บทที่ 4
ผลการศึกษา

การวิเคราะห์ข้อมูลนี้มีวัตถุประสงค์ เพื่อหาตัวแบบพยากรณ์ที่เหมาะสมในการเลือกวิธีการพยากรณ์ เพื่อที่จะพยากรณ์ข้อมูลในอนาคต โดยจะศึกษาเปรียบเทียบค่าเฉลี่ยร้อยละของความคลาดเคลื่อนสัมบูรณ์ ที่ได้จากการพยากรณ์ด้วยวิธีการวิเคราะห์การถดถอยเชิงพหุคูณ การวิเคราะห์อนุกรมเวลาด้วยวิธีการปรับให้เรียบแบบเอ็กซ์โพแนนเชียลสองครั้ง และการวิเคราะห์อนุกรมเวลาด้วยวิธีการของโฮลท์ โดยใช้วิธีค่าสัมบูรณ์ต่ำสุด เพื่อหาผลสรุปว่าวิธีใดให้ค่าเฉลี่ยร้อยละของค่าความคลาดเคลื่อนสัมบูรณ์จากการพยากรณ์ต่ำสุด

จากการศึกษาถึงรูปแบบสมการการพยากรณ์ด้วยวิธีต่างๆดังกล่าว จะพิจารณาจากความแตกต่างและค่าเฉลี่ยร้อยละของค่าความคลาดเลื่อนสัมบูรณ์จากการพยากรณ์เป็นเกณฑ์ในการพิจารณาในการวัดวิธีการพยากรณ์ ซึ่งผลการพยากรณ์ครั้งนี้จะเสนอเป็นตารางและกราฟ

4.1 การเก็บรวบรวมข้อมูล

ในการเปรียบเทียบตัวแบบการพยากรณ์จำนวนผู้ใช้อินเทอร์เน็ตในประเทศไทยครั้งนี้ ใช้ข้อมูลทุติยภูมิ (Secondary data) ซึ่งเป็นข้อมูลอนุกรมเวลา (Time series data) ที่มีแนวโน้มในช่วงปี พ.ศ.2540 ถึง พ.ศ. 2551 รวมเป็นระยะเวลา 12 ปี โดยมีการเก็บรวบรวมข้อมูลจากหน่วยงานต่างๆ ดังนี้

1. ข้อมูลจำนวนผู้ใช้งานอินเทอร์เน็ตในประเทศไทย รวบรวมจาก หน่วยปฏิบัติการวิจัยเทคโนโลยีเครือข่าย ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ
2. ข้อมูลจำนวนเว็บไซต์ที่จดทะเบียนในประเทศไทยภายใต้โดเมน (.th) รวบรวมจาก หน่วยปฏิบัติการวิจัยเทคโนโลยีเครือข่าย ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ
3. ข้อมูลจำนวนนิติบุคลจัดตั้งใหม่ รวบรวมจาก กรมพัฒนาธุรกิจการค้า กระทรวงพานิชย์
4. ข้อมูลจำนวนประชากรจากการจดทะเบียนของประเทศไทย รวบรวมจาก กรมการปกครอง กระทรวงมหาดไทย
5. ข้อมูลผลิตภัณฑ์มวลรวมประชาชาติ รวบรวมจาก ธนาคารแห่งประเทศไทย

6. ข้อมูลอัตราแลกเปลี่ยนเงินบาทต่อดอลลาร์สหรัฐ รวบรวมจาก ธนาคารแห่งประเทศไทย

4.2 ตัวแบบพยากรณ์จำนวนผู้ใช้อินเทอร์เน็ตในประเทศไทย

หลังจากทำการวิเคราะห์หาตัวแบบพยากรณ์โดยใช้วิธีการวิเคราะห์การถดถอยเชิงพหุคูณ การวิเคราะห์อนุกรมเวลาด้วยวิธีการปรับให้เรียบแบบเอ็กซ์โพแนนเชียลสองครั้ง และการวิเคราะห์อนุกรมเวลาด้วยวิธีการของโฮลท์ ได้ตัวแบบพยากรณ์จำนวนผู้ใช้อินเทอร์เน็ตในประเทศไทย สำหรับแต่ละวิธีการพยากรณ์ต่างๆ ข้างต้น หลังจากนั้นจึงทำการเปรียบเทียบค่าเฉลี่ยร้อยละของความคลาดเคลื่อนสัมบูรณ์ของตัวแบบพยากรณ์ที่ได้จากแต่ละวิธีการพยากรณ์ ทั้ง 3 วิธี แล้วจึงเลือกตัวแบบพยากรณ์ที่ให้ค่าเฉลี่ยร้อยละของความคลาดเคลื่อนสัมบูรณ์ต่ำที่สุด ดังจะนำเสนอผลการวิเคราะห์แบ่งตามขั้นตอนดังนี้

4.2.1 การวิเคราะห์อนุกรมเวลาด้วยวิธีการปรับให้เรียบแบบเอ็กซ์โพแนนเชียลสองครั้ง

ข้อมูลจำนวนผู้ใช้อินเทอร์เน็ตในประเทศไทยที่นำมาวิเคราะห์เป็นข้อมูลรายปี ตั้งแต่ปี พ.ศ. 2540 ถึง พ.ศ.2551 จำนวน 12 ปี พบว่าลักษณะของข้อมูลอนุกรมเวลาชุดนี้มีอิทธิพลเนื่องจากแนวโน้ม จึงเหมาะสมที่จะใช้เทคนิคปรับให้เรียบแบบเอ็กซ์โพแนนเชียลสองครั้ง (รายละเอียดแสดงในภาคผนวก ข)

จากการหาค่าคงที่ทำโดยใช้โปรแกรมสำเร็จรูปคำนวณเพื่อให้ได้ค่าพยากรณ์ ทำให้ทราบค่าทำให้เรียบสำหรับข้อมูลชุดนี้ มีค่าคงที่ 1 ค่าคือ ค่าทำให้เรียบระหว่างข้อมูลกับค่าพยากรณ์ (α) มีค่าเท่ากับ 0.359

4.2.2 การวิเคราะห์อนุกรมเวลาด้วยวิธีการของโฮลท์

ข้อมูลจำนวนผู้ใช้อินเทอร์เน็ตในประเทศไทยที่นำมาวิเคราะห์เป็นข้อมูลรายปี ตั้งแต่ปี พ.ศ. 2540 ถึง พ.ศ.2551 จำนวน 12 ปี พบว่าลักษณะของข้อมูลอนุกรมเวลาชุดนี้มีอิทธิพลเนื่องจากแนวโน้ม จึงเหมาะสมที่จะใช้เทคนิคปรับให้เรียบด้วยวิธีการของโฮลท์ (รายละเอียดแสดงในภาคผนวก ข)

จากการหาค่าคงที่ทำโดยใช้โปรแกรมสำเร็จรูปคำนวณเพื่อให้ได้ค่าพยากรณ์ ทำให้ทราบค่าทำให้เรียบสำหรับข้อมูลชุดนี้ มีค่าคงที่ 2 ค่าคือ ค่าทำให้เรียบระหว่างข้อมูลกับค่าพยากรณ์ (α) มีค่าเท่ากับ 0.710 และค่าปรับให้เรียบระหว่างแนวโน้มกับค่าประมาณแนวโน้ม (γ) มีค่าเท่ากับ 0.810

4.2.3 การวิเคราะห์การถดถอยเชิงพหุคูณ

สมการที่ 4.1 แสดงตัวแบบพยากรณ์ที่ได้ จากการวิเคราะห์การถดถอยเชิงพหุคูณ

$$\hat{y}_t = -2162975 + 379.683x_2 + 76.389x_4 \quad (4.1)$$

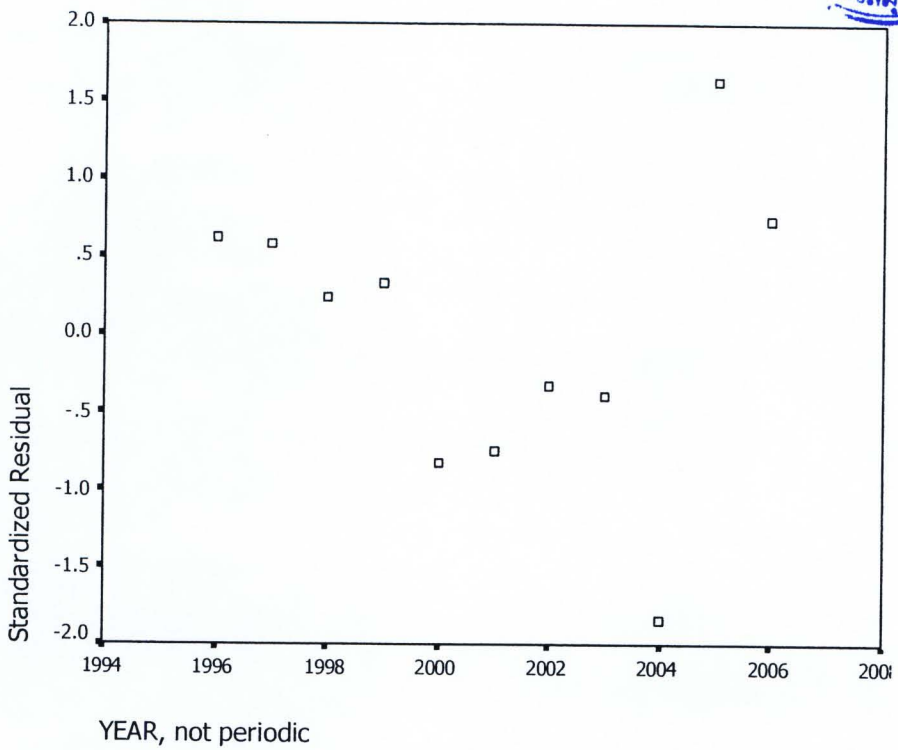
โดยที่ $\hat{y}_t$ คือ จำนวนผู้ใช้อินเทอร์เน็ต ณ ปีที่ t

x_2 คือ จำนวนเว็บไซต์ที่จดทะเบียนในประเทศไทยภายใต้โดเมน (.th)

x_4 คือ จำนวนการจดทะเบียนนิติบุคคลต่อกระทรวงพาณิชย์ในประเทศไทย

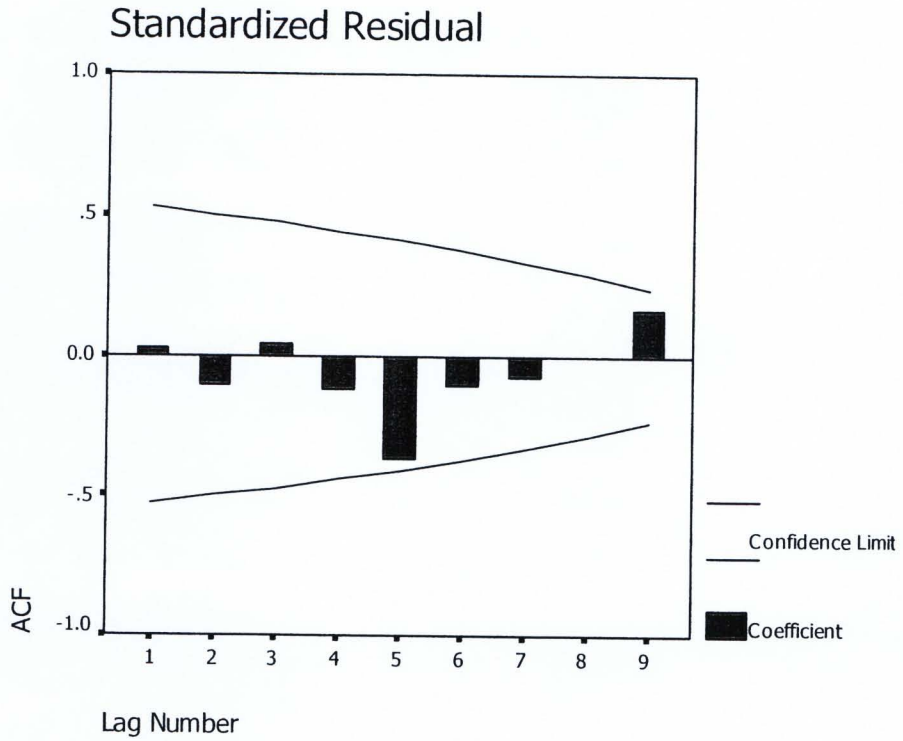
ข้อมูลที่นำมาวิเคราะห์เป็นข้อมูลรายปี ตั้งแต่ปี พ.ศ. 2540 ถึง พ.ศ.2551 จำนวน 12 ปี นำมาวิเคราะห์หาตัวแบบพยากรณ์ด้วยโปรแกรมสำเร็จรูปโดยใช้วิธีการคัดเลือกตัวแปรด้วยวิธี STEPWISE หลังจากดำเนินการตามขั้นตอนการสร้างตัวแบบพยากรณ์ของวิธีการวิเคราะห์การถดถอยเชิงพหุคูณ โดยพิจารณาจากตัวแบบอิสระต่างๆ (ดังที่แสดงรายละเอียดในบทที่ 3)

ทำการตรวจสอบความเหมาะสมของตัวแบบพยากรณ์ โดยพิจารณากราฟของเศษเหลือตกค้างกับแกนเวลา พบว่าค่าคลาดเคลื่อนกระจายตัวอยู่รอบค่าศูนย์ แสดงว่าความคลาดเคลื่อนมีค่าเฉลี่ยเท่ากับศูนย์ และความแปรปรวนคงที่ ดังแสดงในภาพที่ 4.1

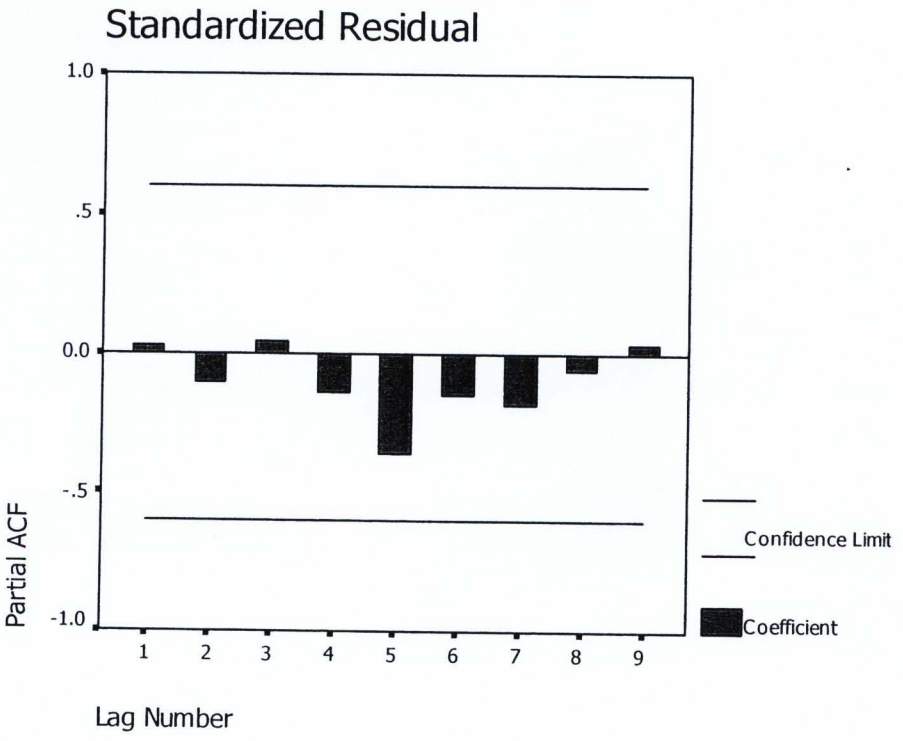


ภาพที่ 4.1 ความคลาดเคลื่อนมีค่าเฉลี่ยเท่ากับศูนย์ และความแปรปรวนคงที่

จากนั้นพิจารณากราฟ ACF และ PACF ของค่าคลาดเคลื่อนพบว่าฟังก์ชันสหสัมพันธ์ในตัวเองของความคลาดเคลื่อนจะตกอยู่ในขอบเขตความเชื่อมั่นร้อยละ 95 ทุกค่า lag แสดงว่าค่าความคลาดเคลื่อนไม่มีสหสัมพันธ์ ซึ่งจากที่กล่าวมาข้างต้น สรุปได้ว่าตัวแบบการวิเคราะห์การถดถอยเชิงพหุคูณมีความเหมาะสมแล้ว ดังแสดงในภาพที่ 4.2 และ ภาพที่ 4.3 ตามลำดับ



ภาพที่ 4.2 ACF ของฟังก์ชันสหสัมพันธ์ในตัวเองของความคลาดเคลื่อน



ภาพที่ 4.3 PACF ของฟังก์ชันสหสัมพันธ์ในตัวเองของความคลาดเคลื่อน

4.3 การทดสอบสมมติฐานของกลุ่มตัวอย่าง (t-test)

ผลจากการทดสอบค่าที (t-test) พบว่าค่า Sig. (2-tailed) > 0.05 แสดงว่าตัวแบบที่ได้จากวิธีการปรับให้เรียบแบบเอ็กซ์โพแนนเชียลสองครั้ง วิธีการของโฮลท์ และการวิเคราะห์การถดถอยเชิงพหุคูณ ให้ผลการพยากรณ์ไม่แตกต่างกัน ที่ระดับนัยสำคัญ 0.05 (ดังแสดงรายละเอียดในภาคผนวก ข)

4.4 การเปรียบเทียบค่าเฉลี่ยร้อยละของความคลาดเคลื่อนสัมบูรณ์

ทำการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ด้วยโปรแกรม Excel ซึ่งมีตัวแบบพยากรณ์ทั้งหมด 3 วิธี โดยพิจารณาค่าเฉลี่ยร้อยละของความคลาดเคลื่อนสัมบูรณ์ในช่วงเวลาเดียวกัน ดังแสดงในตารางที่ 4.1 โดยค่าเฉลี่ยร้อยละของความคลาดเคลื่อนสัมบูรณ์ สามารถหาได้จากสมการที่ 4.2

$$MAPE = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{e_t}{Y_t}\right| \qquad (4.2)$$

โดยที่ e_t คือ ความคลาดเคลื่อนของการพยากรณ์

ซึ่งคำนวณได้ดังนี้ $e_t = Y_t - \hat{Y}_t$

Y_t คือ ค่าจริงของข้อมูลอนุกรมเวลา ณ เวลาที่ t

$\hat{Y}_t$ คือ ค่าพยากรณ์ของข้อมูลอนุกรมเวลา ณ เวลาที่ t

ตารางที่ 4.1 เปรียบเทียบค่าเฉลี่ยร้อยละของความคลาดเคลื่อนสัมบูรณ์ของจำนวนผู้ใช้อินเทอร์เน็ต

ค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ (MAPE)			
ปี พ.ศ.	**Regression**	**Double Exponential**	**Holt**
2540	1.005712977	41.41856929	3.280991735
2541	0.642803454	8.945874634	1.175889552
2542	0.156337635	0.129290273	0.017397445
2543	0.12035837	0.680025059	0.094404996
2544	0.048640122	0.610274142	0.156474118
2545	0.000590609	0.553292467	0.089281996
2546	0.06507909	0.454457237	0.007925288
2547	0.13799412	0.347676012	0.045952053
2548	0.080133447	0.418660582	0.17420907
2549	0.089727824	0.40664879	0.009465504
2550	0.02284222	0.300989264	0.005876523
2551	0.009556665	0.311535415	0.04074325
MAPE	**19.83147112**	**454.8107763**	**42.48842942**

เมื่อพิจารณาค่าเฉลี่ยร้อยละของความคลาดเคลื่อนสัมบูรณ์ของการพยากรณ์ (MAPE) จะเห็นว่า ค่าพยากรณ์โดยวิธีการปรับให้เรียบแบบเอ็กซ์โพแนนเชียลสองครั้งจะให้ค่า MAPE เท่ากับ 454.810 ค่าพยากรณ์โดยการวิเคราะห์อนุกรมเวลาด้วยวิธีการของโฮลท์จะให้ค่า MAPE เท่ากับ 42.488 และค่าพยากรณ์โดยการวิเคราะห์การถดถอยเชิงพหุคูณจะให้ค่า MAPE เท่ากับ 19.831 ดังนั้นในการคาดคะเนจำนวนผู้ใช้อินเทอร์เน็ตในประเทศไทย ค่าพยากรณ์ที่ได้จากการวิเคราะห์การถดถอยเชิงพหุคูณจะเหมาะสมกับอนุกรมเวลาชุดนี้ เนื่องจากให้ค่า MAPE ต่ำที่สุด โดยมีปัจจัยที่เกี่ยวข้องกับค่าพยากรณ์คือ จำนวนเว็บไซต์ที่จดทะเบียนในประเทศไทยภายใต้โดเมน (.th) และจำนวนการจดทะเบียนนิติบุคคลต่อกระทรวงพาณิชย์ในประเทศไทย

หลังจากได้ตัวแบบสำหรับการพยากรณ์จำนวนผู้ใช้อินเทอร์เน็ตในประเทศไทยด้วยวิธีการวิเคราะห์ความถดถอยแล้ว จากนั้นนำตัวแบบมาตรวจสอบความแม่นยำโดยการคำนวณ ดังแสดงในตารางที่ 4.2

ตารางที่ 4.2 ค่าคาดคะเนของจำนวนผู้ใช้อินเทอร์เน็ตจากตัวแบบพยากรณ์การถดถอยเชิงพหุคูณ

พ.ศ.	จำนวนผู้ใช้อินเทอร์เน็ต (คน)	
	ค่าจริง	ค่าพยากรณ์
2549	11,413,000	10,388,936
2550	13,416,000	13,722,451
2551	16,100,000	16,253,862
2552	-	19,356,811