อย่างมีนัยสำคัญในทุกเงื่อนไข ขณะที่วิธี QR ให้ค่า CP ต่ำกว่าระดับที่กำหนดอย่างชัดเจนเมื่อขนาดตัวอย่างเล็ก เช่น $n = 10$ มีค่า CP เพียง 0.72 ที่ตำแหน่งค่าพยากรณ์ $x_0 = 0$ และลดลงถึง 0.49 เมื่อ $x_0 = 1.5$ แล้วค่อยปรับเข้าใกล้ 0.95 เมื่อขนาดตัวอย่างเพิ่มขึ้นจนผ่านเกณฑ์ที่ $n = 5,000$ เมื่อพิจารณาค่า AIS พบว่าวิธี OLS ให้ค่าต่ำกว่าวิธี QR ในทุกเงื่อนไข แม้กระทั่งในกรณีที่ $n = 5,000$ ทั้งสองวิธีให้ความกว้างช่วง

และความครอบคลุมใกล้เคียงกัน ผลการวิจัยนี้สอดคล้องกับทฤษฎีที่ว่าเมื่อข้อสมมติเป็นจริง ตัวประมาณวิธี OLS มีประสิทธิภาพสูงสุด และสะท้อนให้เห็นว่าวิธี QR ค่าประมาณจะมีความแม่นยำน้อยกว่าเมื่อตัวอย่างมีขนาดเล็ก

ตารางที่ 2 แสดงผลการเปรียบเทียบประสิทธิภาพของช่วงพยากรณ์วิธี OLS และวิธี QR เมื่อความคลาดเคลื่อนมีการแจกแจงการแจกแจงเลขชี้กำลัง (การแจกแจงแบบเบ้) พบว่าวิธี QR เริ่มแสดงประสิทธิภาพเพิ่มขึ้นเมื่อขนาดตัวอย่างเพียงพอ จุดที่สังเกตได้ชัดคือความกว้างเฉลี่ยของช่วงวิธี QR ลดเข้าสู่ค่าประมาณ 18.3 ซึ่งแคบกว่าวิธี OLS ที่ประมาณ

ตารางที่ 5 ค่าความน่าจะเป็นครอบคลุม (CP) ค่าความกว้างช่วง (AW) และค่าคะแนนช่วงเฉลี่ย (AIS) ของช่วงพยากรณ์วิธี OLS และวิธี QR กรณี

 $\mathcal{E}_i \sim \text{Heteroscedasticity}$

n	x_0	CP_OLS	AW_OLS	AIS_OLS	CP_QR	AW_QR	AIS_QR
10	0.0	0.9937	24.93	25.28	0.8692	12.01	20.85 *
	0.5	0.9431	25.92	31.72 *	0.6482	12.02	62.52
	0.8	0.8981	27.40	41.08 *	0.5316	12.03	106.32
	1.2	0.8506	30.19	57.73 *	0.4310	12.04	175.18
	1.5	0.8192	32.76	72.21 *	0.3778	12.05	232.90
30	0.0	0.9982	22.13	22.18	0.9886	18.92	19.43 *
	0.5	0.9527	22.41	26.36 *	0.8817	18.91	30.77
	0.8	0.8930	22.84	35.18 *	0.7798	18.90	50.68
	1.2	0.8117	23.70	55.30 *	0.6578	18.90	90.19
	1.5	0.7578	24.54	74.21 *	0.5790	18.89	128.75
50	0.0	0.9990	21.71	21.74 *	0.9970	21.81	21.95
	0.5	0.9520	21.87	25.75 *	0.9336	21.84	28.06
	0.8	0.8930	22.13	34.63 *	0.8568	21.85	40.68
	1.2	0.8029	22.64	55.74 *	0.7420	21.87	72.17
	1.5	0.7442	23.14	76.19 *	0.6622	21.89	102.59
100	0.0	0.9994	21.44	21.47	0.9981	21.01	21.08 *
	0.5	0.9552	21.52	24.80 *	0.9394	21.01	25.93
	0.8	0.8900	21.65	34.39 *	0.8594	21.01	38.76
	1.2	0.8022	21.90	53.96 *	0.7546	21.02	63.55
	1.5	0.7311	22.16	76.56 *	0.6744	21.02	92.88
1,000	0.0	0.9996	21.23	21.24 *	0.9997	21.83	21.84
	0.5	0.9557	21.24	24.70 *	0.9601	21.83	24.92
	0.8	0.8930	21.25	33.28	0.9008	21.82	32.91 *
	1.2	0.7942	21.27	54.60	0.8025	21.82	53.76 *
	1.5	0.7232	21.30	76.51	0.7317	21.82	75.56 *
5,000	0.0	0.9997	21.20	21.21 *	0.9998	21.84	21.85
	0.5	0.9562	21.21	24.69 *	0.9617	21.84	24.83
	0.8	0.8925	21.21	33.25	0.9018	21.84	32.65 *
	1.2	0.7928	21.21	54.13	0.8047	21.84	52.40 *
	1.5	0.7189	21.22	76.53	0.7325	21.84	74.01 *

หมายเหตุ ตัวเลขหนา หมายถึง ค่า CP ที่ไม่ต่ำกว่า 0.95 อย่างมีนัยสำคัญ ($CP \geq 0.9475$ ซึ่งเป็นขอบล่างทางเดียวที่ระดับนัยสำคัญ 0.05 จากการจำลอง 20,000 รอบ)

* หมายถึง ค่า AIS ที่มีค่าต่ำกว่า (สำหรับกรณีที่ไม่ได้มี CP ค่าใดถึงเกณฑ์ ค่า AIS ที่ต่ำกว่าบ่งชี้เพียงวิธีที่ดีอย่างน้อยกว่าอีกวิธีเท่านั้น)

19.6 อย่างต่อเนื่อง เนื่องจากวิธี QR สร้างช่วงพยากรณ์แบบไม่สมมาตรเพื่อให้สอดคล้องกับความเบ้ของการแจกแจง ส่งผลให้ค่า AIS ของวิธี QR ต่ำกว่าวิธี OLS ที่ขนาดตัวอย่างตั้งแต่ $n = 30$ เป็นต้นไป ในตำแหน่งค่าพยากรณ์ใกล้ศูนย์กลาง ($x_0 = 0$) และครอบคลุมตำแหน่งที่ห่างจากศูนย์กลางมากขึ้น เมื่อขนาดตัวอย่างเพิ่มขึ้นพบว่าที่ $n = 1,000$ และ 5,000 วิธี QR ให้ค่า AIS ต่ำกว่าในทุกตำแหน่งค่าพยากรณ์ ที่สำคัญคือ เมื่อขนาดตัวอย่างใหญ่ ($n \geq 1,000$) วิธี OLS และวิธี QR ให้ค่า CP ผ่านเกณฑ์ทั้งสองวิธี แต่วิธี QR ให้ความกว้างที่แคบกว่าและค่า AIS ที่ต่ำ

กว่า แสดงว่าวิธี QR มีประสิทธิภาพมากกว่าวิธี OLS เมื่อความคลาดเคลื่อนมีลักษณะเบ้

ตารางที่ 3 แสดงผลการเปรียบเทียบประสิทธิภาพของช่วงพยากรณ์วิธี OLS และวิธี QR เมื่อความคลาดเคลื่อนมีการแจกแจง t ที่องศาอิสระเท่ากับ 5 ซึ่งเป็นกรณีที่ละเมิดข้อสมมติการแจกแจงปกติ (การแจกแจงหางหนา) พบว่าวิธี OLS ให้ค่า CP ที่ค่อนข้างคงที่และใกล้เคียงกับระดับเป้าหมาย ในทุกเงื่อนไขแม้ว่าข้อสมมติเรื่องการแจกแจงปกติจะถูกละเมิดก็ตาม ส่วนวิธี QR ในกรณีที่ขนาดตัวอย่างเล็ก ($n = 10$ และ 30) วิธี QR ให้ค่า CP ต่ำกว่าระดับเป้าหมายอย่างมาก (Undercoverage) โดยเฉพาะอย่างยิ่ง

เมื่อตำแหน่งค่าพยากรณ์ (x_0) อยู่ห่างจากค่าเฉลี่ย (เช่น $x_0 = 1.5$ ค่า CP ลดลงเหลือเพียง 0.4531 ที่ $n = 10$) อย่างไรก็ตาม เมื่อขนาดตัวอย่างเพิ่มขึ้น วิธี QR จะปรับตัวได้ดีขึ้น โดยค่า CP จะเข้าสู่ระดับเป้าหมาย และผ่านเกณฑ์อย่างมีนัยสำคัญ เมื่อขนาดตัวอย่างใหญ่ ($n = 5,000$) เมื่อพิจารณาค่า AIS พบว่า วิธี OLS ให้ค่าที่ต่ำกว่าวิธี QR ในทุกเงื่อนไขของขนาดตัวอย่างและตำแหน่งค่าพยากรณ์ แต่เมื่อขนาดตัวอย่างใหญ่ ($n \geq 1,000$) วิธี QR จะให้ค่า AIS ที่ใกล้เคียงกับวิธี OLS

ตารางที่ 4 แสดงผลการเปรียบเทียบประสิทธิภาพของช่วงพยากรณ์วิธี OLS และวิธี QR เมื่อความคลาดเคลื่อนมีการแจกแจงโคชี (การแจกแจงหางหนารุนแรง) ปรากฏความแตกต่างชัดเจนที่สุด เนื่องจากการแจกแจงโคชีไม่มีค่าความแปรปรวน ตัวประมาณ S^2 จึงไม่คงที่ ส่งผลให้ค่า AW ของวิธี OLS ขยายตัวอย่างมากตามขนาดตัวอย่างจนมีค่ามากกว่า 3,000 ที่ $n = 5,000$ และทำให้ค่า CP สูงเกิน 0.99 ซึ่งเป็นการครอบคลุมเกินความจำเป็น ในทางตรงข้าม วิธี QR ให้ค่า AW ที่คงที่มากกว่า (ประมาณ 25 – 26) ในทุกขนาดตัวอย่าง ส่งผลให้ค่า AIS ของวิธี QR ต่ำกว่าวิธี OLS เป็นอย่างมากเมื่อขนาดตัวอย่างใหญ่ เช่น 188.84 เทียบกับ 3,215.11 ที่ $n = 5,000$ ตำแหน่งค่าพยากรณ์ $x_0 = 0$ อย่างไรก็ตาม ที่ขนาดตัวอย่างเล็กถึงปานกลาง ($10 \leq n \leq 50$) ผลการวิจัยมีความผันผวนและไม่เป็นเอกภาพ เนื่องจากความไม่คงที่ของทั้งสองวิธีภายใต้เงื่อนไขของตำแหน่งค่าพยากรณ์ที่ห่างจากจุดศูนย์กลาง ($x_0 \geq 0$) โดยรวมผลการวิจัยในกรณีนี้สนับสนุนว่าวิธี QR มีความแม่นยำเหนือกว่าวิธี OLS เมื่อความคลาดเคลื่อนมีหางหนารุนแรง

ตารางที่ 5 แสดงผลการเปรียบเทียบประสิทธิภาพของช่วงพยากรณ์วิธี OLS และวิธี QR เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ (Heteroscedasticity) ผลการวิจัยพบว่าค่า CP ของวิธี OLS และวิธี QR แปรผันตามตำแหน่งค่าพยากรณ์ (x_0) อย่างมาก กล่าวคือ ณ ตำแหน่งใกล้จุดศูนย์กลาง ($x_0 = 0$ และ 0.5) วิธี OLS ให้ค่า CP ที่สูงและผ่านเกณฑ์ในเกือบทุกขนาดตัวอย่าง โดยมีแนวโน้มครอบคลุมเกินจำเป็น (Overcoverage) จนใกล้เคียง 0.99 ส่วนวิธี QR ให้ค่า CP เข้าสู่ระดับเป้าหมายช้ากว่า โดยผ่านเกณฑ์เมื่อขนาดตัวอย่างมากกว่า 30 ที่ตำแหน่ง $x_0 = 0$ แต่ที่ตำแหน่ง $x_0 = 0.5$ ต้องใช้ขนาดตัวอย่างใหญ่ขึ้น ($n \geq 1,000$) จึงผ่านเกณฑ์ ในทางตรงข้าม ณ ตำแหน่งค่า

พยากรณ์ที่ห่างออกไป ($x_0 = 0.8, 1.2$ และ 1.5) ค่า CP ของทั้งสองวิธีลดลงต่ำกว่าระดับเป้าหมายอย่างมาก (Undercoverage) ในทุกขนาดตัวอย่าง โดยเฉพาะ ณ ตำแหน่งที่ห่างจากศูนย์กลางมากที่สุด ($x_0 = 1.5$) ค่า CP ของวิธี OLS ลดลงเหลือประมาณ 0.7189 และของวิธี QR อยู่ที่ประมาณ 0.7325 แม้เพิ่มขนาดตัวอย่างเป็น $n = 5,000$ แล้วก็ตาม จะเห็นว่าวิธี QR ให้ค่า CP สูงกว่าวิธี OLS เพียงเล็กน้อย และทั้งสองวิธียังต่ำกว่าระดับเป้าหมาย 0.95 อยู่มาก แสดงให้เห็นว่าเมื่อข้อมูลมีความแปรปรวนไม่คงที่ การเพิ่มขนาดตัวอย่างไม่ได้ช่วยให้ค่าความครอบคลุมเข้าใกล้ระดับเป้าหมาย

สำหรับวิธี OLS สาเหตุเกิดจากการใช้ค่าประมาณความแปรปรวนของความคลาดเคลื่อน (S^2) ที่ประมาณจากข้อมูลทั้งหมดเพียงค่าเดียวในการกำหนดความกว้างของช่วงพยากรณ์ ทำให้ช่วงมีความกว้างคงที่เท่ากันทุกตำแหน่ง ขณะที่ความแปรปรวนจริงของข้อมูลในแบบจำลองนี้เปลี่ยนแปลงตามค่าของตัวแปรอิสระ ($|x|$) คือต่ำสุดที่ศูนย์กลางและสูงขึ้นเมื่อห่างออกไป ความไม่สอดคล้องระหว่างความกว้างช่วงคงที่กับความแปรปรวนจริงที่เปลี่ยนแปลงนี้ทำให้เกิดผลสองประการ คือ ประการแรก ณ ตำแหน่งใกล้ศูนย์กลางที่ความแปรปรวนจริงต่ำสุด ความกว้างคงที่กลับกว้างเกินไป จึงครอบคลุมเกินจำเป็นจนค่า CP เกือบเคียง 0.99 ประการที่สอง ณ ตำแหน่งที่ห่างออกไปซึ่งความแปรปรวนจริงสูงขึ้น ความกว้างคงที่กลับแคบเกินไป จึงครอบคลุมต่ำกว่าระดับเป้าหมาย และการเพิ่มขนาดตัวอย่างไม่ได้ช่วยแก้ปัญหาเนื่องจากค่า S^2 มีค่าเข้าสู่ความแปรปรวนเฉลี่ยรวมซึ่งเป็นค่าที่ไม่สอดคล้องกับความแปรปรวนจริง ณ ตำแหน่งส่วนใหญ่ จึงทำให้ปัญหานี้เป็นข้อจำกัดของตัวแบบ OLS ซึ่งต่างจากวิธี QR ที่ประมาณความกว้างของช่วงจากการกระจายของข้อมูลโดยตรงผ่านการถดถอยควอนไทล์ จึงสามารถปรับความกว้างให้สอดคล้องกับความแปรปรวนของข้อมูลได้ดีกว่าและให้ค่า CP ที่เสถียรมากกว่าเล็กน้อย อย่างไรก็ตาม เนื่องจากเส้นควอนไทล์ในงานวิจัยนี้ถูกกำหนดให้อยู่ในรูปแบบเชิงเส้น วิธี QR จึงยังปรับตามการเปลี่ยนแปลงของความแปรปรวนที่ขึ้นกับตำแหน่งได้ไม่เต็มที่ ส่งผลให้ค่าความครอบคลุม ณ ตำแหน่งที่ห่างจากศูนย์กลางยังคงต่ำกว่าระดับเป้าหมายเช่นเดียวกัน

เมื่อพิจารณาด้วยค่าคะแนนช่วงเฉลี่ย (AIS) ซึ่งเป็นเกณฑ์หลักในการตัดสินประสิทธิภาพ ที่ขนาดตัวอย่างใหญ่ ($n = 1,000$ และ $5,000$) วิธี QR ให้ค่า AIS ต่ำกว่า ณ ตำแหน่งที่ห่างจากศูนย์กลาง ($x_0 = 0.8, 1.2$ และ 1.5) ส่วนที่วิธี OLS ให้ค่า AIS ต่ำกว่า ณ ตำแหน่งใกล้ศูนย์กลาง ($x_0 = 0$ และ 0.5) ที่ช่วงพยากรณ์ของวิธี OLS ยังครอบคลุมได้เพียงพอและมีความกว้างน้อยกว่าวิธี QR เล็กน้อย ดังนั้นในกรณีความแปรปรวนไม่คงที่ ไม่มีวิธีใดเหนือกว่าในทุกตำแหน่ง แต่วิธี QR มีข้อได้เปรียบด้านประสิทธิภาพ ณ ตำแหน่งที่ห่างจากศูนย์กลางเมื่อขนาดตัวอย่างเพียงพอ ซึ่งเป็นตำแหน่งที่วิธี OLS มีปัญหาการครอบคลุมมากที่สุด

5. บทสรุป

งานวิจัยนี้เปรียบเทียบประสิทธิภาพของช่วงพยากรณ์ระหว่างวิธี OLS และวิธี QR สำหรับการวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย ผ่านการจำลองมอนติคาร์โลภายใต้การแจกแจงของความคลาดเคลื่อน 5 รูปแบบ ทั้งกรณีทดสอบคล้อยกับข้อสมมติและกรณีที่มีการละเมิด โดยพิจารณาจาก Coverage Probability, Average Width และ Average Interval Score ผลการศึกษาแสดงให้เห็นว่าประสิทธิภาพของช่วงพยากรณ์ขึ้นอยู่กับผลรวมของวิธีการประมาณ ลักษณะการแจกแจงของความคลาดเคลื่อน ขนาดตัวอย่าง และตำแหน่งที่ทำการพยากรณ์ ไม่มีวิธีใดเหนือกว่าในทุกสถานการณ์ เมื่อความคลาดเคลื่อนเป็นไปตามข้อสมมติของ OLS หรือมีลักษณะทางหนาในระดับปานกลาง วิธี OLS ให้ช่วงพยากรณ์ที่มีประสิทธิภาพดีกว่าหรือใกล้เคียงกับวิธี QR แม้ในกรณีตัวอย่างขนาดเล็ก ผลดังกล่าวสอดคล้องกับคุณสมบัติของการแจกแจงแบบ $t(5)$ ที่แม้จะมีหางหนากว่าการแจกแจงปกติ แต่ค่า quantile ที่ระดับ 0.975 เมื่อปรับมาตรฐานแล้วมีค่า $2.5706/1.2910 \approx 1.991$ ซึ่งใกล้เคียงกับค่า 1.960 ของการแจกแจงปกติมาตรฐาน (Casella & Berger, 2002) ความใกล้เคียงนี้อธิบายว่าเหตุใดช่วงพยากรณ์จาก OLS จึงยังคงให้ค่าความน่าจะเป็นครอบคลุมที่ไม่เบี่ยงเบนมากนักแม้ข้อสมมติถูกละเมิดในระดับปานกลาง เมื่อความคลาดเคลื่อนมีการแจกแจงเบ้หรือความแปรปรวนไม่คงที่ วิธี QR มีแนวโน้มให้ช่วงพยากรณ์ที่ดีกว่าเมื่อขนาดตัวอย่างเพียงพอ เนื่องจากประมาณค่าควอนไทล์แบบมีเงื่อนไขโดยตรงโดยไม่ต้องอาศัยข้อสมมติเกี่ยวกับโครงสร้างของความแปรปรวน (Koenker & Bassett, 1978)

อย่างไรก็ตาม ประสิทธิภาพของ QR ไม่ได้เพิ่มขึ้นตามระดับความรุนแรงของหางอย่างเห็นชัดส่วนเสมอไป ในกรณีการแจกแจงโคชี ที่ทั้งค่าเฉลี่ยและความแปรปรวนไม่มีค่าจำกัด ประสิทธิภาพของ OLS ลดลงอย่างชัดเจน และแม้ว่า QR จะมีข้อได้เปรียบในเชิงทฤษฎี แต่ความผันผวนสูงของข้อมูลทำให้การประมาณค่าควอนไทล์บริเวณปลายขอบยังคงมีความไม่เสถียร โดยเฉพาะเมื่อขนาดตัวอย่างมีขนาดเล็กถึงปานกลาง ปัญหานี้ยังรุนแรงขึ้นเมื่อทำการพยากรณ์นอกพิสัยของข้อมูล ซึ่งลดความน่าเชื่อถือของช่วงพยากรณ์จากทั้งสองวิธี

ผลการศึกษาหลายประเด็นสอดคล้องกับงานวิจัยที่ผ่านมา ข้อจำกัดของ QR ในตัวอย่างขนาดเล็กสอดคล้องกับ Yang และ Mo (1990) ข้อได้เปรียบของ QR ภายใต้การแจกแจงที่เบ้หรือหางหนาสนับสนุนแนวคิดของ Mak (2000) และข้อจำกัดของการพยากรณ์นอกพิสัยข้อมูลสอดคล้องกับ Stine (1985) และ Schmoyer (1992) ทั้งนี้ งานวิจัยนี้ขยายองค์ความรู้เดิมโดยเพิ่มการประเมินด้วย Average Interval Score ตามแนวทางของ Gneiting และ Raftery (2007) และทดสอบภายใต้การแจกแจงโคชี ซึ่งเป็นกรณีสุดขีดที่ยังขาดการศึกษาอย่างเป็นระบบในวรรณกรรมที่มีอยู่ในทางประยุกต์ การเลือกวิธีสร้างช่วงพยากรณ์ควรพิจารณา ลักษณะของข้อมูลและบริบทของการพยากรณ์ร่วมกัน หากความคลาดเคลื่อนมีการแจกแจงใกล้เคียงปกติหรือหางหนาระดับปานกลาง โดยเฉพาะเมื่อขนาดตัวอย่างมีขนาดเล็ก วิธี OLS ยังเป็นทางเลือกที่เหมาะสม แต่ถ้าข้อมูลมีลักษณะเบ้หรือความแปรปรวนไม่คงที่และมีขนาดตัวอย่างเพียงพอ วิธี QR ให้ช่วงพยากรณ์ที่มีความยืดหยุ่นมากกว่า อย่างไรก็ตาม ควรใช้ความระมัดระวังในทั้งสองกรณีเมื่อการแจกแจงมีหางหนารุนแรงหรือเมื่อทำการพยากรณ์นอกพิสัยข้อมูล

กิตติกรรมประกาศ

งานวิจัยนี้ได้รับทุนสนับสนุนจากคณะวิทยาศาสตร์ประยุกต์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ เลขที่สัญญา 690047

เอกสารอ้างอิง

- Cade, B. S., & Noon, B. R. (2003). A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment*, 1(8), 412–420.
- Casella, G., & Berger, R. L. (2002). *Statistical Inference*. 2nd ed., Duxbury.

- Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477), 359–378.
- Greene, W. H. (2018). *Econometric analysis*. 8th ed. Pearson.
- He, X. (1997). Quantile curves without crossing. *The American Statistician*, 51(2), 186–192.
- Koenker, R. (2005). *Quantile regression*. Cambridge University Press.
- Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1), 33–50.
- Koenker, R., & Hallock, K. F. (2001). Quantile regression. *Journal of Economic Perspectives*, 15(4), 143–156.
- Meinshausen, N. (2006). Quantile regression forests. *Journal of Machine Learning Research*, 7, 983–999.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis*. 5th ed., Wiley.
- Olive, D. J. (2007). Prediction intervals for regression models. *Computational Statistics & Data Analysis*, 51(6), 3115–3122. <https://doi.org/10.1016/j.csda.2006.02.006>
- R Core Team. (2025). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>.
- Schmoyer, R. L. (1992). Asymptotically valid prediction intervals for linear models. *Technometrics*, 34(4), 399–408.
- Stine, R.A. (1985). Bootstrap prediction intervals for regression. *Journal of the American Statistical Association*, 80(392), 1026–1031.
- Yang, M. C. K. & Mo, C. T. C. (1990). Coverage probability to the original data by prediction intervals in simple linear regression. *Journal of Statistical Computation and Simulation*, 36(4), 231–244.