การเปรียบเทียบวิธีการประมาณสำหรับการวิเคราะห์การถดถอยเชิงเส้นพหุ ที่ค่าความคลาดเคลื่อนมีความแปรปรวนไม่คงที่

พีรวัฒน์ เสรีวัฒนนุกูล[1], อนุภาพ สมบูรณ์สวัสดี[2]

[1]นักศึกษาปริญญาโท สาขาวิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

[2]อาจารย์ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

บทคัดย่อ

ปัญหาที่เกิดจากข้อตกลงตัวแบบถดถอยบางอย่าง เช่น ความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ ถือเป็นปัญหาหนึ่งที่สามารถพบได้ทั่วไปในงานวิจัยในการพยากรณ์หรือประมาณค่าตัวแปรที่สนใจ ซึ่งหากปัญหาดังกล่าวยังไม่ถูกกำจัดให้หมดไปก่อนการวิเคราะห์ข้อมูล จะทำให้ส่งผลถึงความถูกต้องในการพยากรณ์ ทำให้ได้ค่าพยากรณ์ที่ไม่ถูกต้องเท่าที่ควร วิธีการแก้ปัญหาความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ที่นักวิจัยใช้ ได้แก่ วิธีการแปลงของ Box และ Cox กับ วิธี Iterative Reweighted Least Square (IRWLS) ดังนั้นในงานวิจัยนี้มีวัตถุประสงค์ คือ เพื่อทำการศึกษาและเปรียบเทียบวิธีการพยากรณ์ตัวแปรตาม ในตัวแบบเชิงเส้นถดถอยพหุ เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ ในกรณีที่ขึ้นอยู่กับตัวแปรอิสระ และในกรณีที่ขึ้นอยู่กับค่าจริงตัวแปรตาม ที่มีระดับความแปรปรวนของตัวแปรอิสระ ภายใต้สถานการณ์ที่จำลองขึ้น ทั้งหมด 3 วิธี ได้แก่ วิธีกำลังสองน้อยที่สุด วิธีการแปลงของ Box และ Cox และ วิธี IRWLS ซึ่งข้อมูลที่ใช้ในการศึกษามาจากการจำลองค่าด้วยโปรแกรม R ทั้งหมด 5,000 ครั้งในแต่ละสถานการณ์ โดยเปรียบเทียบความถูกต้องของการพยากรณ์ด้วยการพิจารณาค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองโดยเฉลี่ย พบว่า วิธีการพยากรณ์ที่ดีที่สุดคือ วิธี IRWLS ส่วนวิธีการแปลงของ Box และ Cox จะมีความเหมาะสมในการพยากรณ์ตัวแปรตามมากกว่า วิธีกำลังสองน้อยที่สุด ก็ต่อเมื่อข้อมูลมีระดับความแปรปรวนของความคลาดเคลื่อนที่ขึ้นอยู่กับตัวแปรอิสระ และพารามิเตอร์ที่เกี่ยวข้องกับระดับความแปรปรวนของความคลาดเคลื่อน มีค่าสูง

คำสำคัญ: *วิธี Iterative Reweighted Least Square (IRWLS), วิธีการแปลงของ Box และ Cox, ความคลาดเคลื่อนมีความแปรปรวนไม่คงที่*

บทนำ

การวิเคราะห์การถดถอย (Regression Analysis) เป็นวิธีการวิเคราะห์ทางสถิติที่เกี่ยวข้องกับการสร้างตัวแบบทางคณิตศาสตร์ เพื่อแสดงความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระจำนวน 1 ตัวหรือมากกว่า 1 ตัวก็ได้ (วิชิต หล่อจีระชุณห์กุล และ จิราวัลย์ จิตรถเวช, 2548: 203) สำหรับการพยากรณ์หรือประมาณการตัวแปรด้วยการวิเคราะห์การถดถอยเชิงเส้น เมื่อทราบค่าของตัวแปรอิสระ จะมีข้อตกลงเบื้องต้นของความคลาดเคลื่อนแบบสุ่ม (random error) (ε) ซึ่งจะปรากฏในตัวแบบความถดถอย ได้แก่ การเป็นอิสระจากกัน (independence), การมีค่าความแปรปรวนคงที่ (homoscedasticity), การมีการแจกแจงปกติ (normality) ที่มีค่าเฉลี่ยเท่ากับศูนย์ (linearity) ทั้งนี้เพื่อให้การวิเคราะห์และอนุมานข้อมูลได้อย่างถูกต้อง (valid)

ปัญหาที่เกิดจากข้อกำหนดของตัวแบบบางอย่าง เช่น ความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ (heteroscedasticity) ลักษณะเช่นนี้จะทำให้เกิดค่าคลาดเคลื่อนบางกลุ่มมีค่าสูงมากๆหรือต่ำมากๆ ซึ่งจะมีผลต่อค่าสังเกตจากตัวแปรตาม (Y) ส่งผลให้ตัวประมาณของค่าพารามิเตอร์ (β) ยังคงมีคุณสมบัติความไม่เอนเอียง (unbiasedness) มีความคงเส้นคงวา (consistency) แต่ความแปรปรวนจะขาดประสิทธิภาพ (efficiency) กล่าวคือ ความแปรปรวนของตัวประมาณอาจมีค่าสูงเกินไป ทำให้ขาดคุณสมบัติของตัวประมาณการเชิงเส้นตรงที่ไม่เอนเอียงดีที่สุด (Best Linear Unbiased Estimator: BLUE) ทำให้ช่วงความเชื่อมั่นของพารามิเตอร์ที่คำนวณได้กว้างเกินไป เนื่องจากค่าคลาดเคลื่อนของสัมประสิทธิ์การถดถอยมีค่าสูงเกินไป ดังนั้นเพื่อให้ปัญหาความคลาดเคลื่อนมีความแปรปรวนไม่คงที่หมดไปก่อนการวิเคราะห์ข้อมูล จึงได้มีวิธีการพยากรณ์ตัวแปรตามจากการประมาณค่าสัมประสิทธิ์การถดถอย เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ ได้แก่ วิธีการแปลงของ Box และ Cox (Box-Cox Transformations) และ วิธี Iterative Reweighted Least Square (IRWLS)

ดังนั้นในงานวิจัยครั้งนี้จึงได้ทำการเปรียบเทียบประสิทธิภาพของการพยากรณ์ตัวแปรตามในตัวแบบเชิงเส้นถดถอยพหุ โดยใช้เกณฑ์ค่าเฉลี่ยค่าคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) เป็นเกณฑ์เปรียบเทียบการพยากรณ์ที่ได้จากวิธีทางสถิติทั้งหมด 3 วิธี ได้แก่ วิธีกำลังสองน้อยที่สุด (Ordinary Least Square: OLS), วิธีการแปลงของ Box และ Cox, และวิธี Iterative Reweighted Least Square (IRWLS) ภายใต้ข้อมูลที่มีค่าความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ทั้งหมด 2 รูปแบบ ได้แก่ ค่าความคลาดเคลื่อนที่ขึ้นอยู่กับตัวแปรอิสระ กับ ค่าจริงของตัวแปรตาม

วัตถุประสงค์

1. เพื่อศึกษาประสิทธิภาพของการพยากรณ์ตัวแปรตาม เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ในแต่ละสถานการณ์ จากในแต่ละวิธี ทั้งหมด 3 วิธี ได้แก่ วิธีกำลังสองน้อยที่สุด วิธีการแปลงของ Box และ Cox และวิธี Iterative Reweighted Least Square (IRWLS)

2. เพื่อเสนอแนวทางที่เหมาะสมในแต่ละสถานการณ์ของความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ในตัวแบบเชิงเส้นถดถอยพหุ มีการพยากรณ์ตัวแปรตามที่มีความถูกต้องมากขึ้น

วิธีการวิจัย

1. ตัวแปรอิสระที่ศึกษามีทั้งหมด 2 ตัว ซึ่งมีการแจกแจงปกติที่มีค่าเฉลี่ยเท่ากับศูนย์และความแปรปรวนทั้งหมด 3 กรณีดังต่อไปนี้

อัตราส่วนของ $\sigma_1^2 : \sigma_2^2$	σ_1^2	σ_2^2
1:2	300	600
1:1	450	450
2:1	600	300

โดยขนาดตัวอย่าง (n) ที่ศึกษามีทั้งหมด 3 ระดับ ได้แก่ เท่ากับ 25, 50 และ 100 และกำหนดให้ตัวแปรอิสระทุกตัวต่างไม่มีความสัมพันธ์กัน

2. สร้างข้อมูลค่าจริงของตัวแปรตาม โดยใช้ความสัมพันธ์เชิงเส้น คือ $\underset{\sim}{Y}' = X\underset{\sim}{\beta}$ โดยกำหนด $\underset{\sim}{\beta}$ เป็นค่าพารามิเตอร์ที่มีค่าคงที่ โดยที่ $\beta_0 = 500$, $\beta_1 = 1$ และ $\beta_2 = 1$

3. สร้างข้อมูลความคลาดเคลื่อน (ε_i) ที่เป็นอิสระซึ่งกันและกัน มีขนาดตัวอย่างเท่ากับ n และมีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับ ศูนย์ และ ค่าความแปรปรวน ที่แบ่งออกเป็น 2 รูปแบบ คือ

รูปแบบที่ 1: ค่าความแปรปรวน ที่ขึ้นอยู่กับค่าของตัวแปรอิสระตัวที่หนึ่งเพียงตัวเดียว (X_{1i})

$$\mathrm{Var}(\varepsilon_i) = \sigma_i^2 = \sigma^2 \left|X_{1i}\right|^{\delta}$$ โดยที่ $\sigma^2 > 0$, δ เป็นค่าพารามิเตอร์

รูปแบบที่ 2: ค่าความแปรปรวน ที่ขึ้นอยู่กับค่าของตัวแปรตาม (Y_i) เพิ่มขึ้นดังนี้

$$\mathrm{Var}(\varepsilon_i) = \sigma_i^2 = \sigma^2 \left|Y_i'\right|^{\delta}$$ โดยที่ $\sigma^2 > 0$, δ เป็นค่าพารามิเตอร์

โดย ค่าความแปรปรวนในเทอมที่เป็นค่าคงที่ของค่าความคลาดเคลื่อน (σ^2) จะศึกษาทั้งหมด 3 ระดับ ได้แก่ เท่ากับ 450, 900 และ 1800 และ ระดับความแตกต่างของความแปรปรวน กำหนดโดยค่าพารามิเตอร์ δ จำนวน 4 ค่า ได้แก่ 0, 0.1, 0.3 และ 0.5

4. สร้างข้อมูลตัวแปรตาม (Y_i) จากค่าตัวแปรอิสระ X_1 และ X_2 ที่ได้จากขั้นตอนที่ 1 และข้อมูลความคลาดเคลื่อน (ε_i) ที่ได้จากขั้นตอนที่ 2 โดยใช้สมการเชิงเส้นดังนี้

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i \; ; \; i = 1, 2, ..., n$$

โดยที่ β_0, β_1 และ β_2 เป็นค่าพารามิเตอร์ที่กำหนดขึ้นมามีค่าเป็น 500, 1 และ 1 ตามลำดับ

X_{1i} และ X_{2i} คือ ค่าของตัวแปรอิสระ และ ε_i คือ ค่าความคลาดเคลื่อน

5. คำนวณค่าสัมประสิทธิ์การถดถอย (regression coefficients) ด้วยวิธีดังต่อไปนี้

5.1 วิธีกำลังสองน้อยที่สุด (Ordinary Least Square: OLS)

จากการจำลองข้อมูลที่ได้ ตัวประมาณกำลังสองน้อยที่สุดที่ทำให้ผลบวกของค่าคลาดเคลื่อนกำลังสอง (Sum Squared Error: SSE) น้อยที่สุด คือ $\underset{\sim}{\hat{\beta}} = \left(X^T X\right)^{-1} X^T \underset{\sim}{Y}$

5.2 วิธีการแปลง Box-Cox (Box และ Cox, 1964)

คำนวณหาค่า λ ที่เหมาะสม ที่ทำให้ค่า log likelihood มากที่สุด ซึ่งค่า λ ดังกล่าวจะอยู่ในช่วง -3 ถึง 3 จากตัวแบบ

$$Y_i'' = \frac{Y_i^{\lambda} - 1}{\lambda} \qquad \text{เมื่อ } \lambda \neq 0$$

$$Y'' = \log(Y) \qquad \text{เมื่อ } \lambda = 0$$

และ หาค่าของสัมประสิทธิ์การถดถอย โดยใช้วิธีกำลังสองน้อยที่สุด

$$\hat{\beta}_1 = b_1 = \frac{\sum (X_i - \bar{X})(Y_i'' - \bar{Y}'')}{\sum (X_i - \bar{X})^2}$$

$$\hat{\beta}_0 = b_0 = \bar{Y}'' - b_1\bar{X}$$

5.3 วิธี Iterative Reweighted Least Square (IRWLS)

จากการจำลองข้อมูลที่ได้ โดยวิธี Iterative Reweighted Least Square (IRWLS) ให้ทำตามขั้นตอนดังต่อไปนี้ (Nelder และ Wedderburn, 1972) คือ

ขั้นตอนที่ 1 ประมาณค่า $\underset{\sim}{\mu}^{(r)}$ โดยที่ $\underset{\sim}{\mu}^{(r)} = g^{-1}(\underset{\sim}{x}^T\underset{\sim}{\beta}^{(r)})$

เมื่อ g เป็น Link function ที่ทำให้ $g(b'(\theta)) = \theta$

ขั้นตอนที่ 2 คำนวณค่า Working responses (Adjusted dependent variate) $\underset{\sim}{Z}^{(r)}$

$$\underset{\sim}{Z}^{(r)} = \underset{\sim}{\eta}^{(r)} + \frac{\partial\underset{\sim}{\eta}}{\partial\underset{\sim}{\mu}}(\underset{\sim}{Y} - \underset{\sim}{\mu}^{(r)})$$

ขั้นตอนที่ 3 คำนวณค่า Working weights โดยที่

$$W = \text{diag}\left[b''(\theta_i)\left(\frac{\partial\eta_i}{\partial\mu_i}\right)^2\right]^{-1} = \text{diag}\left[\text{Var}(\mu_i^{(r)})\left(\frac{\partial\eta_i}{\partial\mu_i}\right)^2\right]^{-1}$$

ขั้นตอนที่ 4 ด้วย Iteration ของ fisher scoring โดยที่

$$\underset{\sim}{\beta}^{(r+1)} = (X^TW^{(r)}X)^{-1}X^TW^{(r)}\underset{\sim}{Z}^{(r)}$$

ขั้นตอนที่ 5 ทำขั้นตอนที่ 1 ถึงขั้นตอนที่ 4 ไปเรื่อยๆ จนกระทั่งค่าสัมประสิทธิ์การถดถอย (β) ลู่เข้า (ค่าสัมบูรณ์ของผลต่างค่าสัมประสิทธิ์การถดถอยทุกค่าน้อยกว่า 0.001)

6. นำค่าสัมประสิทธิ์การถดถอยที่ได้ มาสร้างสมการเชิงเส้นพหุ เพื่อหาค่าพยากรณ์ตัวแปรตาม

7. คำนวณค่า AMSE ของแต่ละวิธีการประมาณค่าตัวแปรตาม โดยให้กระทำซ้ำในแต่ละสถานการณ์ทุกๆสถานการณ์ทั้งหมด 5,000 รอบ จากนั้นให้ทำการคำนวณค่าเฉลี่ยของค่าความคลาดเคลื่อนกำลังสองโดยเฉลี่ย ซึ่งเป็นเกณฑ์ที่ใช้ในการเปรียบเทียบวิธีการพยากรณ์ตัวแปรตาม เมื่อความแปรปรวนของค่าความคลาดเคลื่อนของตัวแบบมีค่าไม่คงที่ ซึ่งคำนวณได้จาก

$$\text{Average of MSE} = \frac{1}{5,000}\left[\frac{1}{n}\sum_{i=1}^{n}(Y_i' - \hat{Y}_i)^2\right]$$

โดยที่ $\hat{Y}_i$ คือ ค่าประมาณของตัวแปรตาม

Y_i' คือ ค่าจริงของตัวแปรตาม และ n คือ ขนาดตัวอย่าง

วิธีการพยากรณ์ตัวแปรตามที่ดีจะมีค่าค่าเฉลี่ยของค่าความคลาดเคลื่อนกำลังสองเฉลี่ยน้อย

ผลการวิจัย

ตารางที่ 1 แสดงค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองโดยเฉลี่ย (AMSE) เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ที่ขึ้นอยู่กับตัวแปรอิสระตัวที่หนึ่ง

σ_1^2	σ_2^2	n	σ^2	δ	AMSE		
					OLS	Box-Cox	IRWLS
300	600	50	900	0	55.17	58.54	55.17
				0.1	69.95	73.51	69.57
				0.3	119.68	123.63	113.17
				0.5	215.72	215.12	185.23
			1,800	0	107.65	112.80	107.65
				0.1	140.59	145.82	139.35
				0.3	242.12	246.01	227.81
				0.5	430.47	426.81	366.70
		100	450	0	13.53	14.92	13.53
				0.1	17.43	18.86	17.35
				0.3	30.57	32.03	29.13
				0.5	54.92	55.91	48.49
			900	0	27.04	28.68	27.04
				0.1	34.60	36.59	34.33
				0.3	60.80	62.76	57.37
				0.5	107.57	108.48	91.73
			1,800	0	54.31	56.81	54.31
				0.1	70.95	74.09	70.36
				0.3	119.32	122.31	111.60
				0.5	215.65	216.20	182.16

σ_1^2	σ_2^2	n	σ^2	δ	AMSE		
					OLS	Box-Cox	IRWLS
450	450	50	900	0	53.64	57.07	53.64
				0.1	71.26	74.99	70.79
				0.3	128.87	132.33	121.84
				0.5	232.00	231.75	200.68
			1,800	0	107.03	112.81	107.03
				0.1	138.66	144.52	137.59
				0.3	258.10	262.36	242.02
				0.5	473.67	471.00	399.57
		100	450	0	13.66	15.01	13.66
				0.1	18.07	19.47	17.99
				0.3	31.80	33.50	30.30
				0.5	59.96	61.05	52.41
			900	0	26.69	28.50	26.69
				0.1	35.35	37.31	35.11
				0.3	64.07	65.64	60.04
				0.5	118.51	118.73	101.23
			1,800	0	53.61	56.33	53.61
				0.1	71.67	74.75	71.28
				0.3	127.00	130.00	118.92
				0.5	236.82	235.36	201.05

σ_1^2	σ_2^2	n	σ^2	δ	AMSE		
					OLS	Box-Cox	IRWLS
600	300	50	900	0	52.56	55.72	52.56
				0.1	72.93	76.59	72.51
				0.3	132.35	135.94	124.93
				0.5	258.40	259.27	222.60
			1,800	0	106.66	111.61	106.66
				0.1	146.63	151.65	145.42
				0.3	269.33	273.60	251.65
				0.5	506.84	504.99	432.24
		100	450	0	13.56	14.89	13.56
				0.1	18.43	19.80	18.32
				0.3	33.27	34.75	31.72
				0.5	65.31	66.26	56.71
			900	0	26.77	28.56	26.77
				0.1	36.08	38.02	35.83
				0.3	66.97	68.84	63.18
				0.5	128.00	128.26	109.43
			1,800	0	54.26	56.92	54.26
				0.1	72.83	75.80	72.15
				0.3	130.88	134.02	122.57
				0.5	257.24	256.67	215.83

ตารางที่ 2 แสดงค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองโดยเฉลี่ย (AMSE) เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ที่ขึ้นอยู่กับค่าจริงของตัวแ

σ_1^2	σ_2^2	n	σ^2	δ	AMSE OLS	AMSE Box-Cox	AMSE IRWLS
300	600	50	450	0	53.79	57.97	53.79
				0.1	99.53	105.41	99.53
				0.3	343.51	356.41	343.44
				0.5	1210.68	1252.99	1208.39
			900	0	110.45	116.19	110.45
				0.1	198.57	206.01	198.57
				0.3	715.17	738.81	714.83
				0.5	2410.04	2505.67	2408.88
			1,800	0	214.64	223.42	214.64
				0.1	400.63	415.18	400.64
				0.3	1394.82	1438.08	1394.42
				0.5	4879.00	4944.99	4876.45
		100	450	0	26.77	29.32	26.77
				0.1	49.82	52.89	49.82
				0.3	173.00	180.46	173.03
				0.5	608.05	634.59	607.58
			900	0	54.38	57.78	54.38
				0.1	99.72	104.95	99.72
				0.3	353.17	366.50	353.04
				0.5	1194.57	1248.78	1193.75
			1,800	0	107.95	112.82	107.95
				0.1	201.28	209.61	201.29
				0.3	689.43	715.95	688.95
				0.5	2385.12	2453.42	2383.56
450	450	50	450	0	54.15	58.37	54.15
				0.1	101.43	106.58	101.43
				0.3	357.58	371.40	357.38
				0.5	1231.85	1279.47	1230.97
			900	0	109.45	115.10	109.45
				0.1	198.53	206.04	198.53
				0.3	684.89	709.97	684.52
				0.5	2451.13	2544.99	2449.52
			1,800	0	216.69	225.70	216.69
				0.1	402.62	416.46	402.61
				0.3	1376.39	1418.08	1376.70
				0.5	4793.76	4928.36	4972.85
		100	450	0	26.81	29.36	26.81
				0.1	50.60	53.72	50.60
				0.3	173.63	180.52	173.54
				0.5	592.03	618.42	591.63
			900	0	53.72	56.86	53.72
				0.1	100.08	104.48	100.08
				0.3	350.44	366.07	350.42
				0.5	1192.97	1247.22	1191.20
			1,800	0	106.01	111.33	106.01
				0.1	201.20	209.36	201.21
				0.3	688.16	720.03	687.97
				0.5	2457.87	2512.74	2456.50
600	300	50	450	0	54.11	58.17	54.11
				0.1	99.43	105.42	99.43
				0.3	342.62	356.67	342.52
				0.5	1198.88	1246.33	1197.74
			900	0	107.09	112.46	107.09
				0.1	201.46	209.86	201.45
				0.3	695.64	718.34	695.44
				0.5	2397.85	2487.60	2395.74
			1,800	0	215.92	224.81	215.92
				0.1	400.75	413.98	400.79
				0.3	1395.49	1439.86	1395.03
				0.5	4774.96	4912.14	4769.47
		100	450	0	26.76	29.38	26.76
				0.1	50.68	53.98	50.68
				0.3	173.08	181.50	173.07
				0.5	602.95	628.87	602.47
			900	0	55.11	58.23	55.11
				0.1	99.55	104.48	99.56
				0.3	345.47	359.07	345.51
				0.5	1187.52	1239.37	1187.34
			1,800	0	107.59	112.81	107.59
				0.1	198.27	206.90	198.26
				0.3	686.17	715.02	685.93
				0.5	2367.44	2441.26	2365.91
300	600	100	450	0	13.36	14.67	13.36
				0.1	24.49	26.10	24.49
				0.3	86.38	90.53	86.38
				0.5	299.93	313.10	299.73
			900	0	26.89	28.49	26.89
				0.1	49.12	51.98	49.12
				0.3	174.71	181.86	174.65
				0.5	600.03	625.25	599.22
			1,800	0	53.57	56.08	53.57
				0.1	101.98	106.42	101.98
				0.3	344.48	360.30	344.46
				0.5	1231.67	1262.54	1230.94
450	450	100	450	0	13.50	14.84	13.50
				0.1	25.26	26.95	25.26
				0.3	87.97	91.56	87.91
				0.5	300.82	313.42	300.33
			900	0	27.24	28.87	27.24
				0.1	51.15	53.29	51.15
				0.3	175.99	184.98	175.84
				0.5	595.08	620.12	594.56
			1,800	0	54.45	57.19	54.45
				0.1	99.94	103.96	99.93
				0.3	347.13	361.58	347.19
				0.5	1193.17	1220.99	1191.95
600	300	100	450	0	13.54	14.85	13.54
				0.1	25.06	26.75	25.05
				0.3	86.95	90.68	86.96
				0.5	303.35	317.21	302.88
			900	0	27.45	29.13	27.45
				0.1	49.62	52.06	49.62
				0.3	174.55	182.25	174.46
				0.5	593.81	621.73	593.64
			1,800	0	54.41	57.11	54.41
				0.1	100.87	105.02	100.87
				0.3	342.87	356.24	342.72
				0.5	1194.63	1231.37	1194.84

การอภิปรายผล

ในการวิเคราะห์ด้วยตัวแบบเชิงเส้นถดถอยพหุ เมื่อข้อมูลที่นำมาพยากรณ์ไม่ได้อยู่ในข้อตกลงของการวิเคราะห์การถดถอย ในกรณีที่ของความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ พบว่า

ปัจจัยที่ส่งผลต่อค่า AMSE ได้แก่ ปัจจัยแรก คือ ขนาดตัวอย่าง กล่าวคือ เมื่อขนาดตัวอย่าง (n) มากขึ้น แล้วจะทำให้ค่า AMSE น้อยลง ปัจจัยที่สอง คือ ระดับความแปรปรวนของความคลาดเคลื่อนในพารามิเตอร์ δ กล่าวคือ เมื่อเพิ่มขนาด δ มากขึ้น แล้วค่า AMSE เพิ่มขึ้น ปัจจัยต่อมา คือ ค่าความแปรปรวนของความคลาดเคลื่อนในส่วนที่เป็นเทอมค่าคงที่ (σ^2) ถ้ามีขนาดมากขึ้น แล้วจะทำให้ค่า AMSE เพิ่มขึ้น และ ที่อัตราส่วนของความแปรปรวนตัวแปรอิสระตัวที่หนึ่งต่อตัวที่สอง ที่ระดับ 1:2, 1:1 และ 2:1 จะทำให้ AMSE มีแนวโน้ม มากขึ้นตามลำดับ หรือกล่าวอีกนัยหนึ่งคือ เมื่อความแปรปรวนของตัวแปรอิสระตัวที่หนึ่ง มากขึ้น จะทำให้ค่า AMSE เพิ่มขึ้น

เมื่อเปรียบเทียบแต่ละวิธีการพยากรณ์หรือประมาณค่าตัวแปรตามโดยใช้เกณฑ์ AMSE พบว่า เมื่อค่าความแปรปรวนของความคลาดเคลื่อน ขึ้นอยู่กับ ตัวแปรอิสระตัวที่หนึ่ง (X_1) และระดับพารามิเตอร์ไม่เท่ากับศูนย์ วิธี IRWLS จะดีที่สุด รองลงมาคือวิธี OLS และ ดีน้อยที่สุดคือวิธีการแปลงของ Box และ Cox ในช่วงค่าพารามิเตอร์ δ และ σ^2 มีค่าน้อยๆ แต่เมื่อเพิ่มระดับ δ และ σ^2 มากขึ้น วิธีการแปลงของ Box และ Cox จึงเริ่มดีกว่า วิธี OLS แต่ เมื่อค่าความแปรปรวนของความคลาดเคลื่อน ขึ้นอยู่กับ ค่าจริงของตัวแปรตาม (Y) ค่า AMSE ของวิธี IRWLS น้อยที่สุดและน้อยกว่าของวิธี OLS อยู่ไม่มาก ส่วนวิธีการแปลงของ Box และ Cox ให้ค่า AMSE มากที่สุด

ข้อเสนอแนะ

<u>ด้านการนำไปใช้ประโยชน์</u>

เมื่อข้อมูลที่นำมาพยากรณ์ไม่ได้อยู่ในข้อตกลงของการวิเคราะห์การถดถอย ในกรณีที่ของความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ ในการวิเคราะห์ตัวแบบเชิงเส้นถดถอยพหุ วิธีการพยากรณ์ที่ควรใช้คือ วิธี IRWLS ในการประมาณค่าตัวแปรตาม ส่วนวิธีการแปลงของ Box-Cox จะมีความเหมาะสมในการพยากรณ์ตัวแปรตามก็ต่อเมื่อข้อมูลมีระดับความแปรปรวนของความคลาดเคลื่อนที่ขึ้นอยู่กับตัวแปรอิสระ และมีพารามิเตอร์ที่เกี่ยวข้องที่ทำให้ระดับความแปรปรวนของความคลาดเคลื่อนมีค่าสูง

<u>ด้านการศึกษาวิจัย</u>

ในงานวิจัยชิ้นนี้ ผู้วิจัยได้ศึกษาเฉพาะกรณีที่ตัวแปรอิสระมีการแจกแจงปกติ และ ไม่มีความสัมพันธ์กัน และ กรณีที่ตัวแปรตามมีความสัมพันธ์กันกับตัวแปรอิสระในรูปแบบเชิงเส้น สำหรับงานวิจัยครั้งต่อไปอาจทำการศึกษากรณีที่ข้อมูลตัวแปรอิสระมีรูปแบบการแจกแจงอื่นๆ หรือ ความสัมพันธ์ของตัวแปรตามกับตัวแปรอิสระ ที่ไม่ได้อยู่ในรูปแบบเชิงเส้น ซึ่งการที่ตัวแปรอิสระและตัวแปรตามมีลักษณะที่ต่างจากเดิม อาจจะทำให้ผลการวิจัยที่แตกต่างจากเดิม

เอกสารอ้างอิง

วิชิต หล่อจีระชุณห์กุล และ จิราวัลย์ จิตรถเวช. 2548. เทคนิคการพยากรณ์. พิมพ์ครั้งที่ 3. กรุงเทพมหานคร: โครงการส่งเสริมและพัฒนาเอกสารวิชาการ สถาบันบัณฑิตพัฒนบริหารศาสตร์

Box, G.E.P. and Cox, D.R. 1964. An Analysis of Transformation. Journal of the Royal Statistical Society , Series B. 26: 211-252.

Nelder, J. and R. Wedderburn. 1972. Generalized linear models. Journal of the Royal Statistical Society, Series A. 132: 370-384.