

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

การวิจัยครั้งนี้เป็นการวิจัยเชิงทดลอง เพื่อศึกษาเปรียบเทียบความถูกต้องของค่าพยากรณ์ในการวิเคราะห์ความถดถอยพหุคูณด้วยวิธีวิธีจักรนิกซ์และวิธีที่ใช้หลักการของโครงข่ายประสาทเทียมในกรณีที่เกิดความสัมพันธ์ระหว่างตัวแปรอิสระ ในสถานการณ์ต่างๆ ดังต่อไปนี้

1. การแจกแจงของความคลาดเคลื่อนมี 3 การแจกแจงคือ การแจกแจงตอกนอร์มอล การแจกแจงปกติป้อมปน และการแจกแจงปกติ
2. ขนาดตัวอย่างที่ใช้ในการวิจัยเท่ากับ 30, 50 และ 100
3. ระดับความสัมพันธ์ระหว่างตัวแปรอิสระเท่ากับ 0.1, 0.3, 0.5, 0.7, 0.9, และ 0.99 เมื่อจำนวนตัวแปรอิสระเท่ากับ 3 และระดับความสัมพันธ์ระหว่างตัวแปรอิสระเท่ากับ (0.1,0.1), (0.3, 0.3), (0.5, 0.5), (0.7, 0.7), (0.9, 0.9), และ (0.99, 0.99)¹ เมื่อจำนวนตัวแปรอิสระเท่ากับ 5

การวิจัยครั้งนี้ผู้วิจัยนำข้อมูลค่ายเทคนิคมอนิเตอร์ไอ โดยใช้โปรแกรมภาษาฟอร์แทรน 77 เพื่อสร้างข้อมูลตามสถานการณ์ที่กำหนด และคำนวณการวิเคราะห์ความถดถอยพหุคูณด้วยวิธีวิธีจักรนิกซ์ และใช้โปรแกรมสำเร็จรูป "Trajan 2.1 Shareware Neural Network Simulator, March 1997" โดยกระทำซ้ำ 400 รอบ ในแต่ละสถานการณ์

¹ กรณีที่จำนวนตัวแปรอิสระเท่ากับ 3 ระดับความสัมพันธ์เท่ากับ 0.1, 0.3, 0.5, 0.7, 0.9 และ 0.99 หมายถึงระดับความสัมพันธ์ระหว่างตัวแปร X_1 กับ X_2 , X_1 กับ X_3 และ X_2 กับ X_3

กรณีที่จำนวนตัวแปรอิสระเท่ากับ 5 ระดับความสัมพันธ์เท่ากับ (0.1,0.1), (0.3,0.3), (0.5,0.5), (0.7,0.7), (0.9,0.9) และ (0.99,0.99) ค่าแรกในวงเล็บ หมายถึง ระดับความสัมพันธ์ระหว่างตัวแปร X_1 กับ X_2 , X_1 กับ X_3 และ X_2 กับ X_3 ค่าหลังในวงเล็บ หมายถึง ระดับความสัมพันธ์ระหว่างตัวแปร X_4 กับ X_5

สรุปผลการวิจัย

1. ความคลาดเคลื่อนในการแจกแจงออกนอร์มอล

ผู้วิจัยศึกษาที่ระดับสัมประสิทธิ์การแปรผันเท่ากับ 22.64%, 59.15% และ 100.69% ตามเหตุที่ใช้ระดับสัมประสิทธิ์การแปรผันดังกล่าวเนื่องจาก ถ้าใช้ระดับสัมประสิทธิ์การแปรผันที่ต่ำกว่า 22.64% จะทำให้การแจกแจงเข้าสู่การแจกแจงปกติ รวมทั้งที่ไม่ใช้ระดับสัมประสิทธิ์การแปรผันที่สูงกว่า 100.69% เพราะว่าวิธี ANN มีแนวโน้มที่จะให้ความถูกต้องในการพยากรณ์ดีกว่าเหมือนเดิม โดยใช้พารามิเตอร์ต่างๆ มีค่าดังนี้ $\mu = 1$ และ $\sigma = 0.2264, 0.5915$ และ 1.0069 ตามลำดับ ซึ่งผลการวิจัยสรุปได้ว่าวิธีโครงข่ายประสาทเทียมให้ค่า AMSE ต่ำกว่าวิธีวิธีเรกเรชัน ยกเว้นในกรณีต่อไปนี้

ขนาดตัวอย่างเท่ากับ 30 ระดับสัมประสิทธิ์การแปรผันเท่ากับ 22.64% จำนวนตัวแปรอิสระเท่ากับ 3 ทุกระดับความสัมพันธ์ และจำนวนตัวแปรอิสระเท่ากับ 5 ที่ระดับความสัมพันธ์ (0.1, 0.1), (0.3, 0.3), (0.5, 0.5) และ (0.7, 0.7)

ขนาดตัวอย่างเท่ากับ 50 ระดับสัมประสิทธิ์การแปรผันเท่ากับ 22.64% จำนวนตัวแปรอิสระเท่ากับ 3 ที่ระดับความสัมพันธ์ 0.1, 0.3, 0.5 และ 0.7 และจำนวนตัวแปรอิสระเท่ากับ 5 ที่ระดับความสัมพันธ์ (0.1, 0.1), (0.3, 0.3) และ (0.5, 0.5)

และขนาดตัวอย่างเท่ากับ 100 ระดับสัมประสิทธิ์การแปรผันเท่ากับ 22.64% จำนวนตัวแปรอิสระเท่ากับ 3 ที่ระดับความสัมพันธ์ 0.1 วิธีวิธีเรกเรชันให้ค่า AMSE ต่ำกว่าวิธีโครงข่ายประสาทเทียม

ค่า AMSE ของวิธีวิธีเรกเรชันมีการเปลี่ยนแปลงดังนี้

- แปรผันตามระดับความสัมพันธ์ของตัวแปรอิสระ ระดับสัมประสิทธิ์การแปรผันของความคลาดเคลื่อน และจำนวนตัวแปรอิสระ เรียงตามลำดับความสำคัญจากมากไปน้อย
- แปรผกผันกับขนาดตัวอย่าง

และค่า AMSE ของวิธีโครงข่ายประสาทเทียมมีการเปลี่ยนแปลงดังนี้

- แปรผันตามระดับสัมประสิทธิ์การแปรผันของความคลาดเคลื่อน
- แปรผกผันกับขนาดตัวอย่าง จำนวนตัวแปรอิสระ และระดับความสัมพันธ์ของตัวแปรอิสระ โดยเรียงตามลำดับความสำคัญจากมากไปน้อย

ซึ่งวิธีโครงข่ายประสาทเทียมจะใช้ได้ดีกว่าวิธีรีดจ์รีเกรสชัน เมื่อขนาดตัวอย่าง ระดับสัมประสิทธิ์การแปรผันของตัวแปรอิสระ² จำนวนตัวแปรอิสระ และระดับความสัมพันธ์ของตัวแปรอิสระสูงขึ้น โดยเรียงตามลำดับความสำคัญจากมากไปน้อย และเพื่อความสะดวกในการตัดสินใจเลือกใช้วิธีการใดในการพยากรณ์ ผู้วิจัยได้แสดงผลการเปรียบเทียบวิธีการพยากรณ์ทั้งสองวิธีดังตารางที่ 5.1



สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

² จะเห็นว่าระดับสัมประสิทธิ์การแปรผันของตัวแปรอิสระ มีความสำคัญมากกว่า จำนวนตัวแปรอิสระ และระดับความสัมพันธ์ เนื่องจากค่า AMSE ที่ได้จากวิธี RR มีค่าสูงขึ้นมากกว่าวิธี ANN เมื่อระดับสัมประสิทธิ์การแปรผันสูงขึ้น จึงส่งผลให้วิธี ANN ใช้ได้ดีกว่าวิธี RR เมื่อระดับสัมประสิทธิ์การแปรผันสูงขึ้น นอกจากนี้แล้วเมื่อพิจารณาการเปลี่ยนแปลงค่า DIFF ของตัวแปรทั้งสาม จะได้ว่ามีค่าลดลงเมื่อระดับสัมประสิทธิ์การแปรผัน จำนวนตัวแปรอิสระและระดับความสัมพันธ์ มีค่าสูงขึ้นตามลำดับ

ตารางที่ 5.1 สรุปผลการเปรียบเทียบค่าพยากรณ์ของวิธีวิธีจักรสขันและวิธีโครงข่ายประสาทเทียม เมื่อความคลาดเคลื่อนมีการแจกแจงดอกรนอร์มอล

# X	CORR	n = 30			n = 50			n = 100		
		$\sigma = 0.2264$	$\sigma = 0.5915$	$\sigma = 1.0069$	$\sigma = 0.2264$	$\sigma = 0.5915$	$\sigma = 1.0069$	$\sigma = 0.2264$	$\sigma = 0.5915$	$\sigma = 1.0069$
3	0.1	*	✓	✓	*	✓	✓	*	✓	✓
	0.3	*	✓	✓	*	✓	✓	✓	✓	✓
	0.5	*	✓	✓	*	✓	✓	✓	✓	✓
	0.7	*	✓	✓	*	✓	✓	✓	✓	✓
	0.9	*	✓	✓	✓	✓	✓	✓	✓	✓
	0.99	*	✓	✓	✓	✓	✓	✓	✓	✓
5	(0.1, 0.1)	*	✓	✓	*	✓	✓	✓	✓	✓
	(0.3, 0.3)	*	✓	✓	*	✓	✓	✓	✓	✓
	(0.5, 0.5)	*	✓	✓	*	✓	✓	✓	✓	✓
	(0.7, 0.7)	*	✓	✓	✓	✓	✓	✓	✓	✓
	(0.9, 0.9)	✓	✓	✓	✓	✓	✓	✓	✓	✓
	(0.99, 0.99)	✓	✓	✓	✓	✓	✓	✓	✓	✓

หมายเหตุ ✓ หมายถึง วิธีโครงข่ายประสาทเทียมให้ค่า AMSE ต่ำกว่า

* หมายถึง วิธีวิธีจักรสขันให้ค่า AMSE ต่ำกว่า

X หมายถึง จำนวนตัวแปรอิสระ

CORR หมายถึง ระดับความสัมพันธ์ของตัวแปรอิสระ

และ σ หมายถึง ส่วนเบี่ยงเบนมาตรฐาน

2. ความคลาดเคลื่อนมีการแจกแจงปกติปลอมปน

ผู้วิจัยศึกษาที่สเกลแฟคเตอร์เท่ากับ 3 และ 10 เปอร์เซ็นต์การปลอมปนเท่ากับ 5 และ 10 เมื่อระดับสัมประสิทธิ์การแปรผันเท่ากับ 10%, 30% และ 50% สาเหตุที่ใช้ระดับสัมประสิทธิ์การแปรผันดังกล่าวนี้เหมือนกับมีการแจกแจงปกติ เนื่องจากมีการแจกแจงปกติปลอมปนยังคงมีลักษณะของการแจกแจงปกติเหมือนเดิม และใช้พารามิเตอร์ $\mu = 1$ และ $\sigma = 0.1, 0.3$ และ 0.5 ตามลำดับ ซึ่งผลการวิจัยสรุปได้ว่า ส่วนใหญ่วิธีรีดจีเรชันให้ค่า AMSE ต่ำกว่าวิธีโครงข่ายประสาทเทียม โดยที่แนวโน้มของค่า AMSE ของวิธีโครงข่ายประสาทเทียมจะต่ำกว่าวิธีรีดจีเรชันเมื่อขนาดตัวอย่าง ระดับสัมประสิทธิ์การแปรผัน จำนวนตัวแปรอิสระ ระดับความสัมพันธ์ของตัวแปรอิสระ สเกลแฟคเตอร์ และเปอร์เซ็นต์การปลอมปนสูงขึ้น โดยเรียงตามลำดับความสำคัญจากมากไปน้อย

ค่า AMSE ของวิธีรีดจีเรชันมีการเปลี่ยนแปลงดังนี้

- แปรผันตามระดับความสัมพันธ์ของตัวแปรอิสระ ระดับสัมประสิทธิ์การแปรผันของความคลาดเคลื่อน จำนวนตัวแปรอิสระ สเกลแฟคเตอร์ และเปอร์เซ็นต์การปลอมปน เรียงตามลำดับความสำคัญจากมากไปน้อย
- แปรผกผันกับขนาดตัวอย่าง

และค่า AMSE ของวิธีโครงข่ายประสาทเทียมมีการเปลี่ยนแปลงดังนี้

- แปรผันตามระดับสัมประสิทธิ์การแปรผันของความคลาดเคลื่อน สเกลแฟคเตอร์ และเปอร์เซ็นต์การปลอมปน เรียงตามลำดับความสำคัญจากมากไปน้อย
- แปรผกผันกับขนาดตัวอย่าง จำนวนตัวแปรอิสระ และระดับความสัมพันธ์ของตัวแปรอิสระ เรียงตามลำดับความสำคัญจากมากไปน้อย

ซึ่งวิธีโครงข่ายประสาทเทียมจะใช้ได้ดีกว่าวิธีรีดจีเรชัน เมื่อขนาดตัวอย่าง ระดับสัมประสิทธิ์การแปรผันของตัวแปรอิสระ³ จำนวนตัวแปรอิสระ ระดับความสัมพันธ์ของตัวแปรอิสระ สเกลแฟคเตอร์ และเปอร์เซ็นต์การปลอมปนสูงขึ้น โดยเรียงตามลำดับความสำคัญจากมากไปน้อย และเพื่อความสะดวกในการตัดสินใจเลือกใช้วิธีการใดในการพยากรณ์ ผู้วิจัยได้แสดงผลการเปรียบเทียบวิธีการพยากรณ์ทั้งสองวิธีดังตารางที่ 5.2

³ จะเห็นว่าระดับสัมประสิทธิ์การแปรผันของตัวแปรอิสระ มีความสำคัญมากกว่าตัวแปรอื่นๆ ยกเว้นขนาดตัวอย่าง เนื่องจากค่า AMSE ที่ได้จากวิธี RR มีค่าสูงขึ้นมากกว่าวิธี ANN เมื่อระดับสัมประสิทธิ์การแปรผันสูงขึ้น จึงส่งผลให้วิธี ANN ใช้ได้ดีกว่าวิธี RR เมื่อระดับสัมประสิทธิ์การแปรผันสูงขึ้น นอกจากนี้แล้วเมื่อพิจารณาการเปลี่ยนแปลงค่า DIFP จะได้ว่ามีค่าลดลงเมื่อระดับสัมประสิทธิ์การแปรผัน จำนวนตัวแปรอิสระ ระดับความสัมพันธ์ สเกลแฟคเตอร์ และเปอร์เซ็นต์การปลอมปน มีค่าสูงขึ้นตามลำดับ

ตารางที่ 5.2 สรุปผลการเปรียบเทียบค่าพยากรณ์ของวิธีคิดจีเรตชันและวิธีโครงข่ายประสาทเทียม เมื่อความคลาดเคลื่อนมีการแจกแจงปกติปดอมน

ขนาด ตัวอย่าง	สเกล แพคเตอร์	เปอร์เซ็นต์ การปดอมน	ส่วนเบี่ยงเบน มาตรฐาน	# X CORR	3						5						
					0.1	0.3	0.5	0.7	0.9	0.99	(0.1, 0.1)	(0.3, 0.3)	(0.5, 0.5)	(0.7, 0.7)	(0.9, 0.9)	(0.99, 0.99)	
30	3	5	0.1		*	*	*	*	*	*	*	*	*	*	*	*	
			0.3		*	*	*	*	*	*	*	*	*	*	*	*	*
			0.5		*	*	*	*	*	*	*	*	*	*	*	*	*
		10	0.1		*	*	*	*	*	*	*	*	*	*	*	*	*
			0.3		*	*	*	*	*	*	*	*	*	*	*	*	✓
			0.5		*	*	*	*	*	*	✓	*	*	*	*	*	*
	10	5	0.1		*	*	*	*	*	*	*	*	*	*	*	*	
			0.3		*	*	*	*	*	*	*	*	*	*	*	*	*
			0.5		*	*	*	*	*	*	✓	*	*	*	*	*	✓
		10	0.1		*	*	*	*	*	*	*	*	*	*	*	*	✓
			0.3		*	*	*	*	*	✓	✓	*	*	*	✓	✓	✓
			0.5		*	*	*	✓	✓	✓	*	✓	✓	✓	✓	✓	✓

หมายเหตุ ✓ หมายถึง วิธีโครงข่ายประสาทเทียมให้ค่า AMSE ต่ำกว่า

* หมายถึง วิธีรีดจีเรตชันให้ค่า AMSE ต่ำกว่า

X หมายถึง จำนวนตัวแปรอิสระ

และ CORR หมายถึง ระดับความสัมพันธ์ของตัวแปรอิสระ

ตารางที่ 5.2 (ต่อ)

ขนาดตัวอย่าง	สมมติแฟกเตอร์	เปอร์เซ็นต์การปิดอมปน	ส่วนเบี่ยงเบนมาตรฐาน	# X CORR	3						5						
					0.1	0.3	0.5	0.7	0.9	0.99	(0.1, 0.1)	(0.3, 0.3)	(0.5, 0.5)	(0.7, 0.7)	(0.9, 0.9)	(0.99, 0.99)	
50	3	5	0.1		*	*	*	*	*	*	✓	*	*	*	*	*	✓
			0.3		*	*	*	*	*	*	✓	*	*	*	*	✓	✓
			0.5		*	*	*	*	✓	✓	*	*	*	*	✓	✓	
		10	0.1		*	*	*	*	*	*	✓	*	*	*	*	*	✓
			0.3		*	*	*	*	✓	✓	*	*	*	*	✓	✓	
			0.5		*	*	*	*	✓	✓	*	*	*	✓	✓	✓	
	10	5	0.1		*	*	*	*	*	*	✓	*	*	*	*	✓	✓
			0.3		*	*	*	*	✓	✓	*	*	*	✓	✓	✓	
			0.5		*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
		10	0.1		*	*	*	*	*	*	✓	*	*	*	*	✓	✓
			0.3		*	*	*	✓	✓	✓	✓	✓	✓	✓	✓	✓	
			0.5		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	

หมายเหตุ ✓ หมายถึง วิธีโครงข่ายประสาทเทียมให้ค่า AMSE ต่ำกว่า

* หมายถึง วิธีรีดจี้เกรสชันให้ค่า AMSE ต่ำกว่า

X หมายถึง จำนวนตัวแปรอิสระ

และ CORR หมายถึง ระดับความสัมพันธ์ของตัวแปรอิสระ

ตารางที่ 5.2 (ต่อ)

ขนาดตัวอย่าง	สเกลแฟคเตอร์	เปอร์เซ็นต์การปลอมปน	ส่วนเบี่ยงเบนมาตรฐาน	# X CORR	3						5					
					0.1	0.3	0.5	0.7	0.9	0.99	(0.1, 0.1)	(0.3, 0.3)	(0.5, 0.5)	(0.7, 0.7)	(0.9, 0.9)	(0.99, 0.99)
100	3	5	0.1		*	*	*	*	*	✓	*	*	*	*	✓	✓
			0.3		*	*	*	*	✓	✓	*	*	✓	✓	✓	✓
			0.5		*	*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		10	0.1		*	*	*	*	*	✓	*	*	*	*	✓	✓
			0.3		*	*	*	*	✓	✓	*	✓	✓	✓	✓	✓
			0.5		*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	10	5	0.1		*	*	*	*	✓	✓	*	*	✓	✓	✓	✓
			0.3		*	*	*	✓	✓	✓	✓	✓	✓	✓	✓	✓
			0.5		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		10	0.1		*	*	*	*	✓	✓	*	*	✓	✓	✓	✓
			0.3		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
			0.5		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

หมายเหตุ ✓ หมายถึง วิธีโครงข่ายประสาทเทียมให้ค่า AMSE ต่ำกว่า

* หมายถึง วิธีรีดจีเรทชันให้ค่า AMSE ต่ำกว่า

X หมายถึง จำนวนตัวแปรอิสระ

และ CORR หมายถึง ระดับความสัมพันธ์ของตัวแปรอิสระ

8. ความคลาดเคลื่อนมีการแจกแจงปกติ

ผู้วิจัยศึกษาที่ระดับสัมประสิทธิ์การแปรผันเท่ากับ 10%, 30% และ 50% สาเหตุที่ใช้ระดับสัมประสิทธิ์การแปรผันดังกล่าว เพื่อศึกษาว่าที่ระดับสัมประสิทธิ์การแปรผันช่วงต่างๆ กัน ให้ผลแตกต่างกันหรือไม่ นอกจากนี้แล้วการที่ข้อมูลตัวแปรตามมีการกระจายสูง การที่ตัวแปรอิสระจะพหุอาการ ตัวแปรตามได้ค่านั้นก็เป็น ได้ยากด้วย โดยเฉพาะในกรณีนี้ซึ่งตัวแปรอิสระมีพหุสัมพันธ์กัน และใช้พารามิเตอร์ $\mu = 1$ และ $\sigma = 0.1, 0.3$ และ 0.5 ตามลำดับ ซึ่งผลการวิจัยสรุปได้ว่าวิธีเรกิสรีเกรสชันให้ค่า AMSE ต่ำกว่าวิธีโครงข่ายประสาทเทียม ยกเว้นในกรณีดังต่อไปนี้

ขนาดตัวอย่างเท่ากับ 50

ระดับสัมประสิทธิ์การแปรผันเท่ากับ 10% จำนวนตัวแปรอิสระเท่ากับ 3 ที่ระดับความสัมพันธ์ 0.99 และจำนวนแปรอิสระเท่ากับ 5 ที่ระดับความสัมพันธ์ (0.99, 0.99)

ระดับสัมประสิทธิ์การแปรผันเท่ากับ 30% จำนวนตัวแปรอิสระเท่ากับ 3 ที่ระดับความสัมพันธ์ 0.99 และจำนวนแปรอิสระเท่ากับ 5 ที่ระดับความสัมพันธ์ (0.99, 0.99)

ระดับสัมประสิทธิ์การแปรผันเท่ากับ 50% จำนวนตัวแปรอิสระเท่ากับ 3 ที่ระดับความสัมพันธ์ 0.99 และจำนวนแปรอิสระเท่ากับ 5 ที่ระดับความสัมพันธ์ (0.9, 0.9) และ (0.99, 0.99)

และขนาดตัวอย่างเท่ากับ 100

ระดับสัมประสิทธิ์การแปรผันเท่ากับ 10% จำนวนตัวแปรอิสระเท่ากับ 3 ที่ระดับความสัมพันธ์ 0.99 และจำนวนแปรอิสระเท่ากับ 5 ที่ระดับความสัมพันธ์ (0.99, 0.99)

ระดับสัมประสิทธิ์การแปรผันเท่ากับ 30% จำนวนตัวแปรอิสระเท่ากับ 3 ที่ระดับความสัมพันธ์ 0.9 และ 0.99 และจำนวนแปรอิสระเท่ากับ 5 ที่ระดับความสัมพันธ์ (0.9, 0.9) และ (0.99, 0.99)

ระดับสัมประสิทธิ์การแปรผันเท่ากับ 50% จำนวนตัวแปรอิสระเท่ากับ 3 ที่ระดับความสัมพันธ์ 0.7, 0.9 และ 0.99 และจำนวนแปรอิสระเท่ากับ 5 ที่ระดับความสัมพันธ์ (0.5, 0.5), (0.7, 0.7), (0.9, 0.9) และ (0.99, 0.99) วิธีโครงข่ายประสาทเทียมให้ค่า AMSE ต่ำกว่าวิธีเรกิสรีเกรสชัน

ค่า AMSE ของวิธีเรกิสรีเกรสชันมีการเปลี่ยนแปลงดังนี้

- แปรผันตามระดับความสัมพันธ์ของตัวแปรอิสระ ระดับสัมประสิทธิ์การแปรผันของความคลาดเคลื่อน และจำนวนตัวแปรอิสระ เรียงตามลำดับความสำคัญจากมากไปน้อย
- แปรผกผันกับขนาดตัวอย่าง

และค่า AMSE ของโครงข่ายประสาทเทียมมีการเปลี่ยนแปลงดังนี้

- แปรผันตามระดับสัมประสิทธิ์การแปรผันของความคลาดเคลื่อน
- แปรผกผันกับขนาดตัวอย่าง จำนวนตัวแปรอิสระ และระดับความสัมพันธ์ของตัวแปรอิสระ เรียงตามลำดับความสำคัญจากมากไปน้อย

ซึ่งวิธีโครงข่ายประสาทเทียมจะใช้ได้ดีกว่าวิธีวิธีเกรสชัน เมื่อขนาดตัวอย่าง ระดับสัมประสิทธิ์การแปรผันของตัวแปรอิสระ⁴ จำนวนตัวแปรอิสระ และระดับความสัมพันธ์ของตัวแปรอิสระสูงขึ้น โดยเรียงตามลำดับความสำคัญจากมากไปน้อย และเพื่อความสะดวกในการตัดสินใจเลือกใช้วิธีการใดในการพยากรณ์ ผู้วิจัยได้แสดงผลการเปรียบเทียบวิธีการพยากรณ์ทั้งสองวิธีดังตารางที่ 5.3



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

⁴ จะเห็นว่าระดับสัมประสิทธิ์การแปรผันของตัวแปรอิสระ มีความสำคัญมากกว่า จำนวนตัวแปรอิสระ และระดับความสัมพันธ์ เนื่องจากค่า AMSE ที่ได้จากวิธี RR มีค่าสูงขึ้นมากกว่าวิธี ANN เมื่อระดับสัมประสิทธิ์การแปรผันสูงขึ้น จึงส่งผลให้วิธี ANN ใช้ได้ดีกว่าวิธี RR เมื่อระดับสัมประสิทธิ์การแปรผันสูงขึ้น นอกจากนี้แล้วเมื่อพิจารณาการเปลี่ยนแปลงค่า DIFP ของตัวแปรทั้งสาม จะได้ว่ามีค่าลดลงเมื่อระดับสัมประสิทธิ์การแปรผัน จำนวนตัวแปรอิสระและระดับความสัมพันธ์ มีค่าสูงขึ้นตามลำดับ

ตารางที่ 5.8 สรุปผลการเปรียบเทียบค่าพยากรณ์ของวิธีวิธีจักรสถันและวิธีโครงข่ายประสาทเทียม เมื่อความคลาดเคลื่อนมีการแจกแจงปกติ

# X	CORR	n = 30			n = 50			n = 100		
		$\sigma = 0.1$	$\sigma = 0.3$	$\sigma = 0.5$	$\sigma = 0.1$	$\sigma = 0.3$	$\sigma = 0.5$	$\sigma = 0.1$	$\sigma = 0.3$	$\sigma = 0.5$
3	0.1	*	*	*	*	*	*	*	*	*
	0.3	*	*	*	*	*	*	*	*	*
	0.5	*	*	*	*	*	*	*	*	*
	0.7	*	*	*	*	*	*	*	*	✓
	0.9	*	*	*	*	*	*	*	✓	✓
	0.99	*	*	*	✓	✓	✓	✓	✓	✓
5	(0.1,0.1)	*	*	*	*	*	*	*	*	*
	(0.3, 0.3)	*	*	*	*	*	*	*	*	*
	(0.5, 0.5)	*	*	*	*	*	*	*	*	✓
	(0.7, 0.7)	*	*	*	*	*	*	*	*	✓
	(0.9, 0.9)	*	*	*	*	*	✓	*	✓	✓
	(0.99, 0.99)	*	*	*	✓	✓	✓	✓	✓	✓

หมายเหตุ ✓ หมายถึง วิธีโครงข่ายประสาทเทียมให้ค่า AMSE ต่ำกว่า

* หมายถึง วิธีวิธีจักรสถันให้ค่า AMSE ต่ำกว่า

X หมายถึง จำนวนตัวแปรอิสระ

CORR หมายถึง ระดับความสัมพันธ์ของตัวแปรอิสระ

และ σ หมายถึง ส่วนเบี่ยงเบนมาตรฐาน

อภิปรายผล

การวิจัยครั้งนี้ผู้วิจัยต้องการเปรียบเทียบค่าความถูกต้องของการพยากรณ์ตัวแปรตามในการวิเคราะห์การถดถอยพหุคูณ โดยพิจารณาความถูกต้องจากค่า AMSE และค่า DIFF ของวิธีการพยากรณ์ทั้งสองวิธี โดยค่า AMSE ของวิธีโครงข่ายประสาทเทียมแปรผันตามระดับสัมประสิทธิ์การแปรผันของความคลาดเคลื่อน สเกดแฟคเตอร์ และเปอร์เซ็นต์การปลอมปน และแปรผกผันกับขนาดตัวอย่าง จำนวนตัวแปรอิสระ และระดับความสัมพันธ์ของตัวแปรอิสระ ส่วนค่า DIFF แปรผกผันกับระดับความสัมพันธ์ของตัวแปรอิสระ ขนาดตัวอย่าง ระดับสัมประสิทธิ์การแปรผันของความคลาดเคลื่อน จำนวนตัวแปรอิสระ สเกดแฟคเตอร์ และเปอร์เซ็นต์การปลอมปนสูงขึ้นไป และค่า AMSE ของวิธีวิธีจักรตชันแปรผันตามระดับความสัมพันธ์ของตัวแปรอิสระ ระดับสัมประสิทธิ์การแปรผันของความคลาดเคลื่อน จำนวนตัวแปรอิสระ สเกดแฟคเตอร์ และเปอร์เซ็นต์การปลอมปน และแปรผกผันกับขนาดตัวอย่าง ส่วนค่า DIFF แปรผันตามระดับความสัมพันธ์ของตัวแปรอิสระ ขนาดตัวอย่าง ระดับสัมประสิทธิ์การแปรผันของความคลาดเคลื่อน จำนวนตัวแปรอิสระ สเกดแฟคเตอร์ และเปอร์เซ็นต์การปลอมปน

ปัจจัยที่ศึกษาทุกปัจจัยมีผลกับวิธี โครงข่ายประสาทเทียมเรียงลำดับความสำคัญจากมากไปน้อย คือการแจกแจงของความคลาดเคลื่อน ขนาดตัวอย่าง ระดับสัมประสิทธิ์การแปรผันของความคลาดเคลื่อน จำนวนตัวแปรอิสระ สเกดแฟคเตอร์ เปอร์เซ็นต์การปลอมปน และระดับความสัมพันธ์ของตัวแปรอิสระ ตามลำดับ และปัจจัยที่ศึกษาทุกปัจจัยมีผลกับวิธีวิธีจักรตชันเรียงลำดับความสำคัญจากมากไปน้อย คือการแจกแจงของความคลาดเคลื่อน ขนาดตัวอย่าง ระดับความสัมพันธ์ของตัวแปรอิสระ ระดับสัมประสิทธิ์การแปรผันของความคลาดเคลื่อน จำนวนตัวแปรอิสระ สเกดแฟคเตอร์ และเปอร์เซ็นต์การปลอมปน ตามลำดับ

จากการวิจัยครั้งนี้จะเห็นได้ว่าวิธี โครงข่ายประสาทเทียม จะใช้ได้ดีกว่าวิธีวิธีจักรตชัน เมื่อขนาดตัวอย่าง ระดับสัมประสิทธิ์การแปรผันของความคลาดเคลื่อน จำนวนตัวแปรอิสระ สเกดแฟคเตอร์ เปอร์เซ็นต์การปลอมปน และระดับความสัมพันธ์ของตัวแปรอิสระสูงขึ้นไป ถึงแม้ว่าการที่ตัวแปรอิสระมีระดับความสัมพันธ์สูงเป็นข้อมูลที่ไม่เหมาะสมในการวิเคราะห์ความถดถอยพหุคูณก็ตาม แต่สำหรับวิธี โครงข่ายประสาทเทียมแล้ว จะเห็นว่าถ้าข้อมูลที่นำมาใช้มีความชัดเจนเพียงพอ แม้ว่าจะเป็นความชัดเจนของข้อมูลในทางลบก็ตาม ซึ่งในการศึกษานี้หมายถึงการที่ตัวแปรอิสระมีระดับความสัมพันธ์สูง สาเหตุที่เป็นเช่นนี้เนื่องจากแบบการส่งของวิธี โครงข่ายประสาทเทียมซึ่งใช้ฟังก์ชันซิกมอยด์ (sigmoid function) ที่สามารถปรับรูปร่างฟังก์ชันให้ใกล้เคียงกับฟังก์ชันเป้าหมาย หรือสมการเป้าหมาย และมีแบบการเรียนรู้ซึ่งใช้ความคลาดเคลื่อนในการพยากรณ์มาช่วยในการปรับ

ฟังก์ชันเป้าหมายให้ได้ใกล้เคียงมากที่สุดด้วยทำให้บางระดับความสัมพันธ์ของตัวแปรอิสระค่า AMSE ของวิธีโครงข่ายประสาทเทียมมีค่าเพิ่มขึ้น ดังนั้นอาจกล่าวได้ว่าระดับความสัมพันธ์ไม่มีผลกับวิธีโครงข่ายประสาทเทียมเท่ากับความจริงของข้อมูล และสาเหตุที่บางกรณีวิธีโครงข่ายประสาทเทียมให้ค่า AMSE ที่สูงกว่าวิธีรีดจีเรชันเนื่องจากขนาดตัวอย่างต่ำ ซึ่งส่งผลให้ความชัดเจนของข้อมูลไม่เพียงพอ วิธีโครงข่ายประสาทเทียมจะพยายามทำให้ค่าความคลาดเคลื่อนน้อยที่สุด โดยการหาฟังก์ชันเป้าหมายที่ผ่านข้อมูลทุกๆ จุด หรือผ่านมากที่สุดเท่าที่เป็นไปได้ อาจทำให้ค่าความคลาดเคลื่อนที่เกิดขึ้นมีค่าสูงกว่าวิธีรีดจีเรชัน หรืออาจเกิดจากการสุ่มค่า initial weight ที่ไม่ดีพอซึ่งเป็นผลจากการที่ข้อมูลต่ำ ให้ความชัดเจนของข้อมูลไม่เพียงพอ

นอกจากนี้แล้วจะเห็นได้ว่าวิธีโครงข่ายประสาทเทียม จะใช้ได้ดีกว่าวิธีรีดจีเรชัน ในกรณีที่มีความคลาดเคลื่อนมีการแจกแจงดอกรมอด การแจกแจงปกติปดอมปน และการแจกแจงปกติโดยเรียงตามลำดับจากมากไปน้อย ซึ่งในการวิเคราะห์หาคออยพหุคูณเมื่อเกิดพหุสัมพันธ์ในตัวแปรอิสระนั้น ส่วนใหญ่จะเป็นข้อมูลทางด้านเศรษฐศาสตร์ เช่นการกระจายรายได้ ซึ่งมีการแจกแจงดอกรมอดเป็นส่วนใหญ่ และข้อสังเกตของวิธีโครงข่ายประสาทเทียมก็คือ ถ้าข้อมูลที่ใช้มีรูปแบบที่ชัดเจน ไม่ว่าจะเป็นรูปแบบที่ไม่เป็นเชิงเส้นก็ตาม วิธีโครงข่ายประสาทเทียมก็สามารถให้ค่าพยากรณ์ที่ถูกต้องได้ โดยไม่ต้องใช้ข้อมูลมาก และถ้าใช้ข้อมูลเพิ่มมากขึ้นก็จะให้ค่าพยากรณ์ที่ถูกต้องมากขึ้นด้วย นอกจากนี้แล้ววิธีโครงข่ายประสาทเทียมจะให้ตัวแบบที่เหมาะสมซึ่งไปได้อยู่ในรูปเชิงเส้น ทำให้ไม่สามารถอธิบายความสัมพันธ์ของตัวแปรอิสระต่างๆ กับตัวแปรตามได้ชัดเจนเท่ากับวิธีรีดจีเรชัน ซึ่งในทางปฏิบัติอาจไม่สะดวกนัก ดังนั้นในการเลือกวิธีการพยากรณ์ควรพิจารณาวัตถุประสงค์ในการใช้งานควบคู่ไปด้วย ถ้าต้องการค่าพยากรณ์ที่ถูกต้องเพียงอย่างเดียวควรเลือกใช้วิธีโครงข่ายประสาทเทียม เพราะการที่ตัวแปรอิสระแต่ละตัวไม่มีสัมประสิทธิ์การคออยพหุคูณนั้น ทำให้ไม่สามารถบอกได้ว่าการเปลี่ยนแปลงค่าของตัวแปรอิสระทำให้ค่าตัวแปรตามเปลี่ยนแปลงไปอย่างไร แต่ก็สามารถหาค่านี้ได้ โดยการแทนค่าตัวแปรอิสระชุดใหม่ที่ต้องการทราบค่าตัวแปรตามกลับไปโปรแกรม เช่น การพยากรณ์ปรากฏการณ์ทางธรรมชาติต่างๆ การพยากรณ์ความคุ้มทุนในการผลิตสินค้าตัวใหม่ เป็นต้น หรือถ้าต้องการตัวแบบที่นำมาอธิบายข้อมูลได้อย่างชัดเจนก็ควรเลือกใช้วิธีรีดจีเรชัน เพราะในบางครั้งก็มีความจำเป็นที่จะต้องทราบการเปลี่ยนแปลงของตัวแปรอิสระแต่ละตัวที่มีต่อตัวแปรตาม เพื่อให้ประกอบในการตัดสินใจลดหรือเพิ่มตัวแปรอิสระนั้นๆ

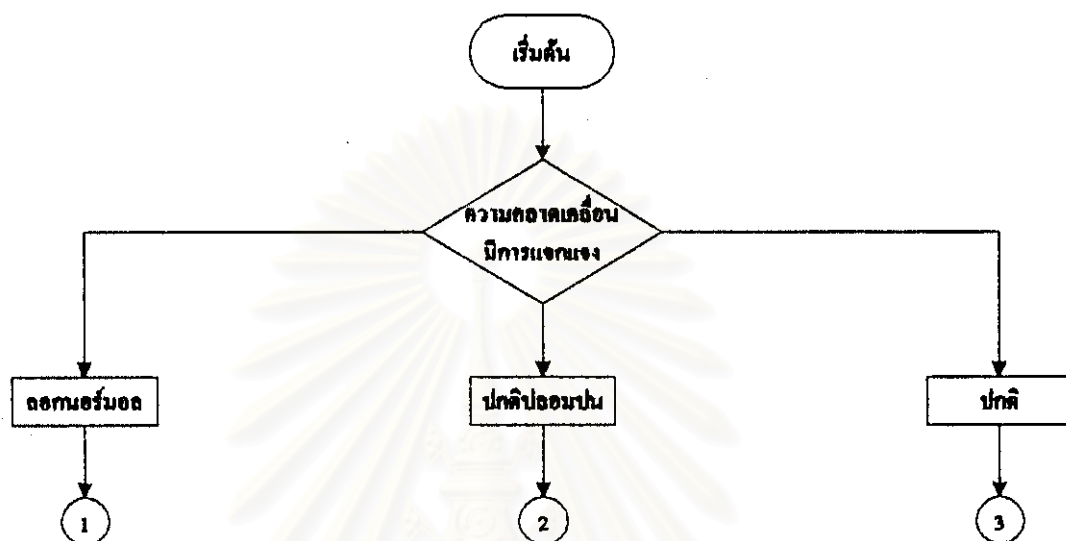
ข้อเสนอแนะ

1. ควรศึกษาเมื่อข้อมูลไม่ใช่ตัวแบบเชิงเส้น โดยการปรับให้ข้อมูลเป็นเชิงเส้นก่อนหรืออาจใช้วิธีการของการวิจัยดำเนินการแล้วนำมาเปรียบเทียบกับวิธีโครงข่ายประสาทเทียม เพราะวิธีโครงข่ายประสาทเทียมสามารถใช้ในการพยากรณ์ตัวแบบที่ไม่ใช่เชิงเส้นได้ นอกจากนี้แล้ววิธีการทางสถิติต้องมีการหาตัวแบบและมีการปรับข้อมูลทำให้ยุ่งยาก และเพื่อเป็นอีกทางเลือกหนึ่งในการตัดสินใจเมื่อข้อมูลไม่ใช่ตัวแบบเชิงเส้น
2. ควรศึกษาเปรียบเทียบกับวิธีทางสถิติอื่นๆ เช่น การวิเคราะห์องค์ประกอบหลัก (Principal Components) ที่ให้ประสิทธิภาพดีเมื่อเกิดหตุสัมพันธ์ในตัวแปรอิสระ
3. ควรศึกษาเปรียบเทียบกับวิธีโครงข่ายประสาทเทียมกับวิธีทางสถิติอื่นๆ ที่ใช้ในการพยากรณ์ เช่นเดียวกัน เช่น การวิเคราะห์อนุกรมเวลา เมื่อข้อมูลเกิดปัญหาหตุสัมพันธ์

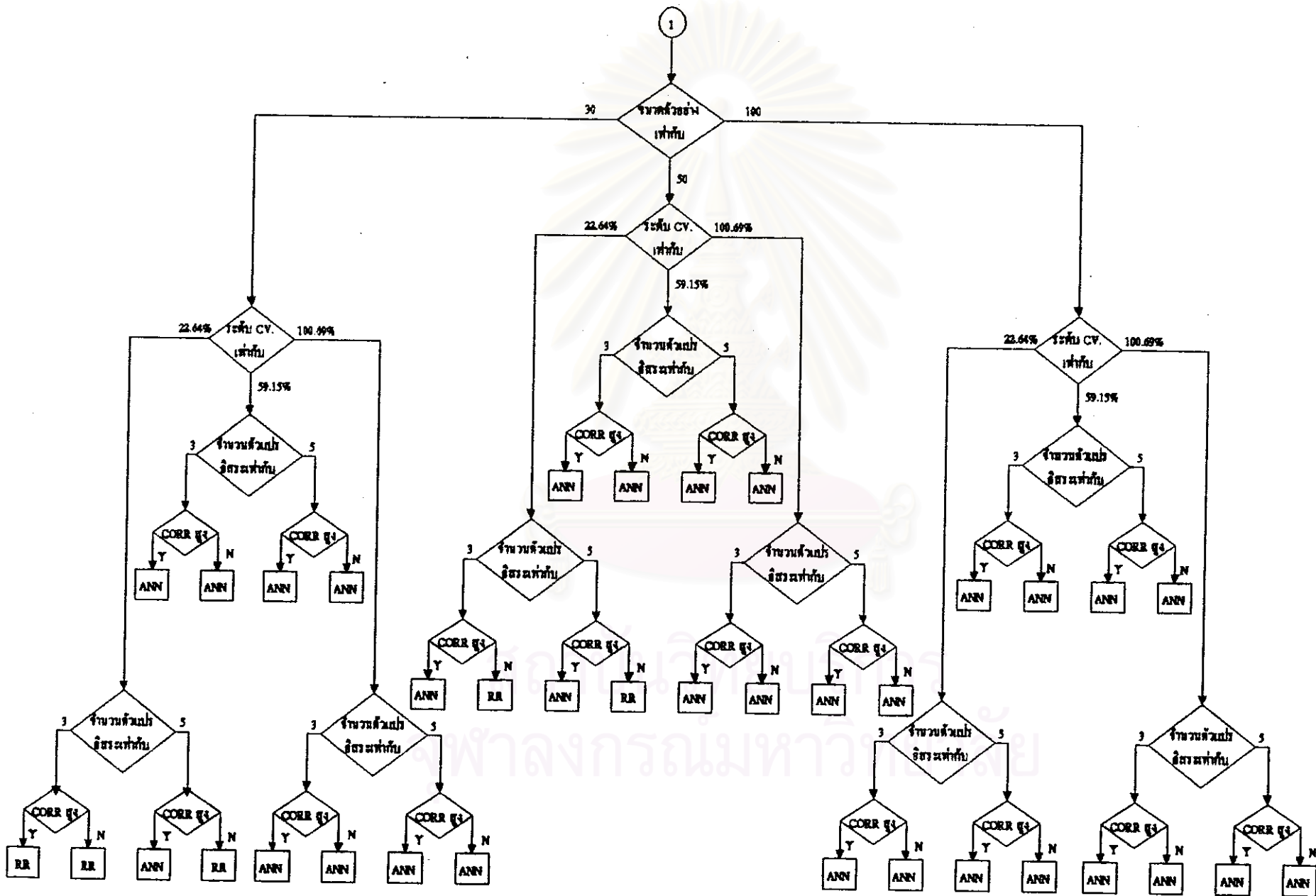
ผู้วิจัยแสดงแผนผังการเลือกใช้วิธีการพยากรณ์ในการวิเคราะห์ความถดถอยพหุคูณ ในกรณีที่เกิดหตุสัมพันธ์ระหว่างตัวแปรอิสระ ในสถานการณ์ต่างๆ โดยใช้สัญลักษณ์ต่อไปนี้

CORR	หมายถึง ระดับความสัมพันธ์ของตัวแปรอิสระ
Y	หมายถึง ใช่
N	หมายถึง ไม่ใช่
ANN	หมายถึง การพยากรณ์โดยใช้วิธีที่ให้ผลการของโครงข่ายประสาทเทียม
RR	หมายถึง การพยากรณ์โดยใช้วิธีเชิงวิเคราะห์
C	หมายถึง สเกลแฟกเตอร์
P	หมายถึง เปอร์เซ็นต์การปลอมปน
CV.	หมายถึง สัมประสิทธิ์การแปรผัน

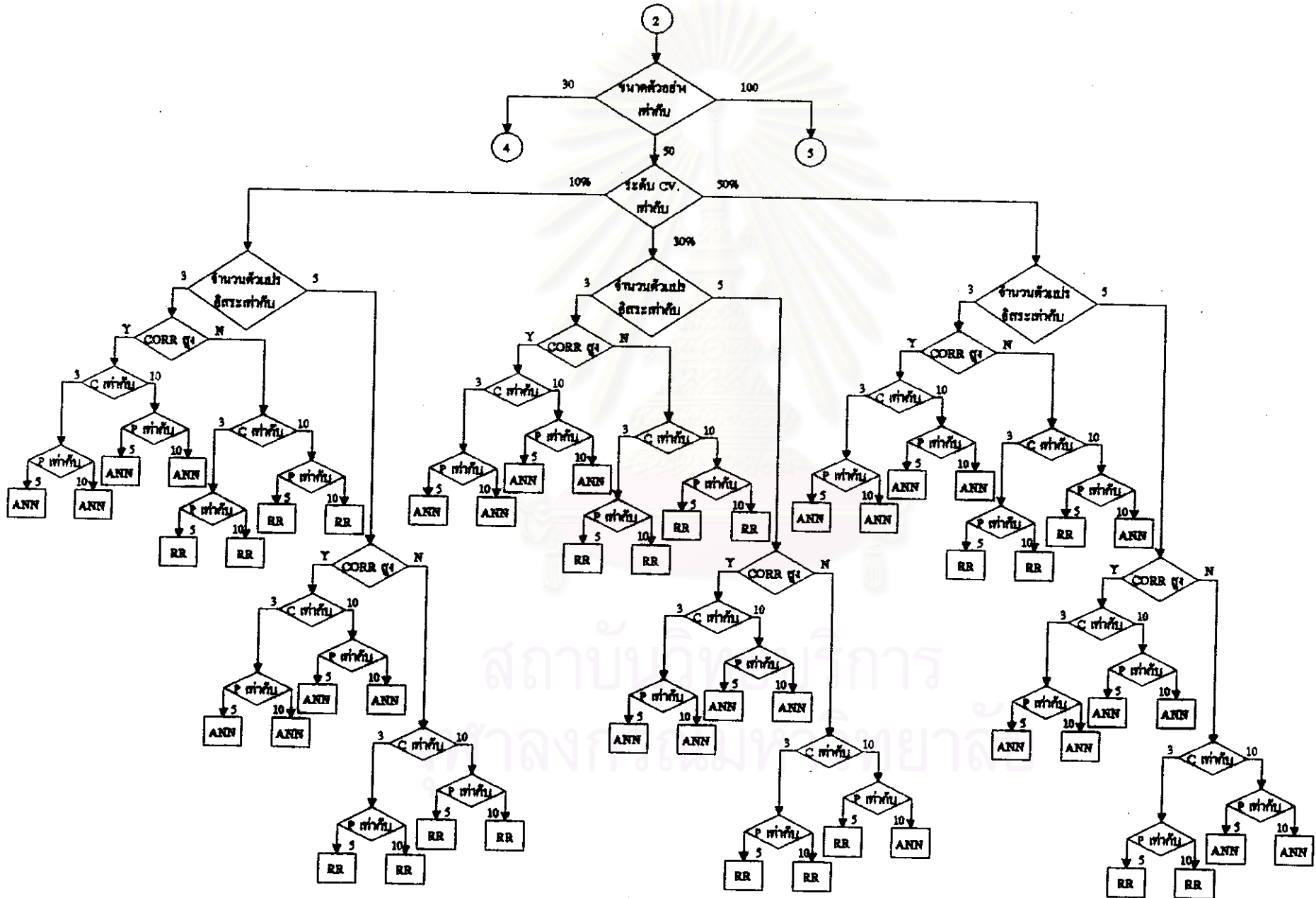
รูปที่ 6.1 แผนผังการเลือกใช่วิธีการพยากรณ์ในการวิเคราะห์ความถดถอยพหุคูณ ในกรณีที่เกิด
พหุสัมพันธ์ระหว่างตัวแปรอิสระ



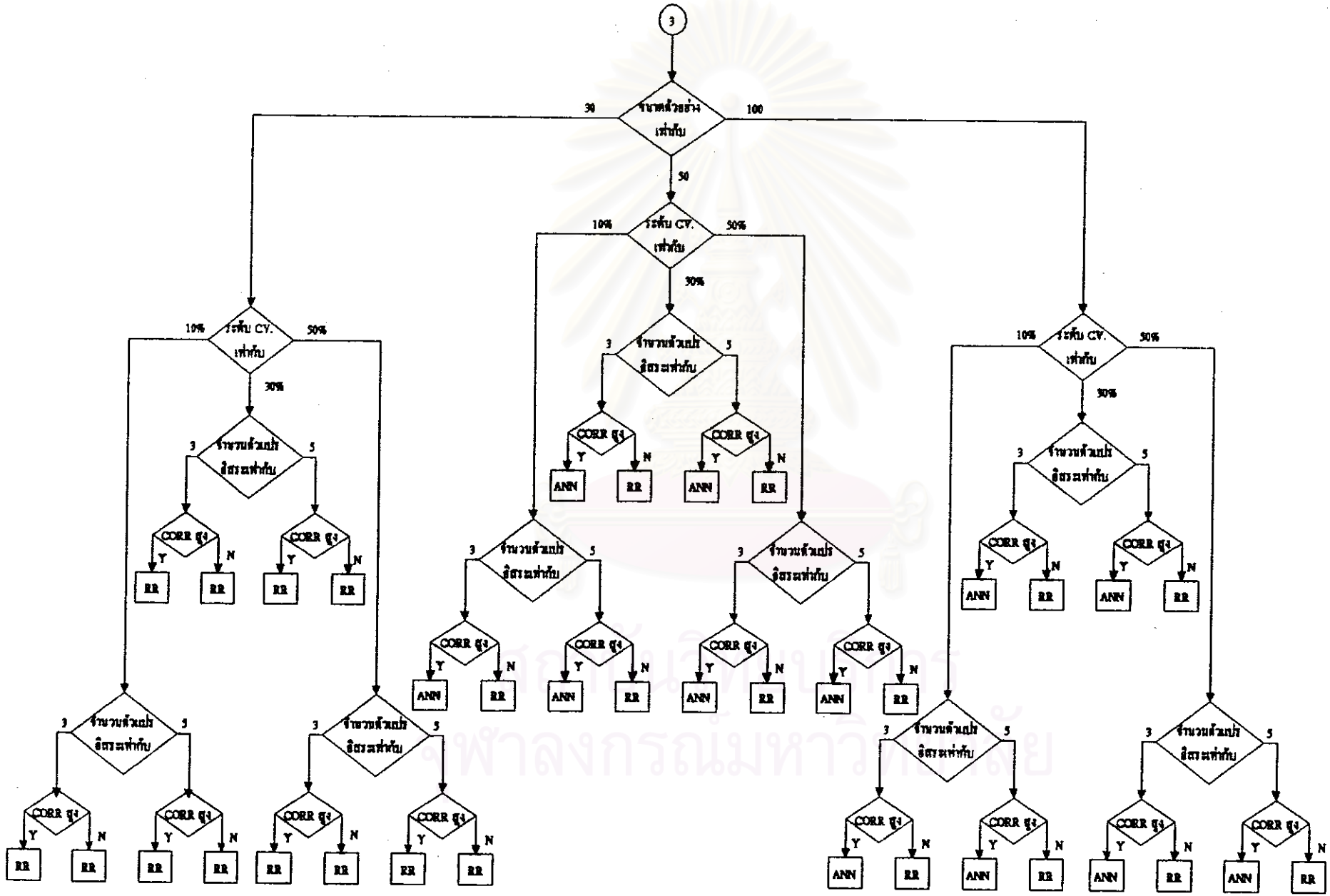
รูปที่ 5.1 (ต่อ)



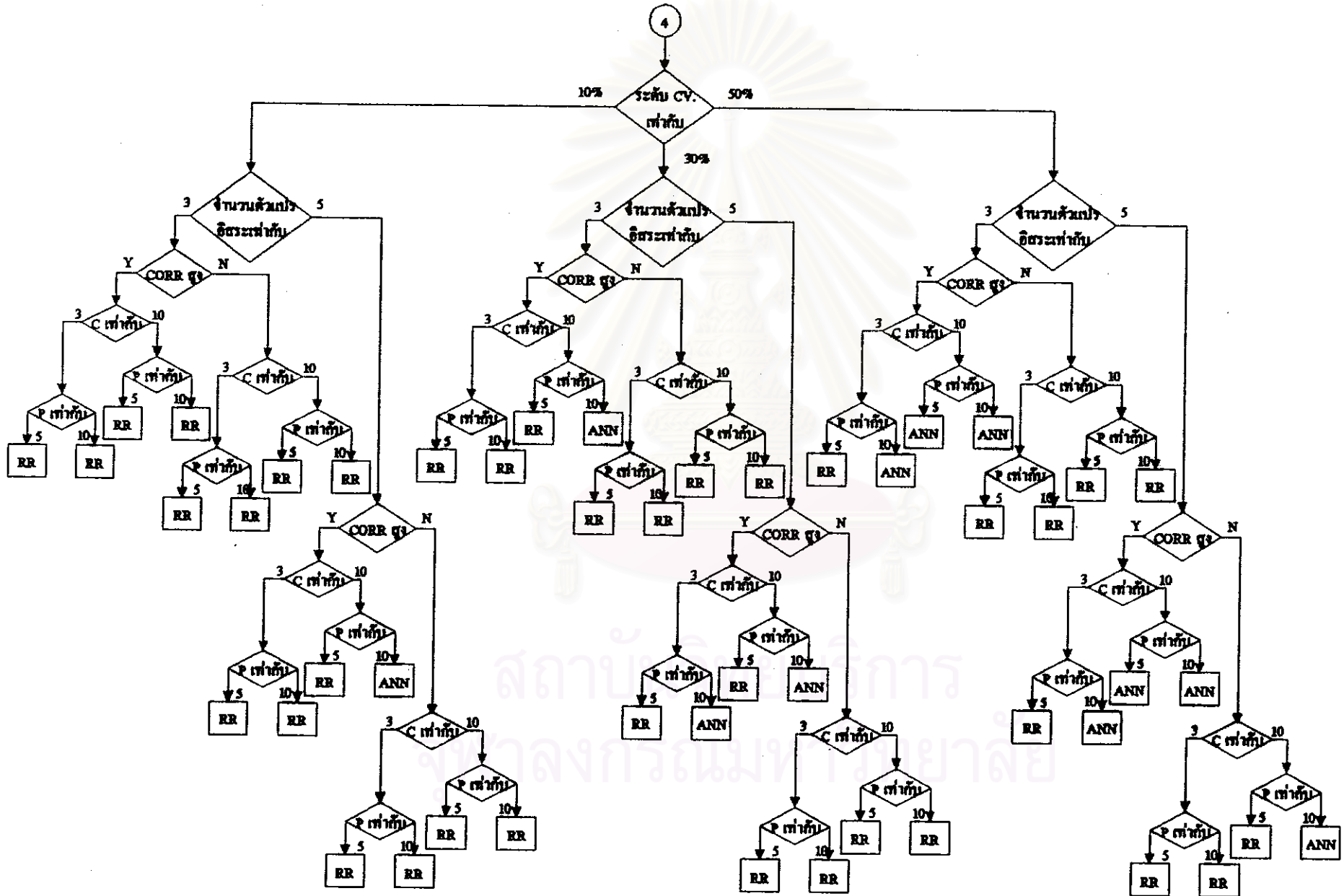
รูปที่ 5.1 (ต่อ)



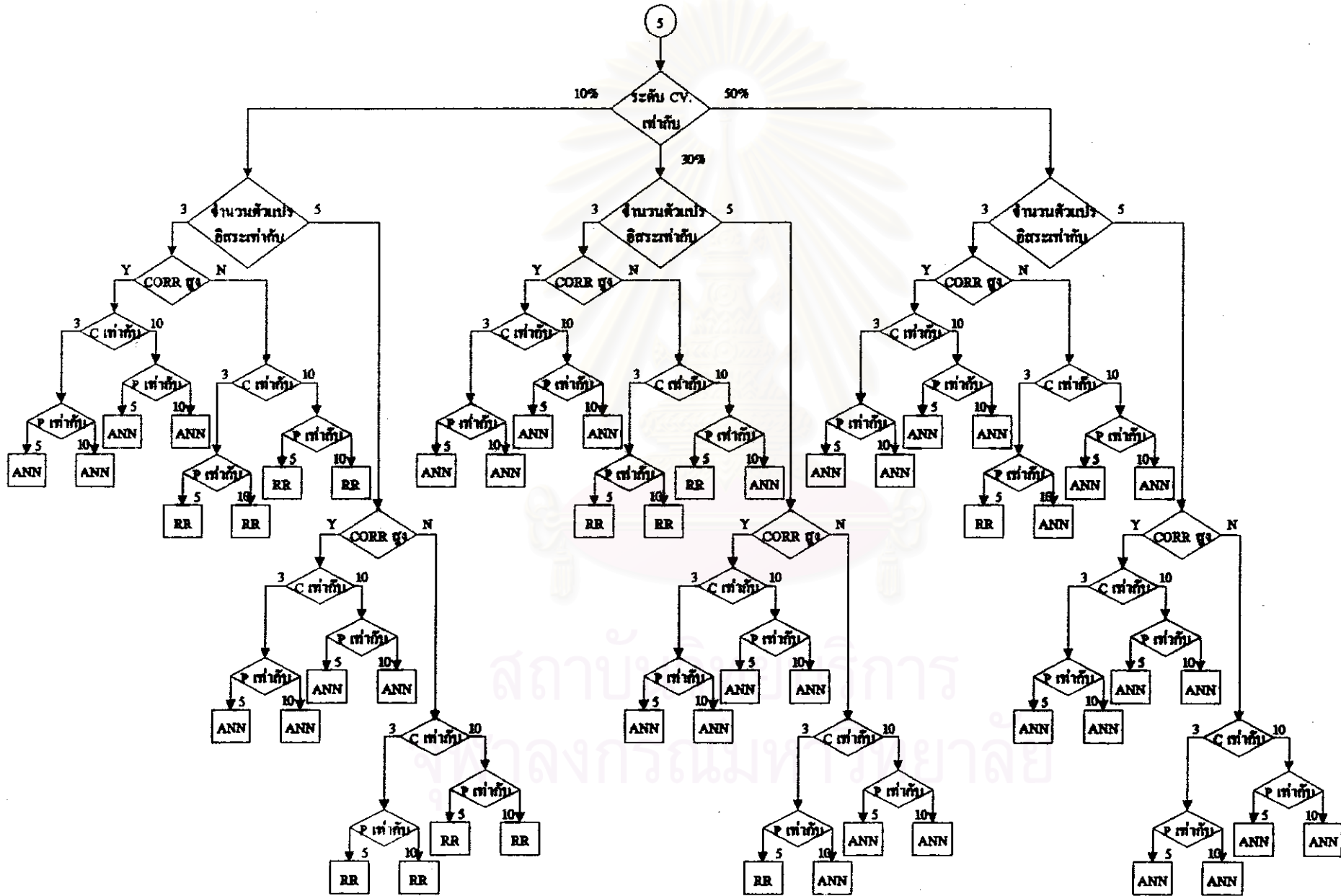
รูปที่ 5.1 (ต่อ)



รูปที่ 5.1 (ต่อ)



รูปที่ 5.1 (ต่อ)



จากรูปที่ 5.1 อาจมีความยุ่งยากในการตัดสินใจเลือกใช้วิธีการพยากรณ์ เนื่องจากในทางปฏิบัติจะทราบเพียงจำนวนตัวแปรอิสระ และขนาดตัวอย่าง⁵ ที่ใช้เท่านั้น ดังนั้นจากข้อมูลเบื้องต้นดังกล่าวสามารถสรุปผลได้ดังนี้

จำนวนตัวแปรอิสระ	ขนาดตัวอย่าง		
	30	50	100
3	RR	RR	ANN
5	RR	ANN	ANN

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

⁵ ในการวิจัยได้ใช้ขนาดตัวอย่างเท่ากับ 70 ผลปรากฏว่าเมื่อขนาดตัวอย่างเพิ่มขึ้นเป็น 70 แนวโน้มของค่า AMSE ที่ได้มีค่าลดลง เช่นเดียวกับเมื่อเพิ่มขนาดตัวอย่างเป็น 100 จึงตัดการวิจัยในส่วนนี้ออกไป