บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ไฟฟ้าเป็นปัจจัยที่สำคัญที่สุดปัจจัยหนึ่งสำหรับการดำรงชีวิตประจำวันของคนในชาติ ทั้งการสื่อสาร การคมนาคม เศรษฐกิจ การเกษตร และ การศึกษา ไฟฟ้าเป็นปัจจัยที่ช่วยเหลือในการทำงานด้านต่าง ๆ เพื่อความสะดวกสบายที่เพิ่มขึ้น ปัจจุบันมีการนำเทคโนโลยีสารสนเทศและการสื่อสารจำนวนมากเข้ามาใช้เป็นเครื่องมือในการเรียนการสอน รวมถึงช่วยอำนวยความสะดวกสบายในการทำกิจกรรมทุกด้าน ซึ่งเทคโนโลยีทั้งหมดนี้จะทำงานไม่ได้ถ้าหากขาดพลังงานไฟฟ้า ไม่ว่าจะเป็นอาคารเรียน อาคารสำนักงาน ร้านค้า บ้านพักและหอพักของนักศึกษาจำเป็นต้องใช้ไฟฟ้าทั้งสิ้น

มหาวิทยาลัยขอนแก่นเป็นสถาบันอุดมศึกษาสถานแห่งแรกของภาคตะวันออกเฉียงเหนือ ถือเป็นศูนย์รวมทางความคิด สติปัญญาของสังคม และเป็นศูนย์กลางทางการศึกษาระดับอุดมศึกษาสู่ภูมิภาค ตั้งอยู่บนพื้นที่ประมาณ 5,500 ไร่ มีคณะวิชาที่ผลิตบัณฑิตจำนวน 25 คณะวิชา นอกจากนี้ยังมีสิ่งอำนวยความสะดวกต่าง ๆ ได้แก่ โรงพยาบาลศรีนครินทร์ และหน่วยงานเทียบเท่าคณะ ประกอบด้วย ศูนย์สถาบัน สำนักให้บริการวิชาการและบริการชุมชน มีที่ทำการไปรษณีย์ ศูนย์บริการ สหกรณ์ ร้านค้า หอพัก บ้านพัก แฟลต เรือนรับรอง ธนาคาร โรงเรียน และสาธารณูปโภคอื่น ๆ เพื่อให้บริการแก่บุคลากร นักศึกษา และประชนชนทั่วไปอย่างครบครัน และปัจจุบันมีจำนวนคนและอาคารสถานที่เพิ่มขึ้นทั้งที่อยู่ในระหว่างการก่อสร้างและสร้างเสร็จแล้ว จึงทำให้ความต้องการใช้พลังงานไฟฟ้าเพิ่มมากขึ้นตามไปด้วย อีกทั้งรายจ่ายด้านการใช้พลังงานไฟฟ้าของมหาวิทยาลัยสูงขึ้นเกือบทุกปี และยังพบว่ามีจุดชำรุดในสายส่งไฟฟ้าที่จ่ายกระแสไฟฟ้าให้ทางมหาวิทยาลัย จึงจำเป็นต้องงดจ่ายกระแสไฟฟ้าเป็นการชั่วคราวอยู่บ่อยครั้ง ส่งผลต่อความไม่สะดวกต่าง ๆ ในการใช้ชีวิตและการทำกิจกรรมของทุกคนในมหาวิทยาลัย (มหาวิทยาลัยขอนแก่น, 2562: ออนไลน์)

จากข้อมูลที่กล่าวมาข้างต้นจะเห็นว่าจะต้องมีการวางแผนการจัดการพลังงานไฟฟ้าให้เพียงพอต่อจำนวนนักศึกษา บุคลากร และอาคารสถานที่ต่าง ๆ ในอนาคต ซึ่งวิธีการที่จะช่วยในการจัดการกับทรัพยากรไฟฟ้าคือการทราบค่าการพยากรณ์ปริมาณการใช้ไฟฟ้าในอนาคตที่น่าเชื่อถือ ผู้วิจัยจึงมุ่งที่จะศึกษาเพื่อหาเทคนิคการพยากรณ์ที่เหมาะสมกับลักษณะข้อมูลปริมาณการใช้ไฟฟ้า โดยจะใช้การวิเคราะห์อนุกรมเวลา จากการที่ได้ไปศึกษางานวิจัยเกี่ยวกับการพยากรณ์ปริมาณการใช้ไฟฟ้า งานวิจัยส่วนมากจะใช้การเปรียบเทียบวิธีการทำให้เรียบและวิธีของบ็อกซ์และเจนกินส์ ซึ่งเทคนิคการพยากรณ์ที่นำมาใช้ในการวิจัยครั้งนี้ประกอบด้วย วิธีการทำให้เรียบ (Smoothing Method) และ วิธีของบ็อกซ์และเจนกินส์ (Box-Jenkins’s Method)

1.2 วัตถุประสงค์ของการวิจัย

เพื่อหาตัวแบบการพยากรณ์ที่เหมาะสมในการพยากรณ์ปริมาณการใช้ไฟฟ้ารายเดือน ของมหาวิทยาลัยขอนแก่น และแต่ละคณะในมหาวิทยาลัยขอนแก่น

1.3 ขอบเขตของการวิจัย

1.3.1 ข้อมูลที่ใช้ในการวิจัยครั้งนี้เป็นข้อมูลทุติยภูมิรายเดือน ของปริมาณการใช้ไฟฟ้าในมหาวิทยาลัยขอนแก่น และแต่ละคณะในมหาวิทยาลัยขอนแก่น เก็บรวบรวมจากหน่วยไฟฟ้า กองอาคารและสถานที่ มหาวิทยาลัยขอนแก่น ตั้งแต่เดือนมกราคม พ.ศ. 2551 ถึงเดือนพฤษภาคม พ.ศ. 2561

1.3.2 ใช้วิธีการทำให้เรียบ (Smoothing Method) และวิธีของบ็อกซ์และเจนกินส์ (Box-Jenkins's Method) ในการพยากรณ์อนุกรมเวลาสำหรับการพยากรณ์ปริมาณการใช้ไฟฟ้าในมหาวิทยาลัยขอนแก่น

1.4 ความหมายหรือนิยามศัพท์เฉพาะ

1.4.1 การพยากรณ์ (Forecasting) หมายถึง การประมาณ หรือ การคาดคะเนว่าอะไรจะเกิดขึ้นในอนาคต

1.4.2 อนุกรมเวลา (Time Series) หมายถึง ข้อมูลหรือค่าสังเกตที่เปลี่ยนแปลงไปตามลำดับเวลาที่เกิดขึ้นอย่างต่อเนื่อง ข้อมูลเหล่านี้ถูกเก็บรวบรวม ณ ช่วงเวลาต่าง ๆ เช่น รายวัน รายสัปดาห์ รายเดือน รายไตรมาส หรือรายปี

1.4.3 พลังงานไฟฟ้า คือ พลังงานที่เปลี่ยนมาจากพลังงานรูปอื่น พลังไฟฟ้านี้เกิดจากการที่อิเล็กตรอนเคลื่อนที่ผ่านตัวนำไฟฟ้า อิเล็กตรอนจะเคลื่อนที่จากขั้วลบไปขั้วบวก แต่ไฟฟ้าเป็นกระแสสมมุติเคลื่อนสวนทางกับอิเล็กตรอนจากขั้วบวกไปขั้วลบ

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1.5.1 ได้ตัวแบบการพยากรณ์ที่เหมาะสมในการพยากรณ์ปริมาณการใช้ไฟฟ้ารายเดือน ในมหาวิทยาลัยขอนแก่น และแต่ละคณะในมหาวิทยาลัยขอนแก่น

1.5.2 ทราบแนวโน้มของปริมาณการใช้ไฟฟ้ารายเดือน ในมหาวิทยาลัยขอนแก่น และแต่ละคณะในมหาวิทยาลัยขอนแก่น

1.5.3 หน่วยไฟฟ้า กองอาคารและสถานที่ มหาวิทยาลัยขอนแก่น สามารถใช้เป็นแนวทางในการวางแผนการจัดหาพลังงานไฟฟ้าอย่างคุ้มค่ามากที่สุด

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

การศึกษาครั้งนี้ได้ศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้องกับการพยากรณ์ปริมาณการใช้ไฟฟ้าในมหาวิทยาลัยขอนแก่น และแต่ละคณะในมหาวิทยาลัยขอนแก่น โดยแบ่งออกเป็น 2 ส่วนคือ

1. ทฤษฎีที่ใช้ในการวิจัย แสดงวิธีการและทฤษฎีทั้งหมดที่ใช้ในงานวิจัยนี้
2. ผลงานวิจัยที่เกี่ยวข้องเป็นการรวบรวมผลการศึกษาเกี่ยวกับวิธีการหาตัวแบบของปริมาณการใช้ไฟฟ้า

2.1 ทฤษฎีสถิติที่เกี่ยวข้องกับงานวิจัย

2.1.1 การพยากรณ์อนุกรมเวลา เป็นการพยากรณ์ที่ใช้ข้อมูลในอดีตและปัจจุบันของสิ่งที่จะพยากรณ์เท่านั้นมาวิเคราะห์ ไม่ได้ใช้ข้อมูลอย่างอื่นเรียกข้อมูลชนิดนี้ว่า “อนุกรมเวลา (Time Series)” หมายถึง ค่าสังเกตที่สังเกตที่ทุกหน่วยเวลาติดต่อกันตามลำดับ แต่ละหน่วยเวลาห่าง (Lag) เท่ากัน เขียนแทนได้ดังนี้ (ยุภาพร ตงประสิทธิ์, 2559)

ให้ Z_t แทน ค่าสังเกตที่เวลา t ใด ๆ

$Z_1, Z_2, Z_3, ..., Z_n$ แทน ค่าสังเกตที่เวลา t เมื่อ $t = 1, 2, ..., n$

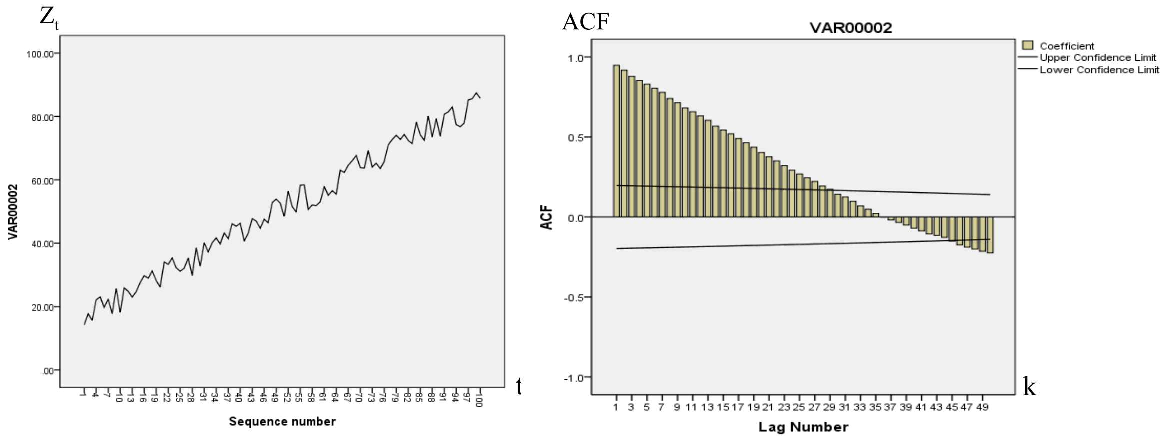
Z_{t-1}, Z_t และ Z_{t+1} แทน ค่าสังเกตที่หน่วยเวลา $t-1$, t และ $t+1$ ตามลำดับ

2.1.1.1 ลักษณะของอนุกรมเวลา อธิบายได้ด้วยส่วนประกอบ 4 ส่วน คือ แนวโน้ม การแปรผันตามฤดูกาล วัฏจักร และความรบกวนสุ่มหรือการแปรผันผิดปกติ ในแต่ละอนุกรมเวลาอาจมีส่วนประกอบทั้ง 4 ส่วน หรือมีเฉพาะบางส่วนก็ได้ การศึกษาลักษณะอนุกรมเวลากระทำได้โดยการพิจารณาจาก กราฟของอนุกรมเวลา และกราฟของฟังก์ชันสหสัมพันธ์ในตัวเอง (Autocorrelation Function: ACF) ดังนี้

1) แนวโน้ม (Trend: T)

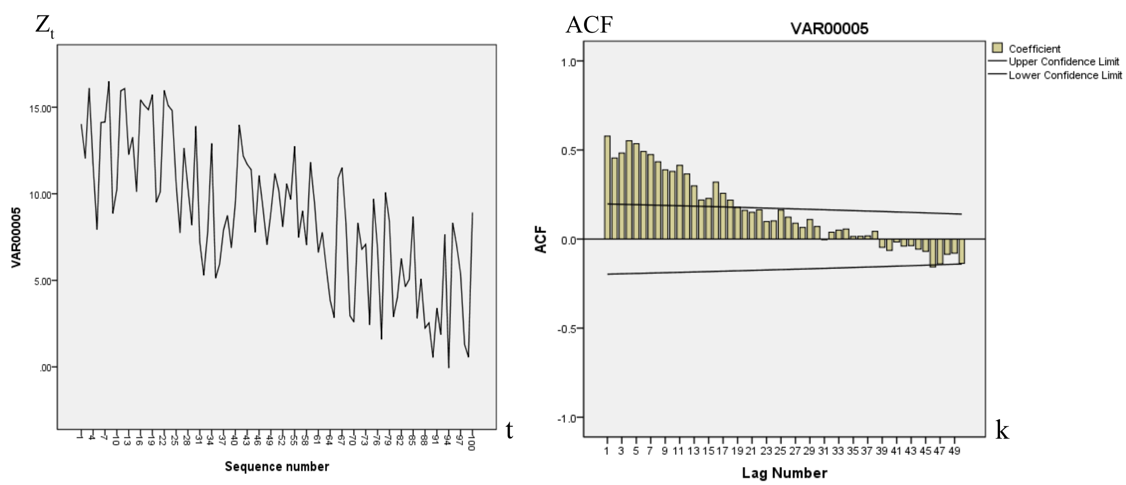
แนวโน้ม หมายถึง การเคลื่อนไหวหรือการเปลี่ยนแปลงของค่าของอนุกรมเวลาในระยะยาวในลักษณะเพิ่มขึ้นหรือลดลง แนวโน้มอาจเป็นเชิงเส้นหรือไม่ใช่เชิงเส้นก็ได้เช่น ราคาปิดของหุ้นรายวัน ปริมาณขายสินค้ารายเดือน จำนวนนักศึกษาทั้งหมดรายปี เป็นต้น

ลักษณะของอนุกรมเวลาที่มีแนวโน้ม สามารถพิจารณาจากกราฟ ดังนี้



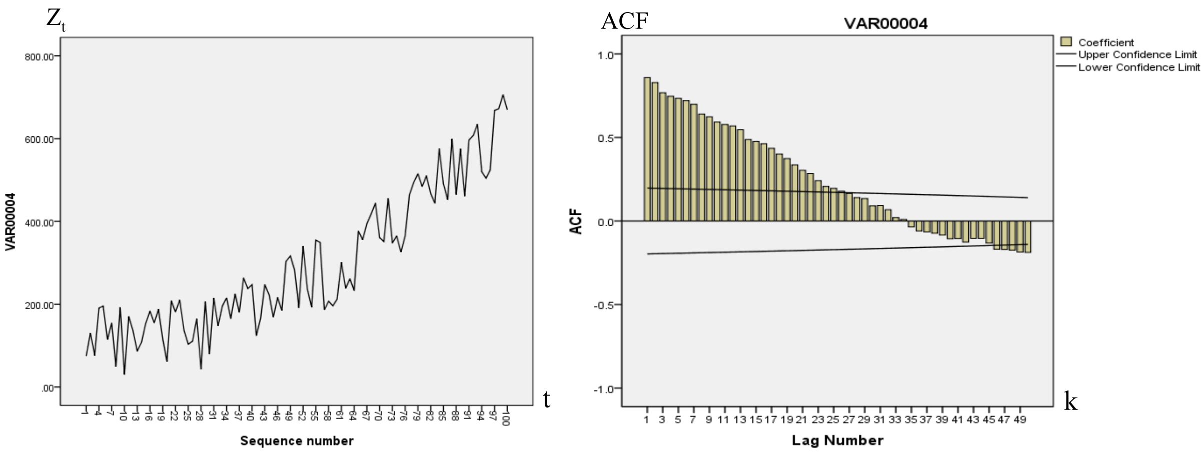
ภาพที่ 2.1 กราฟเส้น และกราฟ ACF ของอนุกรมเวลาที่มีแนวโน้มเชิงเส้นเพิ่มขึ้น

กราฟเส้นของอนุกรมเวลาที่พิจารณาในช่วงเวลา $t = 1,2,...,n$ มีการเคลื่อนไหวหรือการเปลี่ยนแปลงของค่าของอนุกรมเวลาในระยะยาว ในลักษณะเพิ่มขึ้นแบบเชิงเส้น และกราฟ ACF มีค่าแตกต่างไปจากศูนย์อย่างมีนัยสำคัญ และลดลงเข้าสู่ศูนย์อย่างช้า ๆ ดังภาพที่ 2.1



ภาพที่ 2.2 กราฟเส้น และกราฟ ACF ของอนุกรมเวลามีแนวโน้มเชิงเส้นลดลง

กราฟเส้นของอนุกรมเวลาที่พิจารณาในช่วงเวลา $t = 1,2,...,n$ มีการเคลื่อนไหวหรือการเปลี่ยนแปลงของค่าของอนุกรมเวลาในระยะยาว ในลักษณะลดลงแบบเชิงเส้น และกราฟ ACF มีค่าแตกต่างไปจากศูนย์อย่างมีนัยสำคัญ และลดลงเข้าสู่ศูนย์อย่างช้า ๆ ดังภาพที่ 2.2



ภาพที่ 2.3 กราฟเส้น และกราฟ ACF ของอนุกรมเวลาที่มีแนวโน้มไม่ใช่เชิงเส้น

กราฟเส้นของอนุกรมเวลาที่พิจารณาในช่วงเวลา $t = 1,2,..., n$ มีการเคลื่อนไหวหรือการเปลี่ยนแปลงของค่าของอนุกรมเวลาในระยะยาว ในลักษณะเพิ่มขึ้นแบบเชิงเอ็กซ์โปแนนเชียล และกราฟ ACF มีค่าแตกต่างไปจากศูนย์อย่างมีนัยสำคัญ และลดลงเข้าสู่ศูนย์อย่างช้า ๆ ดังภาพที่ 2.3

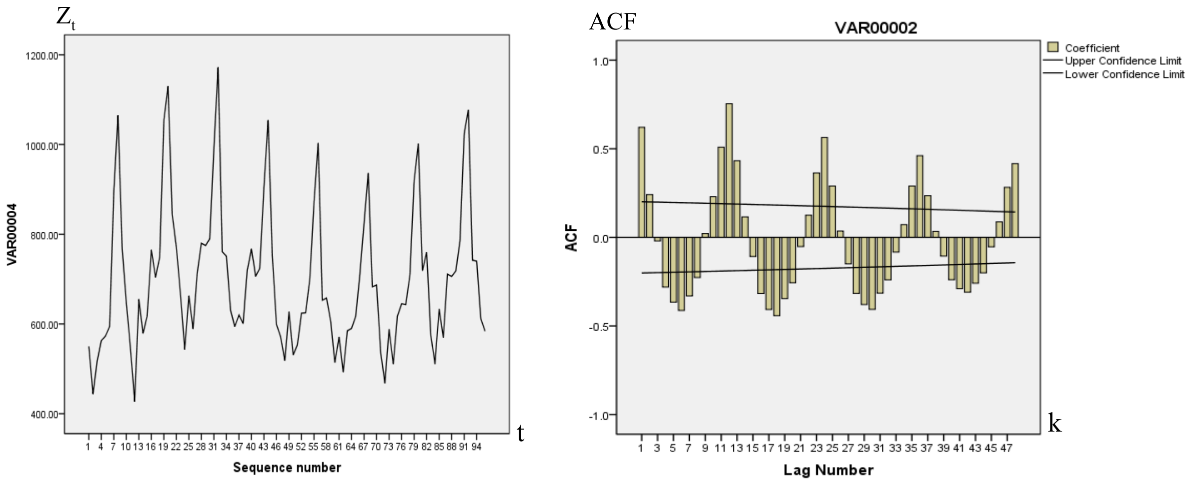
2) การแปรผันตามฤดูกาล (Seasonal Variation: S)

การแปรผันตามฤดูกาล หมายถึง ลักษณะการเคลื่อนไหวเหมือนกันเป็นช่วง ๆ โดยที่แต่ละช่วงนานไม่เกิน 1 ปี เรียกแต่ละช่วงเวลาที่อนุกรมเวลามีลักษณะเหมือนกันนี้ว่า ความยาวหรือคาบของฤดูกาล เช่น ข้อมูลรายเดือน คาบของฤดูกาลอาจเท่ากับ 6 หรือ 12 เดือน ข้อมูลรายไตรมาส คาบของฤดูกาลเท่ากับ 4 ไตรมาส ข้อมูลที่แปรผันตามฤดูกาลเช่นยอดขายเสื้อกันหนาวจะสูงในฤดูหนาวและจะต่ำในฤดูร้อนจำนวนนักท่องเที่ยวพักโรงแรมที่ชายทะเลจะสูงในฤดูร้อนและจะต่ำในฤดูหนาว เป็นต้น

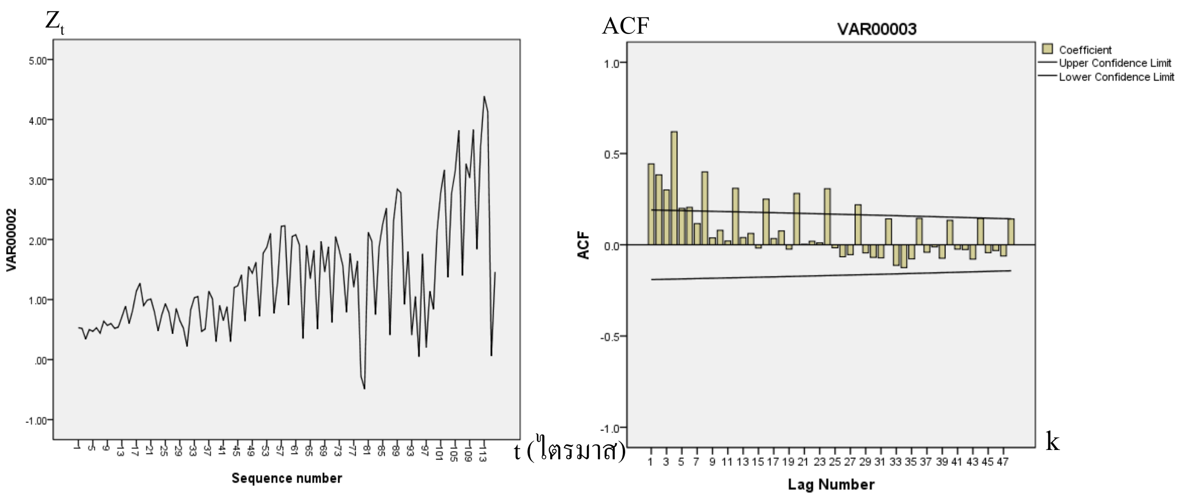
กราฟเส้นของอนุกรมเวลาที่พิจารณาในช่วงเวลา $t = 1,2,..., n$ มีการเปลี่ยนแปลงค่าในลักษณะคล้ายคลึงกันเป็นช่วง ๆ แต่ละช่วงมีขนาดไม่แตกต่างกันมาก โดยแต่ละช่วงนานไม่เกิน 1 ปี ซึ่งได้รับอิทธิพลจากฤดูกาลแสดงว่าข้อมูลมีฤดูกาลดังภาพที่ 2.4 และ 2.5

กราฟแท่งของฟังก์ชันสหสัมพันธ์ในตัวเอง (Autocorrelation Function: ACF) มีค่าแตกต่างไปจากศูนย์อย่างมีนัยสำคัญและลดลงเข้าสู่ศูนย์อย่างรวดเร็วไปแล้วระยะหนึ่ง และกลับมามีค่าแตกต่างไปจากศูนย์อย่างมีนัยสำคัญอีกเป็นช่วง ๆ โดยฤดูกาลมีคาบยาวเท่ากับช่วงห่างของค่า ACF ที่มีค่าแตกต่างไปจากศูนย์อย่างมีนัยสำคัญดังภาพที่ 2.4 และ 2.5

ลักษณะของอนุกรมเวลาที่มีฤดูกาล สามารถพิจารณาจากกราฟ ดังนี้



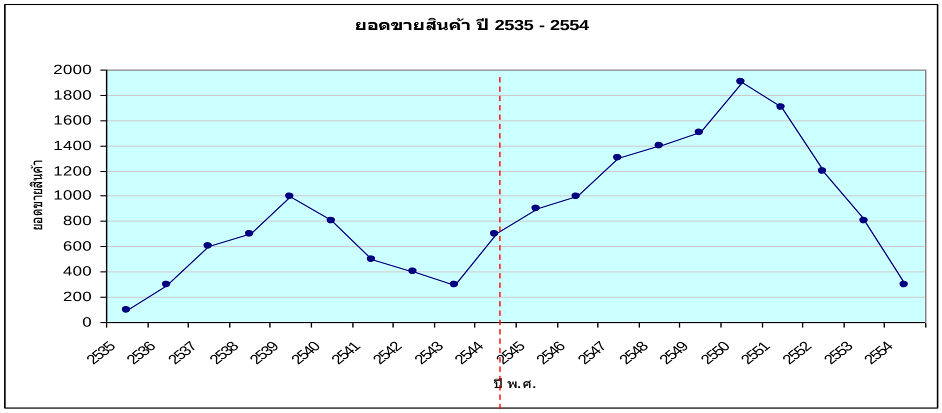
ภาพที่ 2.4 กราฟเส้น และกราฟ ACF ของอนุกรมเวลาที่มีความแปรผันตามฤดูกาล



ภาพที่ 2.5 กราฟเส้น และกราฟ ACF ของอนุกรมเวลาที่มีความแปรผันตามฤดูกาล

3) วัฏจักร (Cycle: C)

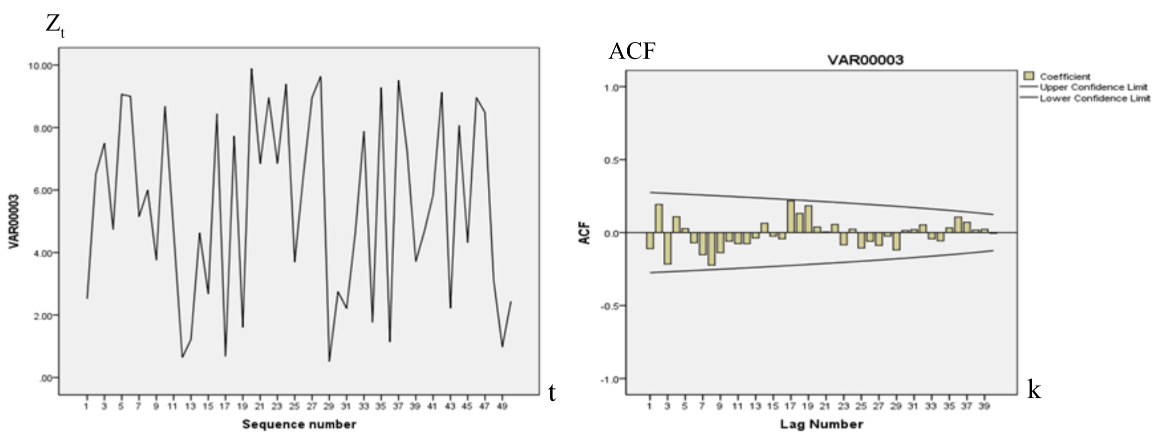
วัฏจักร หมายถึง การเคลื่อนไหวขึ้น ๆ ลง ๆ รอบระดับของแนวโน้มคล้ายกันเป็นช่วง ๆ โดยแต่ละช่วงยาวนานกว่า 1 ปี เช่น 2 - 10 ปี ทั้งความยาวของช่วงของวัฏจักร และขนาดของการเคลื่อนไหวสูงต่ำอาจไม่คงที่เนื่องจากอิทธิพลที่ทำให้เกิดวัฏจักรนั้น เช่น ภาพที่ 2.6 วัฏจักรทางธุรกิจ รอบแรกยาว 9 ปี โดยมีช่วงเศรษฐกิจรุ่งเรือง (ขาขึ้น) 5 ปี และช่วงเศรษฐกิจตกต่ำ (ขาลง) 4 ปี ใน รอบที่สองยาว 11 ปี โดยมีช่วงเศรษฐกิจรุ่งเรือง 7 ปี และช่วงเศรษฐกิจตกต่ำ 4 ปี เป็นต้น ข้อมูลที่มีลักษณะนี้การพยากรณ์ทำได้ยากที่สุด



ภาพที่ 2.6 กราฟเส้นของอนุกรมเวลายอดขายสินค้าชนิดหนึ่งรายปีที่มีวัฏจักรเป็นส่วนประกอบ

4) ความรบกวนสุ่มหรือการแปรผันผิดปกติ (Irregular Variation: I)

หมายถึง การเคลื่อนไหวที่ไม่มีรูปแบบ หรือการเคลื่อนไหวผิดปกติอันเนื่องมาจากเกิดเหตุการณ์ที่ไม่คาดคิด เช่น สงคราม ปฏิวัติ แผ่นดินไหว น้ำท่วม ไฟไหม้ เป็นต้น พิจารณาจาก 2 กราฟ ดังนี้



ภาพที่ 2.7 กราฟเส้น และกราฟ ACF ของอนุกรมเวลาที่มีความรบกวนสุ่มหรือการแปรผันผิดปกติ

กราฟของอนุกรมเวลาที่พิจารณาในช่วงเวลา $t = 1,2,...,n$ มีการเคลื่อนไหวที่ไม่มีรูปแบบและไม่ผันแปรตามเวลา และกราฟของ ACF มีค่าน้อยไม่แตกต่างไปจากศูนย์อย่างมีนัยสำคัญ

2.1.2 การพยากรณ์โดยวิธีการทำให้เรียบ (Smoothing Method)

คือ การพยากรณ์โดยใช้ค่าสังเกตจากอดีตส่วนหนึ่งหรือทั้งหมดในการสร้างสมการ พยากรณ์ ซึ่งน้ำหนักที่ให้กับค่าสังเกตแต่ละค่าจะแตกต่างกัน เหตุผลสำคัญที่มีการใช้วิธีการทำให้เรียบ เนื่องจากอนุกรมเวลาอาจเกิดความผันแปรจากเหตุการณ์ที่ผิดปกติทำให้ไม่เห็นส่วนประกอบของอนุกรมเวลา อื่น ๆ ซึ่งวิธีการทำให้เรียบจะช่วยลดอิทธิพลของความผันแปรดังกล่าวได้ ดังนั้นจึงใช้วิธีสร้างกราฟเพื่อช่วยให้เห็นส่วนประกอบของอนุกรมเวลาแต่ละส่วนจึงปรากฏชัดเจนขึ้น ทำให้สามารถพยากรณ์ค่าของอนุกรมเวลาในอนาคตได้ สำหรับวิธีการทำให้เรียบนั้นมีวิธีการหลายวิธีและการใช้งานจะขึ้นอยู่กับลักษณะของอนุกรมเวลาแยกออกเป็น 3 กลุ่ม ได้แก่ 1. อนุกรมเวลาที่ไม่มีส่วนประกอบของแนวโน้มและฤดูกาล 2. อนุกรมเวลาที่มีเฉพาะส่วนประกอบของแนวโน้ม 3. อนุกรมเวลาที่มีทั้งส่วนประกอบของแนวโน้มและฤดูกาล (วรางคณา เรียนสุทธิ์, 2559) ซึ่งจะมีรายละเอียด ดังนี้

2.1.2.1 อนุกรมเวลาที่ไม่มีส่วนประกอบของแนวโน้มและฤดูกาล

การทำให้เรียบแบบเอ็กซ์โพเนนเชียลชั้นเดียว (Single Exponential) เป็นวิธีการพยากรณ์ที่เหมาะกับข้อมูลที่มีการเคลื่อนไหวอยู่ในระดับคงที่ และเป็นวิธีที่เหมาะสมสำหรับการพยากรณ์ในระยะสั้น ๆ เท่านั้น

สำหรับสมการที่ใช้ในการพยากรณ์ มีดังนี้

$$\hat{Z}_t = \alpha Z_t + \left(1-\alpha\right)\hat{Z}_{t-1}$$

โดยที่ $\hat{Z}_t$ แทน ค่าพยากรณ์ ณ เวลา t

Z_t แทน ค่าสังเกตหรือข้อมูล ณ เวลา t

α แทน ค่าคงที่การทำให้เรียบระหว่างข้อมูลกับค่าพยากรณ์โดยที่ $0 \le \alpha \le 1$

เมื่อกำหนดให้ค่าเริ่มต้น คือ $\bar{Z}_0\left(1\right) = \bar{Z} = \dfrac{1}{t}\sum_{i=1}^{t} Z_i$

2.1.2.2 อนุกรมเวลาที่มีเฉพาะส่วนประกอบของแนวโน้ม

วิธีของโฮลต์ (Holt's Method) มีความเหมาะสมกับอนุกรมเวลาที่มีแนวโน้มเป็นเส้นตรงและไม่มีส่วนประกอบของฤดูกาล เหมาะกับการพยากรณ์ในระยะสั้นถึงพยากรณ์ในระยะปานกลาง มีค่าคงที่การทำให้เรียบ 2 ค่า คือ ค่าคงที่การทำให้เรียบของค่าระดับ (Level: α) และค่าคงที่การปรับเรียบของค่าความชัน (Trend: β)

สำหรับตัวแบบของข้อมูลอนุกรมเวลาที่ใช้กับการพยากรณ์นี้สามารถเขียนตัวแบบได้ดังนี้

$$Z_t = b_0 + b_1 t + a_t$$

โดยที่ Z_t แทน ข้อมูลหรือค่าสังเกต ณ เวลา t

b_0, b_1 แทน ค่าพารามิเตอร์ของตัวแบบ

a_t แทน ค่าความคลาดเคลื่อน ณ เวลา t

สำหรับสมการในการพยากรณ์ของโฮลท์ที่ τ หน่วยเวลาล่วงหน้า คือ

$$\hat{Z}_t(\tau) = \hat{Z}_{t+l} = \hat{b}_{0t} + \hat{b}_{1t}\tau$$

โดยที่ $\hat{Z}_t(\tau)$ แทน ค่าพยากรณ์ ณ เวลา $t+\tau$

$\hat{b}_{0t}$ และ $\hat{b}_{1t}$ แทน ค่าประมาณค่าพารามิเตอร์ b_0 และ b_1 ณ เวลา t

เมื่อ $$\hat{b}_{0t} = \alpha Z_t + (1-\alpha)\left(\hat{b}_{0(t-1)} + \hat{b}_{1(t-1)}\right)$$

$$\hat{b}_{1t} = \beta\left(\hat{b}_{0t} - \hat{b}_{0(t-1)}\right) + (1-\beta)\hat{b}_{1(t-1)}$$

α แทน ค่าคงที่การทำให้เรียบระหว่างข้อมูลกับค่าพยากรณ์โดยที่ $0 \le \alpha \le 1$

β แทน ค่าคงที่การทำให้เรียบระหว่างแนวโน้มจริงกับค่าประมาณของแนวโน้มโดยที่ $0 \le \beta \le 1$

สำหรับการคำนวณค่าเริ่มต้น ของ $\hat{b}_{0t}, \hat{b}_{1t}$ สามารถคำนวณได้ดังนี้

1. ค่า $\hat{b}_{00} = Z_1$ หรือ $\hat{b}_{00} = \bar{Z} = \sum_{t=1}^{n} Z_t / n$

2. ค่า $\hat{b}_{10} = Z_2 - Z_1$ หรือ $\hat{b}_{10} = \dfrac{Z_4 - Z_1}{3}$ หรือ $\hat{b}_{10} = \dfrac{Z_n - Z_1}{n-1}$

ส่วนค่า α, β ผู้วิจัยอาจเป็นผู้กำหนดเอง หรือใช้โปรแกรมสำเร็จรูปค้นหาค่าของ α, β ที่ทำให้ความคลาดเคลื่อนของการพยากรณ์มีค่าต่ำสุด

2.1.2.3 อนุกรมเวลาที่มีทั้งส่วนประกอบของแนวโน้มและฤดูกาล

วิธีของวินเทอร์ (Winter's Method) วิธีการพยากรณ์นี้เหมาะสำหรับอนุกรมเวลาที่มีแนวโน้มและอิทธิพลของฤดูกาล เหมาะกับการพยากรณ์ในระยะสั้นจนถึงการพยากรณ์ในระยะปานกลางเหมาะกับข้อมูลรายสัปดาห์ รายเดือน หรือรายไตรมาส สำหรับวิธีการพยากรณ์นี้ใช้ค่าปรับให้เรียบ 3 ค่า ได้แก่ α เป็นค่าคงที่การทำให้เรียบระหว่างข้อมูลกับค่าพยากรณ์ β เป็นค่าคงที่การทำให้เรียบระหว่างแนวโน้มจริงกับค่าประมาณของแนวโน้ม γ เป็นค่าคงที่การทำให้เรียบระหว่างค่าฤดูกาลจริงกับค่าประมาณฤดูกาล ที่มีค่าอยู่ระหว่าง 0 ถึง 1 สำหรับตัวแบบของข้อมูลอนุกรมเวลาที่ใช้กับการพยากรณ์นี้มีทั้งตัวแบบเชิงคูณและตัวแบบเชิงบวก ซึ่งแต่ละตัวแบบมีรายละเอียดดังนี้ (ยุภาพร ตงประสิทธิ์, 2559)

1) ตัวแบบเชิงคูณ

การแปรผันตามฤดูกาลแต่ละช่วงมีลักษณะแตกต่างกันมาก เมื่อระดับอนุกรมเวลาเพิ่มขึ้นการแปรผันตามฤดูกาลเพิ่มขึ้นด้วย หรือในทางกลับกัน การแปรผันฤดูกาลลดลงเมื่อระดับของอนุกรมเวลาเพิ่มขึ้น อนุกรมเวลานี้ไม่มีแนวโน้มทั้งภายในฤดูกาลและระหว่างฤดูกาล โดยมีตัวแบบ ดังนี้

$$Z_t = \left(b_0 + b_1 t\right) C_t + a_t$$

โดยที่ Z_t แทน ข้อมูลหรือค่าสังเกต ณ เวลา t

b_0 แทน ระดับของข้อมูล

b_1 แทน ความลาดชัน

C_t แทน การผันแปรตามฤดูกาลที่เวลา t

a_t แทน ความคลาดเคลื่อน ที่มีค่าเฉลี่ยเท่ากับศูนย์ ไม่มีสหสัมพันธ์ในตัวและความแปรปรวนคงที่ ณ เวลา t

สำหรับสมการในการพยากรณ์ที่ τ หน่วยเวลาล่วงหน้า คือ

$$\hat{Z}_t\left(\tau\right) = \left(\hat{b}_{0t} + \hat{b}_{1t}\tau\right)\hat{C}_{t-L+\tau} \quad , \tau = 1, 2, ..., L$$

โดยที่ $\hat{b}_{0t}$ แทน ค่าปรมาณของระดับข้อมูลที่เวลา t โดยที่

$$\hat{b}_{0t} = \alpha \frac{Z_t}{\hat{C}_{t-L}} + \left(1-\alpha\right)\left[\hat{b}_{0(t-1)} + \hat{b}_{1(t-1)}\right]$$

เมื่อ α แทน ค่าคงที่ทำให้เรียบค่าที่หนึ่ง; $0 < \alpha < 1$

Z_t แทน ค่าข้อมูลจริงที่เวลา t

$\hat{C}_{t-L}$ แทน ค่าประมาณของดัชนีฤดูกาลที่เวลา $t-L$

$\hat{b}_{0(t-1)}$ แทน ค่าประมาณของระดับข้อมูลที่เวลา $t-1$

$\hat{b}_{1(t-1)}$ แทน ค่าประมาณของความลาดชันที่เวลา $t-1$

โดยที่ $\hat{b}_{1t}$ แทน ค่าประมาณความลาดชันที่เวลา t โดยที่

$$\hat{b}_{1t} = \beta\left[\hat{b}_{0t} - \hat{b}_{0(t-1)}\right] + \left(1-\beta\right)\hat{b}_{1(t-1)}$$

เมื่อ β แทน ค่าคงที่ทำให้เรียบค่าที่สอง ; $0 < \beta < 1$

$\hat{b}_{0t}$ แทน ค่าประมาณของระดับข้อมูลที่เวลา t

$\hat{b}_{0(t-1)}$ แทน ค่าประมาณของระดับข้อมูลที่เวลา $t-1$

$\hat{b}_{1(t-1)}$ แทน ค่าประมาณของความลาดชันที่เวลา $t-1$

โดยที่ $\hat{C}_t$ แทน ค่าประมาณดัชนีฤดูกาลที่เวลา t

$$\hat{C}_t = \gamma\left[\frac{Z_t}{\hat{b}_{0t}}\right] + (1-\gamma)\hat{C}_{t-L}$$

เมื่อ γ แทน ค่าคงที่ทำให้เรียบค่าที่สาม ; $0 < \gamma < 1$

Z_t แทน ค่าข้อมูลจริงที่เวลา t

$\hat{b}_{0t}$ แทน ค่าประมาณของระดับข้อมูลที่เวลา t

$\hat{C}_{t-L}$ แทน ค่าประมาณของดัชนีฤดูกาลที่เวลา $t-L$

สำหรับการคำนวณค่าเริ่มต้น ของ $\hat{b}_{0t}$, $\hat{b}_{1t}$, $\hat{C}_t$ ที่เวลา $t = 1, 2, 3, \ldots\ldots, L$ สามารถคำนวณได้ดังนี้

ให้อนุกรมเวลามีข้อมูล mL ค่า, เมื่อ m คือ จำนวนฤดูกาล และ L คือ ความยาวของฤดูกาล หาค่าเฉลี่ยของข้อมูลในฤดูกาลที่ j จากสูตร

$$\bar{Z}_i = \frac{1}{L}\sum_{j=1}^{L} Z_i \quad ; i = 1, 2, 3, \ldots\ldots, m$$

1. ค่าเริ่มต้นของ $\hat{b}_{1t}$ คือ $\hat{b}_{10}$ ซึ่ง $\hat{b}_{10} = \dfrac{\bar{Z}_m - \bar{Z}_1}{(m-1)L}$

2. ค่าเริ่มต้นของ $\hat{b}_{0t}$ คือ $\hat{b}_{00}$ ซึ่ง $\hat{b}_{00} = \bar{Z}_1 - \left(\dfrac{L+1}{2}\right)\hat{b}_{10}$

3. ค่าเริ่มต้นของ $\hat{C}_t$ คือ $\hat{C}_j$ ที่เวลา t จากสูตร

- หาค่า $\tilde{C}_{ij} = \tilde{C}_t = \dfrac{Z_t}{\bar{Z}_i - \left[\dfrac{L+1}{2} - j\right]\hat{b}_{10}}$; $t = 1, 2, 3, \ldots\ldots, mL$

เมื่อ $t = (i-1)L + j$; $i = 1, 2, \ldots, m$ และ $j = 1, 2, \ldots., L$

- หาค่าเฉลี่ยของค่าประมาณของดัชนีฤดูกาลที่หน่วยเวลา i นับจากต้นฤดูกาล จากสูตร

$$\bar{C}_j = \frac{1}{m}\sum_{i=1}^{m} \tilde{C}_{(i-1)L+j} = \frac{1}{m}\sum_{i=1}^{m} \tilde{C}_{ij} \quad ; i = 1, 2, \ldots, m \text{ และ } j = 1, 2, \ldots., L$$

- นำค่า $\bar{C}_j$ มาปรับให้ผลรวมของดัชนีฤดูกาลทั้ง L ค่า ให้มีค่าเท่ากับ L จะได้ค่าประมาณเริ่มต้นของดัชนีฤดูกาล ดังนี้

$$\hat{C}_j = \bar{C}_j\left(\frac{L}{\sum_{j=1}^{L}\bar{C}_j}\right) \quad ; j = 1, 2, \ldots, L$$

2) ตัวแบบเชิงบวก

การผันแปรตามฤดูกาลแต่ละช่วงจะลักษณะเช่นเดิม หรือแตกต่างจากเดิมไม่มากนักตามอิทธิพลจากฤดูกาล อนุกรมเวลานี้ไม่มีแนวโน้มภายในฤดูกาล แต่มีแนวโน้มระหว่างฤดูกาล โดยมีตัวแบบ ดังนี้

$$Y_t = b_0 + b_1 t + C_t + a_t$$

โดยที่ Z_t แทน ข้อมูลหรือค่าสังเกต ณ เวลา t

b_0 แทน ระดับของข้อมูล

b_1 แทน ความลาดชัน

C_t แทน การผันแปรตามฤดูกาลที่เวลา t

a_t แทน ความคลาดเคลื่อน ที่มีค่าเฉลี่ยเท่ากับศูนย์ ไม่มีสหสัมพันธ์ในตัวและความแปรปรวนคงที่ ณ เวลา t

สำหรับสมการในการพยากรณ์ที่ τ หน่วยเวลาล่วงหน้า คือ

$$\hat{Z}_t(\tau) = \hat{b}_{0t} + \hat{b}_{1t}\tau + \hat{C}_{t-L+\tau} \quad , \tau = 1, 2, ..., L$$

โดยที่ $\hat{b}_{0t}$ แทน ค่าปรมาณของระดับข้อมูลที่เวลา t โดยที่

$$\hat{b}_{0t} = \alpha\left[Z_t - \hat{C}_{t-L}\right] + (1-\alpha)\left[\hat{b}_{0(t-1)} + \hat{b}_{1(t-1)}\right]$$

เมื่อ α แทน ค่าคงที่ทำให้เรียบค่าที่หนึ่ง; $0 < \alpha < 1$

Z_t แทน ค่าข้อมูลจริงที่เวลา t

$\hat{C}_{t-L}$ แทน ค่าประมาณของดัชนีฤดูกาลที่เวลา $t-L$

$\hat{b}_{0(t-1)}$ แทน ค่าประมาณของระดับข้อมูลที่เวลา $t-1$

$\hat{b}_{1(t-1)}$ แทน ค่าประมาณของความลาดชันที่เวลา $t-1$

โดยที่ $\hat{b}_{1t}$ แทน ค่าประมาณความลาดชันที่เวลา t

$$\hat{b}_{1t} = \beta\left[\hat{b}_{0t} - \hat{b}_{0(t-1)}\right] + (1-\beta)\hat{b}_{1(t-1)}$$

เมื่อ β แทน ค่าคงที่ทำให้เรียบค่าที่สอง ; $0 < \beta < 1$

$\hat{b}_{0t}$ แทน ค่าประมาณของระดับข้อมูลที่เวลา t

$\hat{b}_{0(t-1)}$ แทน ค่าประมาณของระดับข้อมูลที่เวลา $t-1$

$\hat{b}_{1(t-1)}$ แทน ค่าประมาณของความลาดชันที่เวลา $t-1$

โดยที่ $\hat{C}_t$ แทน ค่าประมาณดัชนีฤดูกาลที่เวลา t

$$\hat{C}_t = \gamma\left[Z_t - \hat{b}_{ot}\right] + (1-\gamma)\hat{C}_{t-L}$$

เมื่อ γ แทน ค่าคงที่ทำให้เรียบค่าที่สาม ; $0 < \gamma < 1$

Z_t แทน ค่าข้อมูลจริงที่เวลา t

$\hat{b}_{0t}$ แทน ค่าประมาณของระดับข้อมูลที่เวลา t

$\hat{C}_{t-L}$ แทน ค่าประมาณของดัชนีฤดูกาลที่เวลา $t-L$

สำหรับการคำนวณค่าเริ่มต้น ของ $\hat{b}_{0t},\ \hat{b}_{1t},\ \hat{C}_t$ ที่เวลา $t=1,2,3,.....,L$ ใช้วิธีกำลังสองน้อยที่สุด จะได้

ให้ m จำนวนฤดูกาลของข้อมูล

L ความยาวของฤดูกาล

mL จำนวนข้อมูลทั้งหมด

1. ค่าเริ่มต้นของ $\hat{b}_{00}$ และ $\hat{b}_{10}$ หาจากการแก้สมการ ดังนี้

$$mL\hat{b}_0+\frac{mL(mL+1)}{2}\hat{b}_1 = \sum_{t=1}^{mL} Z_t$$

$$\frac{m(m-1)L^2}{2}\hat{b}_0+\frac{m(m-1)L^2\left[(4m+1)L+3\right]}{12}\hat{b}_1 = \sum_{t=1}^{mL} tZ_t-\sum_{j=1}^{L}\sum_{i=1}^{m} jZ_{(i-1)L+j}$$

2. ค่าเริ่มต้นของ $\hat{C}_t$ คือ $\hat{C}_j$ จากสมการ ดังนี้

$$\hat{C}_j=\left[\frac{1}{m}\sum_{i=1}^{m} Z_{(i=1)L+j}\right]-\hat{b}_0-\hat{b}_1\left[j+\frac{(m-1)L}{2}\right] \quad ;\ j=1,2,....,L$$

ส่วนค่า $\alpha,\ \beta$ และ γ ผู้วิจัยอาจเป็นผู้กำหนดเอง หรือใช้โปรแกรมสำเร็จรูปค้นหาค่าของ $\alpha,\ \beta$ และ γ ที่ทำให้ความคลาดเคลื่อนของการพยากรณ์มีค่าต่ำสุด

2.1.3 วิธีของบ็อกซ์และเจนกินส์ (Box-Jenkins's Method)

เป็นวิธีวิเคราะห์อนุกรมเวลาที่พัฒนาโดย George E.P. Box และ Gwilym M. Jenkins เมื่อปี ค.ศ. 1970 โดยนำเสนอตัวแบบ ARIMA และดำเนินการปรับปรุงในปี ค.ศ. 1994 โดยใช้หลักการ คือ เลือกตัวแบบที่ใช้ในการพยากรณ์โดยพิจารณาจาก ลักษณะของสหสัมพันธ์ในตัวเองและสหสัมพันธ์ในตัวเองบางส่วนของอนุกรมเวลา และกำหนดตัวแบบสโตแคสติคที่เป็นไปได้ที่มีลักษณะของสหสัมพันธ์ในตัวเองและสหสัมพันธ์ในตัวเองบางส่วนที่เหมือนกัน วิธีการนี้อาจมีตัวแบบที่เป็นไปได้มากกว่า 1 ตัวแบบ ดังนั้นจึงจำเป็นต้องมีขั้นตอนของการตรวจสอบเพื่อให้ได้ตัวแบบที่เหมาะสมสำหรับนำไปใช้ในการพยากรณ์ต่อไป (ยุภาพร ตงประสิทธิ์, 2559)

อนุกรมเวลาที่จะนำมาใช้เพื่อการพยากรณ์มักมีลักษณะการเคลื่อนไหวที่แตกต่างกัน วิธีของบ็อกซ์และเจนกินส์ได้แบ่งออกเป็น 2 ลักษณะ ตามการเคลื่อนไหวของอนุกรมเวลาดังนี้

2.1.3.1 อนุกรมเวลาที่สเตชันนารี (Stationary series)

เป็นอนุกรมเวลาที่มีค่าเฉลี่ยและค่าความแปรปรวนของ Y_t คงที่ นอกจากนั้นค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเองที่ช่วงเวลาห่างกัน k หน่วย ขึ้นอยู่กับ k อย่างเดียวหรือ ไม่ขึ้นกับเวลา t อนุกรมเวลาที่จะกำหนดรูปแบบ $ARMA(p,q)$ ให้ต้องเป็นอนุกรมเวลที่เป็นสเตชันนารี (ทรงศิริ แต้สมบัติ, 2549) การพิจารณาว่าอนุกรมเวลาชุดใดชุดหนึ่ง สเตชันนารีหรือไม่ พิจารณาจาก

ก. การพล็อตกราฟของอนุกรมเวลา แล้วดูการเคลื่อนไหวของอนุกรมเวลา ถ้าการเคลื่อนไหวของอนุกรมเวลามีแนวโน้ม หรือฤดูกาล หรือมีทั้ง 2 อย่าง แสดงว่าอนุกรมเวลาชุดนั้นไม่สเตชันนารี

ข. พิจารณาคอเรลโรแกรมของ r_k มีลักษณะลดลงอย่างรวดเร็ว เมื่อ k มีค่าเพิ่มขึ้น แต่ถ้า r_k มีลักษณะลดลงค่อนข้างช้าหรือมีค่าค่อนข้างสูงที่ $k = L, 2L, 3L, \dots$ เมื่อ L เป็นจำนวนฤดูกาล แสดงว่าอนุกรมเวลานั้นไม่สเตชันนารี (สมเกียรติ เกตุเอี่ยม, 2546)

2.1.3.2 อนุกรมเวลาที่ไม่สเตชันนารี (Nonstationary time series)

เป็นอนุกรมเวลาที่ค่าเฉลี่ยและค่าความแปรปรวนของ Y_t ไม่คงที่ หรือค่าใดค่าหนึ่งไม่คงที่ และค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเองที่ช่วงเวลาห่างกัน k หน่วย ไม่ขึ้นอยู่กับเฉพาะ k อย่างเดียว แต่ขึ้นอยู่กับเวลา t ด้วย สำหรับอนุกรมเวลาที่มีแนวโน้ม หรือฤดูกาล การหารูปแบบ $ARMA(p,q)$ ให้กับอนุกรมเวลาที่ไม่สเตชันนารีทำไม่ได้ในทันที ต้องแปลงอนุกรมเวลาที่ไม่สเตชันนารีให้เป็นอนุกรมเวลาใหม่ที่มีคุณสมบัติสเตชันนารีก่อน รายละเอียดของการแปลงอนุกรมเวลามีดังนี้

ก. หาผลต่าง ถ้าอนุกรมเวลามีการเคลื่อนไหวเนื่องจากแนวโน้ม จะแปลงอนุกรมเวลาเดิมเป็นอนุกรมเวลาใหม่ที่ไม่มีแนวโน้ม โดย $Z_t = \nabla^d Y_t$ ซึ่ง d เป็นจำนวนครั้งของการหาผลต่างจำนวนครั้งที่หาผลต่างขึ้นกับว่า เมื่อหาผลต่างแล้วอนุกรมเวลาใหม่เป็นสเตชันนารีแล้วหรือไม่ ถ้ายังไม่เป็นสเตชันนารีต้องหาผลต่างต่อไป

ข. หาผลต่างฤดูกาล ถ้าอนุกรมเวลามีการเคลื่อนไหวเนื่องจากฤดูกาลจะแปลงอนุกรมเวลาเดิมเป็นอนุกรมเวลาใหม่ที่ไม่มีฤดูกาล โดย $Z_t = \nabla_L{}^D Y_t$ ซึ่ง D เป็นจำนวนครั้งของการหาผลต่างฤดูกาล และ L เป็นจำนวนฤดูกาลต่อปี จำนวนครั้งที่หาผลต่างฤดูกาลขึ้นกับว่า เมื่อหาผลต่างฤดูกาลแล้วอนุกรมเวลาใหม่เป็นสเตชันนารีแล้วหรือไม่ ถ้ายังไม่เป็นสเตชันนารีต้องหาผลต่างฤดูกาลต่อไป

ค. หาผลต่างและผลต่างฤดูกาล กรณีที่อนุกรมเวลามีการเคลื่อนไหวเนื่องจากแนวโน้มและฤดูกาลจะแปลงอนุกรมเวลาเดิมเป็นอนุกรมเวลาใหม่ที่ไม่มีแนวโน้มและฤดูกาลโดยหาผลต่างและผลต่างฤดูกาล ซึ่ง $Z_t = \nabla^d \nabla_L^D Y_t$ จำนวนครั้งที่หาผลต่างและผลต่างฤดูกาลที่กำหนดด้วย d และ D ตามลำดับขึ้นอยู่กับว่าอนุกรมเวลาใหม่ที่สร้างขึ้นเป็นสเตชันนารีแล้วหรือไม่

ง. หากอนุกรมเวลาที่แปลงแล้ว ความผันแปรของอนุกรมเวลาไม่คงที่ นั่นคือค่าความแปรปรวนของ Y_t ไม่คงที่จะแปลงให้เป็นอนุกรมเวลาใหม่ด้วย $Z_t = \ln(Y_t)$ หรือ $Z_t = \sqrt{Y_t}$ อนุกรมเวลาที่แปลงแล้วจะมีค่าความแปรปรวนคงที่ นอกจากนั้นผลจากการแปลงอนุกรมเวลาที่มีรูปแบบแนวโน้มฤดูกาลแบบคูณจะได้อนุกรมเวลาใหม่ที่มีรูปแบบแนวโน้มฤดูกาลแบบบวก (ทรงศิริ แต้สมบัติ, 2549)

2.1.3.3 สหสัมพันธ์ในตัวเอง (Autocorrelation: ACF) เมื่อเวลาย้อนหลังไป k หน่วยเวลา

สหสัมพันธ์ระหว่างตัวแปรสุ่มที่หน่วยเวลาห่างกัน k หน่วยของกระบวนการเดียวกันเรียกว่า สหสัมพันธ์ในตัวเอง (Autocorrelation) เป็นมาตรวัดความสัมพันธ์ระหว่างตัวแปรสุ่ม Z_t และ Z_{t-k} ซึ่ง ห่างกัน k หน่วยเวลา ในอนุกรมเวลาเดียวกัน เขียนแทนด้วย ρ_k เรียกว่า สัมประสิทธิ์สหสัมพันธ์ในตัวเองเมื่อเวลาย้อนหลังไป k หน่วย

ในทางปฏิบัติ ค่า ρ_k จะประมาณจากอนุกรมเวลาตัวอย่างขนาด n ค่า แทนด้วย $Z_1, Z_2, ..., Z_n$ และเรียกว่า สหสัมพันธ์ในตัวเองของตัวอย่างที่ช่วงเวลาห่างกัน k หน่วยเขียนแทนด้วย r_k คำนวณได้จาก (ยุภาพร ตงประสิทธิ์, 2559)

$$r_k = \frac{\sum_{t=1}^{n-k}(Z_t - \bar{Z})(Z_{t+k} - \bar{Z})}{\sum_{t=1}^{n}(Z_t - \bar{Z})^2} = \frac{\sum_{t=k+1}^{n}(Z_t - \bar{Z})(Z_{t-k} - \bar{Z})}{\sum_{t=1}^{n}(Z_t - \bar{Z})^2}, k = 0, 1, 2, 3, ... \; ; -1 \le r_k \le 1$$

โดยที่ Z_t คือ ข้อมูลหรือค่าสังเกต ณ เวลา t

k คือ จำนวนช่วงเวลาที่ข้อมูลอยู่ห่างกัน $k = 0,1,2,3,...$

n คือ จำนวนข้อมูลทั้งหมด

$\bar{Z}$ คือ ค่าเฉลี่ยของข้อมูลทั้งหมดโดยที่ $\bar{Z} = \frac{1}{n}\sum_{t=1}^{n} Z_t$

2.1.3.4 สหสัมพันธ์ในตัวเองบางส่วน (Partial Autocorrelation: PACF) เมื่อเวลาย้อนหลังไป k หน่วยเวลา

การพิจารณาสหสัมพันธ์ระหว่างตัวแปรสุ่ม Z_t และ Z_{t-k} อาจเป็นไปได้ว่า สหสัมพันธ์ดังกล่าวเกิดจากสหสัมพันธ์ระหว่างสองตัวแปรนี้กับตัวแปรที่เหลือ คือ $Z_{t-1}, Z_{t-2}, ..., Z_{t-k+1}$ ด้วย ดังนั้น เพื่อให้ได้สหสัมพันธ์ที่ไม่รวมความสัมพันธ์ระหว่าง Z_t และ Z_{t-k} กับ $Z_{t-1}, Z_{t-2}, ..., Z_{t-k+1}$ สามารถวัดด้วย "สหสัมพันธ์ในตัวเองบางส่วนระหว่างตัวแปรสุ่ม Z_t และ Z_{t-k}"

สหสัมพันธ์ในตัวเองบางส่วนระหว่างตัวแปรสุ่ม Z_t และ Z_{t-k} เขียนแทนด้วย φ_{kk} หมายถึง ความสัมพันธ์ระหว่างตัวแปร Z_t และ Z_{t-k} ที่ไม่รวมความสัมพันธ์ระหว่างตัวแปรทั้งสองนี้กับตัวแปร $Z_{t-1}, Z_{t-2}, ..., Z_{t-k+1}$ (ยุภาพร ตงประสิทธิ์, 2559)

ค่าประมาณของ φ_{kk} คือ $\hat{\phi}_{kk}$ หาได้จาก

$$\hat{\phi}_{kk} = \begin{cases} r_1 & ,k=1 \\ \dfrac{r_k - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} r_j} & ,k=2,3,4,\ldots \end{cases}$$

เมื่อ $\hat{\varphi}_{kj} = \hat{\varphi}_{k-1,j} - \hat{\varphi}_{kk}\hat{\varphi}_{k-1,k-j} \quad j=1,2,3,\ldots,k-1, \;, \; k=2,3,4,\ldots$

2.1.3.5 กำหนดตัวแบบ (Identification) พิจารณาจากลักษณะของสหสัมพันธ์ในตัวเอง (ρ_k) และสหสัมพันธ์ในตัวเองบางส่วน (ϕ_{kk}) ของอนุกรมเวลาที่พิจารณา โดยกำหนดตัวแบบสโตแคสติคที่มีลักษณะของสหสัมพันธ์ในตัวเอง (ACF) และสหสัมพันธ์ในตัวเองบางส่วน (PACF) ใกล้เคียงกันหรือเหมือนกัน ซึ่งอาจจะกำหนดได้มากกว่า 1 ตัวแบบ (ยุภาพร ตงประสิทธิ์, 2559) ซึ่งมีตัวแบบดังนี้

ตัวแบบที่ 1 กระบวนการถดถอยในตัวเอง (Autoregressive Process of Order p : $AR(p)$)

คือ กระบวนการเชิงเส้นทั่วไปที่อธิบายตัวแปรสุ่ม Z_t ในเทอมของค่าสังเกตในอดีต และความรบกวนสุ่มในปัจจุบัน a_t โดยใช้น้ำหนัก π_j กระบวนการถดถอยในตัวเองอันดับ p **(Autoregressive Process of Order p :** $AR(p)$) คือ กระบวนการเชิงเส้นทั่วไปที่อธิบายตัวแปรสุ่ม Z_t ในเทอมของค่าสังเกตในอดีตจำนวน p ค่า และความรบกวนสุ่มในปัจจุบัน a_t โดยการกำหนดน้ำหนัก $\pi_j = \phi_j$, $j=1,2,\ldots,p$ มีตัวแบบดังนี้

$$(1-\phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)(Z_t - \mu) = a_t$$

ลักษณะของ ρ_k และ ϕ_{kk} ของกระบวนการ $AR(p)$ เป็นดังนี้

1) ρ_k มีจำนวนอนันต์ และลดลงเข้าสู่ศูนย์แบบเอกซ์โปเนนเชียล หรือลดลงเข้าสู่ศูนย์แบบคลื่นไซน์

2) ϕ_{kk} Cut-off เมื่อ $k > p$ หรือ $AR(p)$ มีความจำในอดีตเพียง p หน่วยเวลา และ $\phi_{kk} \neq 0$, เมื่อ $k \leq p$

กระบวนการ $AR(1)$ มีตัวแบบดังนี้ $(1-\phi_1 B)(Z_t - \mu) = a_t$ ลักษณะของ ρ_k มีค่าเข้าใกล้ 0 เมื่อ k มีค่ามากขึ้น และ $\phi_{kk} \neq 0, k=1$ และ $\phi_{kk} = 0, k>1$

กระบวนการ $AR(2)$ มีตัวแบบดังนี้ $(1-\phi_1 B - \phi_2 B^2)(Z_t - \mu) = a_t$ ลักษณะของ ρ_k มีค่าเข้าใกล้ 0 เมื่อ k มีค่ามากขึ้น และ $\phi_{kk} \neq 0, k=1,2$ และ $\phi_{kk} = 0, k>2$

ตัวแบบที่ 2 กระบวนการค่าเฉลี่ยเคลื่อนที่ (Moving Average Process of Order q : $MA(q)$)

คือ กระบวนการเชิงเส้นทั่วไปที่อธิบายค่า Z_t ในเทอมของความรบกวนสุ่ม a_t ย้อนหลังไปในอดีตจำนวน q เทอม เมื่อกำหนดให้น้ำหนัก $\psi_j = -\theta_j$ ค่า Z_t ได้มาจากการให้น้ำหนัก $1, -\theta_1, -\theta_2, \ldots$ กับตัวแปรสุ่มอิสระ $a_t, a_{t-1}, a_{t-2}, \ldots$ และค่า Z_{t+1} ได้มาจากการให้น้ำหนักชุดเดียวกันนี้กับตัวแปรสุ่มอิสระ $a_{t+1}, a_{t+2}, \ldots, a_{t+q}$ นั่นคือ การเลื่อนน้ำหนักชุดนี้ไปทีละหนึ่งหน่วยเวลา จึงเรียกกระบวนการนี้ว่า กระบวนการค่าเฉลี่ยเคลื่อนที่ ซึ่งมีตัวแบบดังนี้

$$Z_t - \mu = (1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q) a_t$$

ลักษณะของ ρ_k และ ϕ_{kk} ของกระบวนการ $MA(q)$ เป็นดังนี้

1) ρ_k Cut-off, เมื่อ $k > q$ หรือกระบวนการมีความจำเพียง q หน่วยเวลา และ $\rho_k \neq 0$, เมื่อ $k \leq q$

2) ϕ_{kk} มีจำนวนอนันต์ และลดลงเข้าสู่ศูนย์แบบเอกซ์โปเนนเชียล หรือลดลงเข้าสู่ศูนย์แบบคลื่นไซน์

กระบวนการ $MA(1)$ มีตัวแบบดังนี้ $Z_t - \mu = (1-\theta_1 B)a_t$ ลักษณะของ $\rho_{kk} \neq 0, k=1$ และ $\rho_{kk} = 0, k>1$ และ ϕ_{kk} มีค่าเข้าใกล้ 0 เมื่อ k มีค่ามากขึ้น

กระบวนการ $MA(2)$ มีตัวแบบดังนี้ $Z_t - \mu = (1-\theta_1 B - \theta_2 B^2)a_t$ ลักษณะของ $\rho_{kk} \neq 0, k=1,2$ และ $\rho_{kk} = 0, k>2$ และ ϕ_{kk} มีค่าเข้าใกล้ 0 เมื่อ k มีค่ามากขึ้น

ตัวแบบที่ 3 กระบวนการผสมถดถอยในตัวเองกับค่าเฉลี่ยเคลื่อนที่อันดับ p และ q (Mixed Autoregressive-Moving Average Process of Order p and q : $ARMA(p,q)$)

คือ กระบวนการที่ค่าสังเกตในปัจจุบัน Z_t เป็นทั้งฟังก์ชันของค่าสังเกตในอดีต p หน่วยเวลาย้อนหลัง และตัวแปรสุ่ม a_t ที่เป็นอิสระกันในปัจจุบันและย้อนหลังไปในอดีต q หน่วยเวลา กระบวนการนี้จึงมีทั้งเทอม AR จำนวน p เทอมและเทอม MA จำนวน q เทอมกระบวนการ $ARMA(p,q)$ มีตัวแบบ ดังนี้

$$(1-\phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)(Z_t - \mu) = (1-\theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q)a_t$$

ลักษณะ ρ_k ของกระบวนการ $ARMA(p,q)$ เหมือนกับลักษณะ ρ_k ของกระบวนการ $AR(p)$ นั่นคือ ρ_k ลดลงเข้าสู่ศูนย์แบบเอกซ์โปเนนเชียล

ลักษณะ ϕ_{kk} ของกระบวนการ $ARMA(p,q)$ เหมือนกับ ลักษณะ ϕ_{kk} ของกระบวนการ $MA(q)$ เมื่อ $k > p-q$ นั่นคือ ϕ_{kk} ลดลงเข้าสู่ศูนย์แบบเอกซ์โปเนนเชียล และ ϕ_{kk} มีลักษณะแตกต่างออกไป เมื่อ $k \leq p-q$

กระบวนการ $ARMA(1,1)$ มีตัวแบบดังนี้ $(1-\phi_1 B)(Z_t - \mu) = (1-\theta_1 B)a_t$ มีลักษณะของ ρ_k มีค่าเข้าใกล้ 0 เมื่อ k มีค่ามากขึ้น และ ϕ_{kk} มีค่าเข้าใกล้ 0 เมื่อ k มีค่ามากขึ้น

ตัวแบบที่ 4 กระบวนการถดถอยในตัวเองบูรณาการกับค่าเฉลี่ยเคลื่อนที่อันดับ *p, d, q* (Autoregressive Integrated Moving Average Process of Order *p, d* and *q*: $ARIMA(p,d,q)$)

คือ อนุกรมเวลาที่ไม่สเตชันนารีเนื่องจากแนวโน้ม เป็นอนุกรมเวลาที่มีค่าคาดหวังของ Z_t ที่แต่ละเวลา t ไม่คงที่ โดยรูปแบบ $ARIMA(p,d,q)$ มี p เป็นอันดับของกระบวนการ AR, d เป็นครั้งที่ของผลต่าง และ q เป็นอันดับของกระบวนการ MA

ตัวแบบของกระบวนการ $ARIMA(p,d,q)$ มีทั้งเทอม AR และ MA ดังนี้

$$(1-\phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)(1-B)^d Z_t = \delta + (1-\theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q)a_t$$

เมื่อ $(1-B)^d Z_t$ แทน ผลต่างครั้งที่ d , $d = 1,2,\ldots$

δ เป็นค่าคงที่, $\delta = (1-\varphi_1 - \varphi_2 - \ldots - \varphi_p)\mu$

เมื่อพิจารณาลักษณะของ ACF ของอนุกรมเวลาที่ไม่สเตชันนารีเนื่องจากแนวโน้มจะเห็นว่ามีลักษณะไม่ต่างกันสำหรับแต่ละรูปแบบ นั่นคือลักษณะ ACF จะลดลงอย่างช้า ๆ เมื่อแปลงอนุกรมเวลาที่ไม่สเตชันนารีเนื่องจาก

แนวโน้ม ให้เป็นอนุกรมเวลาใหม่ที่สเตชันนารี ACF กับ PACF จะบอกลักษณะของอนุกรมเวลาใหม่ได้ (ทรงศิริ แต้สมบัติ, 2549)

ตัวแบบที่ 5 กระบวนการถดถอยในตัวเองแบบมีฤดูกาลอันดับ P (Seasonal Autoregressive Process of Order P: $AR(P)_s$)

$AR(P)_s$ เป็นกระบวนการที่มีการผันแปรตามฤดูกาล มีคาบของฤดูกาล s หน่วยเวลา และค่าสังเกตเฉพาะที่อยู่ห่างกัน s หน่วยเวลา มีความสัมพันธ์กันสูง จึงอธิบายค่าสังเกตที่หน่วยเวลา t ในรูปของค่าสังเกตในอดีตย้อนหลังที่หน่วยเวลา $t-s$ จนถึง $t-Ps$ โดยการกำหนดน้ำหนัก Γ_j , $j=1,2,...,P$ และความรบกวนสุ่มในปัจจุบัน a_t มีตัวแบบดังนี้

$$(1-\Gamma_1 B^s-\Gamma_2 B^{2s}-...-\Gamma_P B^{Ps})Z_t=\delta+a_t$$

ลักษณะของ $\rho_{ks}, k=1,2,...$ ลดลงเข้าสู่ศูนย์แบบเอกซ์โปเนนเชียลหรือแบบคลื่นไซน์เช่น เดียวกับกระบวนการ $AR(p)$

ลักษณะของ $\phi_{ks,ks}$ มีค่าไม่เป็นศูนย์ เมื่อช่วงเวลาห่างกัน $k\le P$ และมีค่าเป็นศูนย์เมื่อช่วงเวลาห่างกัน $k>P$ เช่นเดียวกับกระบวนการ $AR(p)$

ตัวแบบที่ 6 กระบวนการค่าเฉลี่ยเคลื่อนที่แบบมีฤดูกาลอันดับ Q (Seasonal Moving Average Process of Order Q: $MA(Q)_s$)

$MA(Q)_s$ เป็นกระบวนการที่มีการเปลี่ยนแปลงตามฤดูกาล มีคาบของฤดูกาล s หน่วยเวลา และค่าสังเกตเฉพาะที่อยู่ห่างกัน s หน่วยเวลา มีความสัมพันธ์กันสูง จึงอธิบายค่าสังเกตที่หน่วยเวลา t ในรูปของการรบกวนสุ่มในปัจจุบัน a_t และการรบกวนสุ่มในอดีตย้อนหลังที่หน่วยเวลา $t-s$ จนถึง $t-Qs$ โดยการกำหนดน้ำหนัก $-\Delta_j$, $j=1,2,...,Q$ มีตัวแบบดังนี้

$$Z_t=(1-\Delta_1 B^s-\Delta_2 B^{2s}-...-\Delta_Q B^{Qs})a_t+\delta$$

ลักษณะของ ρ_{ks} มีค่ามากกว่าศูนย์ เมื่อ $k=1,2,3,...,Q$ ที่เหลือจะมีค่าเป็นศูนย์ทั้งหมด เหมือนกับกระบวนการ $MA(q)$

ลักษณะของ $\phi_{ks,ks}$ เมื่อ $k=1,2,3,...$ ลดลงเข้าสู่ศูนย์แบบเอกซ์โปเนนเชียลหรือแบบคลื่นไซน์เช่นเดียวกับกระบวนการ $MA(q)$

ตัวแบบที่ 7 กระบวนการผสมถดถอยในตัวเองกับค่าเฉลี่ยเคลื่อนที่แบบมีฤดูกาลอันดับ P และ Q (Seasonal Mixed Autoregressive - Moving Average Process of Order P and Q: $ARMA(P,Q)_s$)

เป็นกระบวนการที่มีการเปลี่ยนแปลงตามฤดูกาล มีคาบของฤดูกาล s หน่วยเวลา และค่าสังเกตเฉพาะที่อยู่ห่างกัน s หน่วยเวลา มีความสัมพันธ์กันสูง จึงอธิบายค่าสังเกตที่หน่วยเวลา t ในรูปของค่าสังเกตในอดีตย้อนหลังที่หน่วยเวลา $t-s$ จนถึง $t-Ps$ และความรบกวนสุ่มในปัจจุบัน a_t และการรบกวนสุ่มในอดีตย้อนหลังที่หน่วยเวลา $t-s$ จนถึง $t-Qs$ ซึ่งมีทั้งเทอม AR และ MA มีตัวแบบดังนี้

$$(1-\Gamma_1 B^s-\Gamma_2 B^{2s}-...-\Gamma_P B^{Ps})Z_t=\delta+(1-\Delta_1 B^s-\Delta_2 B^{2s}-...-\Delta_Q B^{Qs})a_t$$

ลักษณะของ ρ_{ks} ลดลงเข้าสู่ศูนย์แบบเอกซ์โปเนนเชียลหรือแบบคลื่นไซน์

ลักษณะของ $\phi_{ks,ks}$ ลดลงเข้าสู่ศูนย์แบบเอกซ์โปเนนเชียลหรือแบบคลื่นไซน์

ตัวแบบที่ 8 กระบวนการถดถอยในตัวเองบูรณาการกับค่าเฉลี่ยเคลื่อนที่แบบมีฤดูกาลอันดับ P, D และ Q (Seasonal Autoregressive Integrated Moving Average Process of Order P, D and Q: $ARIMA(P,D,Q)_s$)

มีข้อสมมติฐาน คือ ค่าสังเกตที่อยู่ภายในฤดูกาลเดียวกันไม่มีสหสัมพันธ์ต่อกัน แต่จะมีค่าสหสัมพันธ์ในตัวเองแตกต่างจากศูนย์เฉพาะช่วงห่าง $s, 2s, 3s, \ldots$ เท่านั้น

บ็อกซ์และเจนกินส์ ได้แนะนำกระบวนการ $ARIMA(P,D,Q)_s$ โดยที่

P คือ อันดับของ Seasonal Autoregressive

Q คือ อันดับของ Seasonal Moving Average

D คือ ผลต่างระหว่างฤดูกาล (Seasonal difference)

s คือ ความยาวของฤดูกาล

กระบวนการ $ARIMA\left(P,D,Q\right)_s$ ใช้กับอนุกรมเวลาที่มีความแปรผันตามฤดูกาลที่มีการหาผลต่างระหว่างฤดูกาล มีตัวแบบดังนี้

$$(1-\Gamma_1 B^s - \Gamma_2 B^{2s} - \ldots - \Gamma_P B^{Ps})(1-B^s)^D Z_t = \delta + (1-\Delta_1 B^s - \Delta_2 B^{2s} - \ldots - \Delta_Q B^{Qs})a_t$$

ตัวแบบที่ 9 กระบวนการแบบมีฤดูกาลเชิงคูณทั่วไป (Multiplicative seasonal autoregressive integrated moving average process of order $(p,d,q)(P,D,Q)_s$: $ARIMA(p,d,q)(P,D,Q)_s$)

คือ กระบวนการที่ค่าสังเกตอาจมีสหสัมพันธ์ต่อกันภายในฤดูกาลเดียวกัน และมีสหสัมพันธ์ระหว่างฤดูกาล ดังนั้นบ็อกซ์และเจนกินส์ จึงได้แนะนำตัวแบบที่มีฤดูกาลเชิงคูณทั่วไป ซึ่งสามารถใช้ได้ทั้งกับตัวแบบเชิงบวก และตัวแบบเชิงคูณ คือมีตัวแบบดังนี้

$$\begin{aligned}&(1-\Gamma_1 B^s - \Gamma_2 B^{2s} - \ldots - \Gamma_P B^{Ps})(1-\phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)(1-B^s)^D (1-B)^d Z_t \\ &= \delta + (1-\Delta_1 B^s - \Delta_2 B^{2s} - \ldots - \Delta_Q B^{Qs})(1-\theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q)a_t\end{aligned}$$

2.1.3.6 การประมาณค่าพารามิเตอร์

เป็นการหาค่าประมาณของพารามิเตอร์ในรูปแบบด้วยวิธีการประมาณแบบต่าง ๆ ได้แก่

ก. วิธีการประมาณแบบง่าย เริ่มจากการสร้างสมการความสัมพันธ์ระหว่าง ρ_k และพารามิเตอร์ในตัวแบบ จำนวนสมการที่สร้างขึ้นเท่ากับจำนวนพารามิเตอร์ในตัวแบบที่ไม่รวมค่าคงที่ θ หรือเท่ากับ $p+q$ เมื่อแทน ρ_k ด้วย r_k และแทนพารามิเตอร์ θ ด้วยตัวประมาณ $\hat{\theta}$ คำตอบของสมการจะเป็นค่าประมาณของพารามิเตอร์

ข. วิธีกำลังสองน้อยสุด ค่าประมาณของพารามิเตอร์ในตัวแบบที่ได้จากวิธีกำลังสองน้อยที่สุด เป็นค่าประมาณที่ให้ผลรวมกำลังสองของความคลาดเคลื่อนของการพยากรณ์มีค่าน้อยที่สุด การประมาณจะเริ่มจากการกำหนดค่าเริ่มต้นของพารามิเตอร์ซึ่งอาจจะเป็นค่าประมาณจากวิธีการประมาณแบบง่าย การคำนวณจะทำหลายรอบจนกว่าจะได้ค่าประมาณที่คงที่และให้ค่า SSE ที่ต่ำที่สุด

ค. วิธีภาวะน่าจะเป็นสูงสุด มีหลักการทำนองเดียวกับวิธีกำลังสองน้อยที่สุด แต่วิธีภาวะน่าจะเป็นสูงสุดต้องการข้อสมมติเกี่ยวกับลักษณะการแจกแจงแบบปกติของ ε_t ค่าประมาณจากวิธีภาวะน่าจะเป็นสูงสุด เป็นค่าประมาณที่ทำให้ฟังก์ชันภาวะน่าจะเป็นสูงสุด (ทรงศิริ แต้สมบัติ, 2549)

2.1.3.7 ตรวจสอบความเหมาะสมของตัวแบบ โดยการตรวจสอบคุณสมบัติของความคลาดเคลื่อน

1) ความคลาดเคลื่อนไม่มีสหสัมพันธ์ในตัวเอง

โดยการทดสอบ Box-Ljung

ทดสอบสมมติฐาน $H_0 : \rho_1(a_t) = \rho_2(a_t) = \rho_3(a_t) = ... = \rho_m(a_t) = 0$

$H_0 : \rho_k(a_t) \neq 0$ อย่างน้อย 1 ค่า , $k = 1,2,..., m$

ระดับนัยสำคัญ α

ตัวสถิติทดสอบ คือ $Q_m = n(n+2)\sum_{k=1}^{m}\left(\frac{r_k^2(a_t)}{n-k}\right)$

เมื่อ m คือ จำนวน $\rho_k(a_t)$ ที่นำมาทดสอบ

n คือ จำนวน a_t

ถ้าปฏิเสธ H_0 แสดงว่า ความคลาดเคลื่อนมีสหสัมพันธ์ในตัวเอง ที่ช่วงเวลาห่างกัน k หน่วยเวลา อย่างน้อย 1 ค่า

ถ้ายอมรับ H_0 แสดงว่า ความคลาดเคลื่อนไม่มีสหสัมพันธ์ในตัวเอง ที่ช่วงเวลาห่างกัน k หน่วยเวลา ทั้ง m ค่า

2) ค่าเฉลี่ยของความคลาดเคลื่อนเป็นศูนย์

ทดสอบสมมติฐาน $H_0 : \mu_a = 0$

$H_1 : \mu_a \neq 0$

กำหนดระดับนัยสำคัญ α

ตัวสถิติทดสอบ คือ $T = \frac{\bar{a} - 0}{S_a / \sqrt{n}}$

เมื่อ $\bar{a}$ คือ ค่าเฉลี่ยของ $a_t = \frac{1}{n}\sum_{t=1}^{n} a_t$

S_a^2 คือ ความแปรปรวนของ $a_t = \frac{1}{n-1}\sum_{t=1}^{n}(a_t - \bar{a})^2$

n คือ จำนวนของ a_t

ปฏิเสธ H_0 ถ้า $T \geq t_{\alpha/2,(n-1)}$ ดังนั้น μ_a ไม่เท่ากับ 0

ยอมรับ H_0 ถ้า $T < t_{\alpha/2,(n-1)}$ ดังนั้น μ_a เท่ากับ 0

3) ความแปรปรวนของความคลาดเคลื่อนคงที่

โดยการพิจารณาแผนภาพการกระจายระหว่าง a_t กับ t หรือ Z_t หรือ $\hat{Z}_t$ โดยให้แกนนอนแทน t หรือ Z_t หรือ $\hat{Z}_t$ และแกนตั้งแทน a_t ถ้าค่า a_t กระจายอยู่รอบ ๆ ศูนย์ ในแนวขนานกับแกนนอน และไม่มีรูปแบบที่แน่นอน หรือค่า a_t มีค่าในช่วงใดช่วงหนึ่งแคบ ๆ ไม่ว่า t , Z_t , $\hat{Z}_t$ จะเปลี่ยนไปอย่างไร นั่นคือ ความแปรปรวนของ a_t คงที่ ถ้าค่า a_t เพิ่มขึ้น เมื่อ t, Z_t , $\hat{Z}_t$ มีค่ามากขึ้น หรือ ค่า a_t ลดลงเมื่อ t, Z_t , $\hat{Z}_t$ มีค่ามากขึ้นนั่นคือ ความแปรปรวนของ a_t ไม่คงที่

4) ความคลาดเคลื่อนมีการแจกแจงแบบปกติ สามารถพล็อตกราฟดูลักษณะข้อมูลได้หลายวิธี ตัวอย่างเช่น Normal Probability Plot (Normal Q-Q Plot) เป็นกราฟที่พล็อตค่าของข้อมูลจริงที่เกิดขึ้น กับค่าที่คาดไว้เมื่อข้อมูลมีการแจกแจงแบบปกติ ซึ่งค่าที่คาดไว้จะเป็นเส้นตรง ค่าจริงจะอยู่รอบเส้นตรงอย่างสุ่ม จะสรุปว่า ข้อมูลจริงนั้นมีการแจกแจงแบบปกติ (ยุภาพร ตงประสิทธิ์, 2559)

2.1.3.8 **เกณฑ์การคัดเลือกตัวแบบจากวิธีของบ็อกซ์และเจนกินส์**

ใช้เกณฑ์ข้อสนเทศของอาไคเคะ Akaike Information Criterion (AIC) เมื่อขนาดตัวอย่างใหญ่การคัดเลือกตัวแบบโดยใช้เกณฑ์เอไอซีเลือกตัวแบบที่ให้ค่าเอไอซีต่ำสุดเป็นตัวแบบที่ใช้ในการอธิบายตัวแปรตามได้และเกณฑ์เอไอซีสามารถคัดเลือกตัวแบบได้ดีเมื่อตัวอย่างมีขนาดใหญ่ (นันทพร บุญสุข และ จิราวัลย์ จิตรเวช, 2557) โดยหาได้จาก

$$AIC = n \cdot \ln\left(\frac{SSE}{n}\right) + 2p$$

โดยที่ p คือ จำนวนตัวแปรอิสระ

n คือ ขนาดตัวอย่าง

SSE คือ ผลรวมกำลังสองของค่าความคลาดเคลื่อน

2.1.4 เปรียบเทียบค่าพยากรณ์จากวิธีการพยากรณ์ต่าง ๆ เพื่อให้มีเพียงหนึ่งตัวแบบที่ดีที่สุด โดยการเปรียบเทียบค่าความคลาดเคลื่อนต่อไปนี้

2.1.4.1 รากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error, $RMSE$)

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2}$$

2.1.4.2 ค่าสัมบูรณ์ของความเบี่ยงเบนเฉลี่ย (Mean Absolute Error, MAE)

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|e_t\right|$$

2.1.4.3 ค่าสัมบูรณ์ของเปอร์เซนต์ของความคลาดเคลื่อนเฉลี่ย (Mean Absolute Percent Error, $MAPE$) มีหน่วยเป็น ร้อยละ

$$MAPE = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{e_t}{Z_t}\right|$$

วิธีพยากรณ์ที่ให้ค่า $RMSE$, MAE หรือ $MAPE$ มีค่าน้อยกว่าวิธีพยากรณ์อื่น ๆ ย่อมเป็นวิธีที่ดีกว่า

2.2 ผลงานวิจัยที่เกี่ยวข้อง

จินตพร หนิ้วอินปั้น, บุญอ้อม โฉมที และ ประสิทธิ์ พยัคฆพงษ์ (2555) ได้ทำการศึกษาการเปรียบเทียบวิธีการพยากรณ์ 4 วิธี สำหรับความต้องการพลังไฟฟ้าสูงสุดในภาคกลางของประเทศไทย โดยมีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบวิธีการพยากรณ์ 4 วิธี คือ วิธีการปรับให้เรียบเอ็กซ์โปเนนเชียลแบบ Holt-Winter การวิเคราะห์การถดถอยที่ใช้ตัวแปรดัมมี่ วิธีของบอกซ์-เจนกินส์ และการวิเคราะห์การถดถอยแบบฟัซซีที่ใช้ตัวแปรดัมมี่และเพื่อพยากรณ์ความต้องการพลังไฟฟ้าสูงสุดในภาคกลางของประเทศไทย โดยศึกษาข้อมูลจากการไฟฟ้าฝ่ายผลิตแห่งประเทศไทย โดยแบ่งข้อมูลออกเป็น 2 ส่วน ส่วนที่ 1 ข้อมูลตั้งแต่เดือนมกราคม 2545 ถึง เดือนธันวาคม 2550 สำหรับกำหนดรูปแบบและส่วนที่ 2 ข้อมูลตั้งแต่เดือนมกราคม 2551 ถึง เดือนธันวาคม 2551 เพื่อหาช่วงการพยากรณ์ล่วงหน้าที่เหมาะสมที่สุด และนำมาพยากรณ์ล่วงหน้า 3 ช่วง คือ 2 6 และ 12 เดือน จากนั้นเลือกวิธีการพยากรณ์ที่เหมาะสมพิจารณาจากค่าสัมบูรณ์ของเปอร์เซนต์ของความคลาดเคลื่อนเฉลี่ยต่ำที่สุด ผลการศึกษา พบว่า วิธีการพยากรณ์ที่ให้รูปแบบที่เหมาะสมที่สุด คือ วิธีการปรับให้เรียบเอ็กซ์โปเนนเชียลแบบ Holt-Winter จากรูปแบบดังกล่าวนำมาคำนวณช่วงการพยากรณ์ล่วงหน้า 2 6 และ 12 เดือน พบว่าวิธีนี้เหมาะสำหรับการพยากรณ์ระยะสั้นล่วงหน้า 2 เดือน

สุปราณี แซ่ลี (2555) ได้ทำการศึกษาการพยากรณ์การใช้กระแสไฟฟ้าในครัวเรือน โดยวิธีของบอกซ์-เจนกินส์ โดยมีวัตถุประสงค์เพื่อพยากรณ์การใช้กระแสไฟฟ้าในครัวเรือน โดยวิธีบอกซ์-เจนกินส์ โดยใช้ข้อมูลรายเดือนตั้งแต่เดือนมกราคม พ.ศ.2546 ถึงเดือนธันวาคม พ.ศ.2555 รวมทั้งสิ้น 120 เดือน ใช้ข้อมูลจากการไฟฟ้านครหลวง การไฟฟ้าส่วนภูมิภาค กระทรวงมหาดไทย จากการศึกษาพบว่าอนุกรมเวลาคงที่ จึงกำหนดตัวแบบ โดยพิจารณาจากกราฟ ACF และ PACF ได้ตัวแบบที่เหมาะสมคือ ตัวแบบ ARIMA$(1,0,0)(0,1,1)_{12}$ วิธีของบอกซ์-เจนกินส์ที่ได้มานั้นมีความเหมาะสมแล้วจากการพยากรณ์การใช้ไฟฟ้าในครัวเรือนล่วงหน้าจำนวน 12 เดือน คือ ตั้งแต่เดือนมกราคม พ.ศ. 2556 ถึงเดือนธันวาคม พ.ศ. 2556 ได้ค่าความคลาดเคลื่อนรากที่สองเฉลี่ย 92.7362

นิฉา แก้วหาวงษ์ (2558) ได้ทำการศึกษาการพยากรณ์ปริมาณการใช้ไฟฟ้าของประเทศไทย โดยใช้ตัวแบบ $ARIMA_S$ และตัวแบบการถดถอยที่มีความคลาดเคลื่อนเป็นตัวแบบ ARMA โดยมีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบตัวแบบการพยากรณ์อนุกรมเวลา 2 ตัวแบบ คือ ตัวแบบ $ARIMA_S$ และตัวแบบการถดถอยที่มีความคลาดเคลื่อนเป็นตัวแบบ ARMA ข้อมูลที่นำมาใช้ในการศึกษาเป็นข้อมูลอนุกรมเวลารายเดือนของปริมาณการใช้ไฟฟ้าของประเทศไทยซึ่งแบ่งการพยากรณ์เป็น 2 ช่วงเวลาคือ การพยากรณ์ข้อมูลในอดีตตั้งแต่เดือนมกราคม พ.ศ. 2545 ถึงเดือนธันวาคม พ.ศ.2556 เพื่อคัดเลือกตัวแบบที่เหมาะสมที่สุดในแต่ละตัวแบบพยากรณ์ทั้ง 2 ตัวแบบการพยากรณ์ล่วงหน้าตั้งแต่เดือนมกราคม พ.ศ.2557 ถึงเดือนพฤษภาคม พ.ศ.2557 จำนวน 5 เดือน เพื่อคัดเลือกตัวแบบที่ดีที่สุดมา1 ตัวแบบ โดยเกณฑ์ที่ใช้ในการเปรียบเทียบความคลาดเคลื่อนของการพยากรณ์ คือค่าเฉลี่ยของค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อน (MAPE) จากการศึกษาพบว่าตัวแบบที่เหมาะสมที่สุดในการพยากรณ์ปริมาณการใช้ไฟฟ้าของประเทศไทยคือตัวแบบการถดถอยที่มีความคลาดเคลื่อนเป็นตัวแบบ ARMA (2,(6,20)) โดยตัวแบบที่ได้มีค่าความคลาดเคลื่อนจากการพยากรณ์ล่วงหน้าเมื่อวัดด้วยค่า MAPE ต่ำที่สุดเท่ากับ 1.7898 %

ณัฐภัทร ก้อนเครือ และ กัลยา บุญหล้า (2559) ได้ทำการศึกษาการพยากรณ์ปริมาณหน่วยจำหน่ายไฟฟ้าจังหวัดพิษณุโลกงานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาตัวแบบที่เหมาะสมในการพยากรณ์ปริมาณจำหน่ายไฟฟ้าของจังหวัดพิษณุโลก ข้อมูลที่ใช้ในการศึกษา คือ ข้อมูลปริมาณหน่วยจำหน่ายไฟฟ้ารายเดือน ระหว่างเดือนมกราคม พ.ศ. 2544 ถึงเดือนธันวาคม พ.ศ.2558 โดยศึกษาตัวแบบการพยากรณ์ด้วยวิธีการปรับให้เรียบแบบโฮลท์วินเตอร์เชิงผลคูณ และ วิธีบ็อกซ์และเจนกินส์จากนั้นนำมาเปรียบเทียบจากค่าเฉลี่ยของค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อน (MAPE) ต่ำที่สุดผลจากการศึกษาพบว่า วิธีบ็อกซ์และเจนกินส์เป็นวิธีที่เหมาะสมที่สุดสำหรับการพยากรณ์ปริมาณหน่วยจำหน่ายไฟฟ้า มีตัวแบบ $ARIMA(1,1,1)(0,1,1)_{12}$ ซึ่งมีค่าคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ 8.863

ศศิประภา ตาลยงค์ (2560) ได้ทำการศึกษาการเปรียบเทียบตัวแบบการพยากรณ์ปริมาณการใช้ไฟฟ้าของประเทศไทยโดยใช้ตัวแบบวินเทอร์ ตัวแบบอารีมา และตัวแบบวินเทอร์ที่มีความคลาดเคลื่อนเป็นตัวแบบอารีมา งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบตัวแบบการพยากรณ์ปริมาณการใช้ไฟฟ้าของประเทศไทย โดยใช้ตัวแบบวินเทอร์ตัวแบบอารีมา และตัวแบบวินเทอร์ที่มีความคลาดเคลื่อนเป็นตัวแบบอารีมา เพื่อหาตัวแบบการพยากรณ์ที่เหมาะสมที่สุดในการพยากรณ์ปริมาณการใช้ไฟฟ้าของประเทศไทย ซึ่งข้อมูลอนุกรมเวลาที่นำมาใช้ในการวิเคราะห์เป็นข้อมูลปริมาณการใช้ไฟฟ้ารายเดือนโดยแบ่งข้อมูลเป็นสองส่วน คือ 1. ส่วนที่ใช้หาตัวแบบที่เหมาะสมในแต่ละวิธีการพยากรณ์ ใช้ข้อมูลตั้งแต่เดือนมกราคม พ.ศ.2547 ถึงเดือนธันวาคม พ.ศ.2557 จำนวน 132 เดือน 2. ส่วนใช้ในการเปรียบเทียบกับค่าพยากรณ์ล่วงหน้าที่ได้และคัดเลือกตัวแบบที่เหมาะสมที่สุด ใช้ข้อมูลตั้งแต่เดือนมกราคม พ.ศ.2558 ถึงเดือนธันวาคม พ.ศ.2558 จำนวน 12 เดือนและเกณฑ์ที่ใช้ในการเปรียบเทียบความคลาดเคลื่อนของการพยากรณ์ของแต่ละตัวแบบด้วยค่าเฉลี่ยของค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อน (MAPE) โดยตัวแบบที่เหมาะสมจะให้ค่า MAPE ต่ำที่สุด จากผลการวิจัยพบว่า ตัวแบบอารีมามีความเหมาะสมที่สุด

บทที่ 3
วิธีดำเนินการวิจัย

การศึกษาวิจัยเรื่องการพยากรณ์ปริมาณการใช้ไฟฟ้าในมหาวิทยาลัยขอนแก่น ด้วยการวิเคราะห์อนุกรมเวลา โดยใช้วิธีการทำให้เรียบ (Smoothing Method) และ วิธีของบ็อกซ์และเจนกินส์ (Box-Jenkins's Method) เกณฑ์การพิจารณาที่นำมาใช้หาความเหมาะสมของตัวแบบ คือการเปรียบเทียบค่าความคลาดเคลื่อน ซึ่งวิธีการพยากรณ์แบบใดให้ค่าความคลาดเคลื่อนต่ำที่สุด วิธีนั้นเป็นวิธีที่เหมาะสมกับข้อมูลที่นำมาพยากรณ์ ซึ่งมีวิธีดำเนินการวิจัยดังต่อไปนี้

3.1 เครื่องมือที่ใช้ในการวิจัย

3.2 การเก็บรวบรวมข้อมูล

3.3 การทำความสะอาดข้อมูล

3.4 การวิเคราะห์ข้อมูล

3.1 เครื่องมือที่ใช้ในการวิจัย

เครื่องคอมพิวเตอร์, โปรแกรม Excel และโปรแกรม R เหตุผลที่เลือกใช้โปรแกรม R เพราะเป็นโปรแกรมที่ได้รับความนิยมทางวิชาการจำนวนมาก เนื่องจากฟรีและไม่มีลิขสิทธิ์ และในโปรแกรมมีแพ็กเกจจำนวนมากให้เลือกใช้ มีการพัฒนาแพ็กเกจใหม่ ๆ เพิ่มขึ้นทุกวันจากนักวิจัยที่ชำนาญเรื่องนั้นโดยเฉพาะ ดังนั้นผู้วิจัยจึงเลือกใช้โปรแกรม R

3.2 การเก็บรวบรวมข้อมูล

การวิจัยครั้งนี้ได้เก็บรวบรวมข้อมูลจากหน่วยไฟฟ้ากองอาคารและสถานที่ มหาวิทยาลัยขอนแก่น โดยอนุกรมเวลาที่ใช้ในการศึกษาเป็นข้อมูลทุติยภูมิรายเดือนของปริมาณการใช้ไฟฟ้ามหาวิทยาลัยขอนแก่น และแต่ละคณะในมหาวิทยาลัยขอนแก่น ตั้งแต่เดือนมกราคม พ.ศ.2551 ถึงเดือนพฤษภาคม พ.ศ. 2561

3.3 การทำความสะอาดข้อมูล

3.3.1 ตรวจสอบความสมบูรณ์ ความถูกต้องของข้อมูล เช่น ต้องการข้อมูลการใช้ไฟฟ้าเป็นรายเดือนย้อนหลัง 10 ปี เราจะต้องตรวจสอบว่าได้ข้อมูลตามที่ต้องการและได้ครบทุกเดือนในแต่ละปี ไม่มีข้อมูลสูญหาย มีระยะเวลาที่ถูกต้อง

3.3.2 กำหนดเกณฑ์การบันทึกข้อมูลให้เป็นมาตรฐานเดียวกัน เช่น หน่วยปริมาณการใช้ไฟฟ้าเป็น หน่วย

3.3.3 กรณีที่มีข้อมูลสูญหาย เนื่องจากเป็นการพยากรณ์อนุกรมเวลาต้องใช้ข้อมูลจำนวนมากที่มีความต่อเนื่องกัน จึงไม่สามารถตัดข้อมูลที่สูญหายทิ้งได้ ดังนั้นจะต้องกลับไปติดต่อขอข้อมูลในส่วนที่หายใหม่อีกครั้ง

3.4 การวิเคราะห์ข้อมูล

ขั้นตอนในการวิเคราะห์ข้อมูล โดยใช้โปรแกรม R มีรายละเอียดดังนี้

3.4.1 วิธีการทำให้เรียบ (Smoothing Method)

3.4.1.1 นำข้อมูลปริมาณการใช้ไฟฟ้าในมหาวิทยาขอนแก่นมาตรวจสอบว่า มีลักษณะการเคลื่อนไหวแบบใด โดยสร้างกราฟเส้นและกราฟ ACF เพื่อดูลักษณะการเคลื่อนไหวของข้อมูล

3.4.1.2 ถ้าข้อมูลอนุกรมเวลาที่ไม่มีส่วนประกอบของแนวโน้มและฤดูกาล จะเลือกใช้วิธีการทำให้เรียบแบบเอ็กซ์โพเนนเชียลชั้นเดียว (Single Exponential) โดยที่จะเลือก $0 \le \alpha \le 1$ ที่ให้ค่า RMSE, MAE และ MAPE ต่ำสุด

3.4.1.3 ถ้าข้อมูลอนุกรมเวลาที่มีเฉพาะส่วนประกอบของแนวโน้ม จะเลือกใช้วิธีของโฮลต์ (Holt's Method) โดยที่จะเลือก $0 \le \alpha \le 1$ และ $0 \le \beta \le 1$ ที่ให้ค่า RMSE, MAE และ MAPE ต่ำสุด

3.4.1.4 ถ้าข้อมูลอนุกรมเวลาที่มีทั้งส่วนประกอบของแนวโน้มและฤดูกาลวิธีของวินเทอร์ (Winter's Method) ในกรณีที่แยกไม่ออกว่าอนุกรมเวลาเป็นตัวแบบเชิงคูณหรือเชิงบวก ให้สังเกตที่กราฟเส้นและกราฟ ACF โดยที่ตัวแบบเชิงคูณจะไม่มีแนวโน้มทั้งภายในฤดูกาลและระหว่างฤดูกาลและตัวแบบเชิงบวกจะไม่มีแนวโน้มภายในฤดูกาลแต่จะมีแนวโน้มระหว่างฤดูกาล โดยที่จะเลือก $0 \le \alpha \le 1$, $0 \le \beta \le 1$ และ $0 \le \gamma \le 1$ ที่ให้ค่า RMSE, MAE และ MAPE ต่ำสุด

3.4.1.3 ทำการกำหนดตัวแบบที่เหมาะสมโดยใช้โปรแกรม R และหาค่าคลาดเคลื่อน RMSE, MAE และ MAPE ต่ำที่สุด

3.4.2 วิธีของบ็อกซ์และเจนกินส์ (Box-Jenkins's Method)

3.4.2.1 นำข้อมูลปริมาณการใช้ไฟฟ้าในมหาวิทยาขอนแก่นมาตรวจสอบว่า มีคุณสมบัติสเตชันนารีหรือไม่ โดยพิจารณาได้จากลักษณะการเคลื่อนไหวของข้อมูลจากกราฟเส้น และกราฟ ACF ถ้าไม่มีคุณสมบัติสเตชันนารีจะต้องทำให้อนุกรมเวลาที่พิจารณานั้นมีคุณสมบัติสเตชันนารีก่อน ซึ่งคุณสมบัติของกระบวนการสเตชันนารี คือ ไม่มีการเปลี่ยนแปลงไปตามเวลา ไม่มีแนวโน้ม มีค่าเฉลี่ยคงที่ ความแปรปรวนคงที่ และความแปรปรวนร่วมในตัวเองเมื่อเวลาห่างกัน k ขึ้นอยู่กับ k เท่านั้น

3.4.2.2 เมื่อข้อมูลมีคุณสมบัติสเตชันนารีแล้ว ให้นำข้อมูลมาสร้างกราฟ ACF และกราฟ PACF โดยใช้โปรแกรม R เพื่อกำหนดตัวแบบที่เหมาะสมโดยพิจารณาได้จากกราฟ ACF และ PACF ซึ่งอาจมีตัวแบบที่เป็นไปได้มากกว่า 1 ตัวแบบ

3.4.2.3 ประมาณค่าพารามิเตอร์ในตัวแบบ โดยใช้โปรแกรมในการประมาณค่าพารามิเตอร์ และทดสอบค่าพารามิเตอร์ว่ามีค่าแตกต่างจากศูนย์อย่างมีนัยสำคัญ

3.4.2.4 ทำการคัดเลือกตัวแบบที่ให้ค่า AIC ต่ำที่สุดที่ผ่านการทดสอบค่าพารามิเตอร์

3.4.2.5 ตรวจสอบความเหมาะสมของตัวแบบที่กำหนด โดยการตรวจสอบคุณสมบัติของความคลาดเคลื่อน ดังนี้

1) ความคลาดเคลื่อนไม่มีสหสัมพันธ์ในตัวเอง โดยดูจากกราฟเส้น และกราฟ ACF ของค่าความคลาดเคลื่อน และใช้สถิติทดสอบ Box-Ljung

2) ค่าเฉลี่ยของความคลาดเคลื่อนเป็นศูนย์ โดยใช้สถิติทดสอบ T-Test

3) ความแปรปรวนของความคลาดเคลื่อนคงที่ สามารถพิจารณาได้จากกราฟการกระจายของค่าความคลาดเคลื่อน

4) ความคลาดเคลื่อนมีการแจกแจงแบบปกติ สามารถพิจารณาได้จากกราฟฮิสโตแกรม ถ้าหากกราฟสมมาตรแสดงว่า มีการแจกแจงแบบปกติ แต่ถ้าหากกราฟเบ้ซ้าย หรือเบ้ขวา แสดงว่า ค่าความคลาดเคลื่อนไม่มีการแจกแจงแบบปกติ

3.4.3 เปรียบเทียบค่าพยากรณ์

เมื่อได้ตัวแบบจากวิธีการทำให้เรียบและวิธีของบ็อกซ์และเจนกินส์ ผู้วิจัยจะทำการคัดเลือกตัวแบบที่ดีที่สุดที่เหมาะสมกับข้อมูล ที่ให้ค่าความคลาดเคลื่อนต่ำสุด โดยใช้เกณฑ์รากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE), ค่าสัมบูรณ์ของความเบี่ยงเบนเฉลี่ย (MAE), ค่าเฉลี่ยของค่าสัมบูรณ์เปอร์เซนต์ความคลาดเคลื่อน (MAPE)

3.4.4 ทำการพยากรณ์ ใช้ตัวแบบที่เหมาะสมมาพยากรณ์ได้ทั้งแบบจุดและแบบช่วง การพยากรณ์จะใช้สมการพยากรณ์ที่สร้างจากตัวแบบการพยากรณ์ที่กำหนด