การจัดการการพยากรณ์ข้อมูลทางธุรกิจ เมื่อพบค่านอกเกณฑ์

Management Forecasting in Business Data when Outliers Existing

เฉลิมสิน สิงห์สนอง*
Chalermsin Singsanong*

* อาจารย์ประจำ กลุ่มวิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์ประยุกต์ มหาวิทยาลัยธุรกิจบัณฑิตย์
* Lecturer, Section of Mathematics and Statistics, Faculty of Applied Science, Dhurakij Pundit University
* Email: chaloemsin.sig@dpu.ac.th

บทคัดย่อ

บทความวิชาการนี้นำเสนอแนวทางการจัดการข้อมูลธุรกิจและการพยากรณ์ที่เหมาะสมเมื่อมีค่านอกเกณฑ์ การจัดการข้อมูลทางธุรกิจ เป็นการบริหารการจัดเก็บ วิเคราะห์ข้อมูลเกี่ยวกับการดำเนินการทางธุรกิจ ให้ได้มาซึ่งสารสนเทศที่เป็นประโยชน์และช่วยสนับสนุนการตัดสินใจ ค่านอกเกณฑ์เป็นค่าของข้อมูลที่ผิดแผกแตกต่างไปจากข้อมูลค่าอื่นๆ วิธีการประมาณค่าพารามิเตอร์ของประชากร จะไม่สามารถอธิบายคุณลักษณะของประชากรที่มีค่านอกเกณฑ์ปะปนอยู่ได้อย่างครบถ้วน ค่านอกเกณฑ์ส่งผลกระทบโดยตรงต่อค่าเฉลี่ย ความแปรปรวน ค่านอกเกณฑ์ในข้อมูลอนุกรมเวลาหรือข้อมูลธุรกิจ มีผลกระทบต่อการกำหนดตัวแบบการประมาณค่าพารามิเตอร์และการพยากรณ์ ทำให้มีความผิดพลาดสูง ขาดความน่าเชื่อถือ วิธีการตรวจหาและปรับแก้ไขค่านอกเกณฑ์ คือ วิธีกราฟ วิธีการทำซ้ำ วิธีการตรวจสอบแบบตัดออก และ วิธีการทดสอบโดยใช้ Grubbs' test วิธีการพยากรณ์ข้อมูลทางธุรกิจเมื่อตรวจพบค่านอกเกณฑ์ คือ วิธีกำลังสองน้อยที่สุด วิธีตัวประมาณภาวะน่าจะเป็นเมื่อปรับปรุงข้อมูล วิธีกำลังสองน้อยที่สุดเมื่อปรับปรุงข้อมูล วิธีกำลังสองน้อยสุดแบบมีเงื่อนไข วิธีประมาณร่วมพารามิเตอร์ตัวแบบและผลกระทบของข้อมูลผิดปกติ วิธีกำลังสองน้อยสุดแบบถ่วงน้ำหนักด้วยบูทสแทร็พ วิธีการพยากรณ์โดยใช้วิธีกำลังสองน้อยที่สุดแบบค่าเฉลี่ยเวียนเกิด วิธีการพยากรณ์โดยใช้วิธีกำลังสองน้อยที่สุดแบบมัธยฐานเวียนเกิด และวิธีการพยากรณ์โดยใช้วิธีกำลังสองน้อยที่สุดแบบมัธยฐานเวียนเกิดปรับปรุง เป็นต้น ซึ่งแต่ละวิธีของการวิเคราะห์นั้นแตกต่างกันออกไป ค่านอกเกณฑ์ที่มักพบในข้อมูลอนุกรมเวลาและข้อมูลทางธุรกิจเป็นส่วนใหญ่ คือ Additive outlier (AO) และ Innovational outlier (IO)

คำสำคัญ: การจัดการ การพยากรณ์ ข้อมูลธุรกิจ ค่านอกเกณฑ์ อนุกรมเวลา

Abstract

This academic paper discusses the appropriate business data management and forecasting when outliers existing. Business data management is administrative process by which the required data are acquired, validated, stored, protected, and analyzed, and by which its accessibility, reliability, and timeliness are ensured to satisfy the needs of the data users. Outlier data greatly diverges from the rest of other data. The current parameter estimate method cannot represent the population and explain the quality of the mixed up outliers pupation completely. The outliers have the biggest effect on the mean and variance. Outliers in time series or business data may have a moderate to significant impact on the effectiveness of the standard methodology for time series and business data analysis in the aspect of model identification, estimation, and forecasting. That creates error and unreliability. The methods in correcting and detecting outliers in business data are Graph Method, Iterative Procedure, Deletion Diagnostics and Grubbs' test. Forecasting time series data with outliers is through Ordinary Least Square Estimation Method, Maximum Likelihood Estimation Method with Iterative for Adjusted Series, Ordinary Least Square Estimation Method with Iterative for Adjusted Series, Conditional Least Squares Method, Joint Estimation of Model Parameters and Outliers Effect Method, Bootstrap Weighted Least Squares Method, Recursive Mean Ordinary Least Squares Method, Recursive Median Ordinary Least Squares Method and Improved Recursive Median Ordinary Least Squares Method. Outliers regularly found in the business and time series data are Additive outlier (AO) and Innovational outlier (IO).

Keywords : Management, Forecasting, Business Data, Outlier, Time Series

บทนำ

ปัจจุบันสารสนเทศถือเป็นสิ่งสำคัญสำหรับผู้บริหาร เพื่อนำมาใช้ประกอบการการวางแผนและพยากรณ์ทางธุรกิจ กล่าวคือ ในอดีตการวางแผนการจะอาศัยจากการคาดเดาใน 2 ลักษณะ คือ

1) จากประสบการณ์ของผู้บริหาร

2) จากการศึกษาข้อมูลในอดีตว่ามีอัตราการเปลี่ยนแปลงเป็นเช่นไร

วิธีทั้งสองวิธีที่กล่าวมานี้ ค่อนข้างจะเป็นวิธีที่เป็นวิทยาศาสตร์มากขึ้น ซึ่งผู้บริหารส่วนใหญ่จะตัดสินใจโดยทำการศึกษาและอาศัยข้อมูลในอดีตเป็นบรรทัดฐาน และมีภาระหน้าที่ในการวางแผนเพื่อเตรียมการด้านต่างๆ ในอนาคต ด้วยเหตุผลที่ว่าการคาดเดาโดยไม่มีหลักการย่อมจะมีจุดอ่อน ดังนั้นผู้บริหารจำเป็นจะต้องรู้เทคนิคหรือวิธีที่จะพยากรณ์ค่าต่างๆ ในอนาคต เพราะถ้าผู้บริหารไม่รู้วิธีหรือเทคนิคการพยากรณ์ค่าตัวเลขต่างๆ ในอนาคต ผู้บริหารก็จะไม่สามารถคาดคะเนหรือบอกได้ว่าค่าของสิ่งที่ต้องการรู้ในอนาคตนั้นควรอยู่ที่ระดับใด เช่น ยอดขายของปีหน้า ต้นทุนของผลิตสินค้าในปีหน้า ผลกำไรการขายสินค้าปีหน้า เป็นต้น ดังนั้นการพยากรณ์ทางธุรกิจ คือ การคาดการณ์สิ่งที่จะเกิดขึ้นในอนาคต โดยอาศัยการวิเคราะห์ข้อมูลในอดีต ข้อมูลปัจจุบันและจากประสบการณ์ของผู้เชี่ยวชาญ ผลของการพยากรณ์จะทำให้ธุรกิจสามารถทราบถึงแนวโน้มการเปลี่ยนแปลงของธุรกิจหรือสภาพแวดล้อมที่จะมีผลต่อธุรกิจได้ในอนาคต สามารถวางแผนหรือกำหนดนโยบายเพื่อให้บรรลุวัตถุประสงค์ของธุรกิจได้

ข้อมูลธุรกิจ เป็นข้อเท็จจริงที่เกิดขึ้นในเหตุการณ์ต่างๆ ที่เกี่ยวเนื่องกับธุรกิจ อันได้จากการสังเกต การจดบันทึก การสัมภาษณ์และการสอบถาม ข้อมูลเหล่านี้ต้องยังไม่มีการประมวลผล ไม่มีการวิเคราะห์ หรือที่เรียกว่าเป็นข้อมูลดิบ โดยที่ยังไม่สามารถนำไปใช้ประโยชน์ได้ทันที

การจัดการข้อมูลธุรกิจ เป็นการบริหาร การจัดเก็บข้อมูล การวิเคราะห์ข้อมูลที่เกี่ยวข้องกับการดำเนินการทางธุรกิจให้ได้มาซึ่งสารสนเทศที่มีประโยชน์พร้อมจะสามารถนำมาใช้ประโยชน์ได้ในทันที การจัดการข้อมูลธุรกิจจะเกิดประโยชน์สูงสุดหรือมีประสิทธิภาพสูงสุด ก็ต่อเมื่อผู้ใช้ข้อมูลสามารถใช้ข้อมูลที่ต้องการได้อย่างรวดเร็ว ถูกต้องและเป็นกลางมากที่สุด เพื่อจะได้นำข้อมูลเหล่านั้นมาช่วยในการตัดสินใจ นำไปใช้ประโยชน์อื่นๆ ต่อไป ข้อมูลธุรกิจที่นำมาวิเคราะห์เพื่อให้เป็นสารสนเทศ สามารถจำแนกจำแนกออกเป็น 2 ประเภท คือ

1) แหล่งข้อมูลภายในองค์กร ข้อมูลแบบนี้จะเป็นการเก็บข้อมูลที่เกิดขึ้นภายในหน่วยงานของตนเอง เช่น ยอดขายรายเดือนของบริษัท งบประมาณรายจ่ายประจำปีของบริษัท เป็นต้น

2) แหล่งข้อมูลภายนอกองค์กร คือ ข้อมูลที่ได้มาจากแหล่งอื่นที่อยู่นอกองค์กร ซึ่งข้อมูลเหล่านี้ส่วนมากจะเป็นข้อมูลที่มีความพิเศษกว่าข้อมูลภายในองค์กร เพราะต้องอาศัยความสามารถขององค์กรอื่นเข้ามาเกี่ยวข้อง เช่น ต้องการทราบอัตราการใช้ปริมาณน้ำมันเชื้อเพลิงที่มีจำนวนมากที่สุด 10 บริษัท ของประเทศ เป็นต้น จะเห็นว่าความต้องการทราบข้อมูลเหล่านี้จะต้องอาศัยแหล่งข้อมูลของที่อื่น ได้แก่ ข้อมูลจากกรมพลังงาน กระทรวงพลังงาน หรือ สำนักงานสถิติแห่งชาติ เป็นต้น

มนตรี สิงหะวาระ (2549) ได้กล่าวถึงความสำคัญกับการพยากรณ์ทางธุรกิจที่มีต่อองค์กรพอจะสรุปได้

1) ทำให้ผู้บริหารธุรกิจทราบขนาดความต้องการผลิตภัณฑ์ของตลาด และเพื่อลดความไม่แน่นอนของตลาดในอนาคต

2) ทำให้ผู้บริหารใช้ผลการพยากรณ์เป็นข้อมูลในการวางแผนในแผนกงานต่างๆ ของธุรกิจ เพื่อความอยู่รอดหรือสร้างความได้ เปรียบทางธุรกิจ

3) ทำให้ผู้บริหารธุรกิจสามารถใช้ผลการพยากรณ์พิจารณาวิเคราะห์สภาพแวดล้อมการแข่งขันและวิเคราะห์ธุรกิจ รวมทั้งกำหนดวิสัยทัศน์ พันธกิจ เป้าหมาย วัตถุประสงค์ แผนกลยุทธ์และแผนปฏิบัติการได้อย่างเหมาะสมกับสภาวการณ์

4) ทำให้ธุรกิจสามารถจัดสรรทรัพยากรได้อย่างเหมาะสม

5) ทำให้ธุรกิจสามารถลดความสูญเสียที่จะเกิดขึ้นได้

6) ทำให้เกิดการประสานงานและประสานการใช้ทรัพยากรของธุรกิจให้เกิดประสิทธิภาพ

7) สร้างผลตอบแทนให้กับผู้มีส่วนได้ส่วนเสียเพื่อให้เกิดความพึงพอใจสูงสุด

ทั้งหมดที่กล่าวมาข้างต้นเป็นการวิเคราะห์ข้อมูลที่อยู่ภายใต้ข้อมูลที่เป็นปกติ แต่หากมีข้อมูลธุรกิจที่เก็บด้วยการบันทึกหรือเก็บรวบรวมมาเป็นระยะเวลาติดต่อกันตามช่วงเวลาที่ต้องการ ข้อมูลบางช่วงที่ได้รวบรวมมามีค่าต่ำกว่าหรือสูงกว่าปกติ ซึ่งลักษณะข้อมูลดังกล่าวนี้ อาจได้รับผลกระทบจากเหตุการณ์ภายนอก เช่น การจดบันทึกผิดพลาด การเปลี่ยนแปลงนโยบายทางเศรษฐกิจ การลดค่าของเงิน การเกิดภาวะสงคราม การนัดหยุดงาน หรือการเกิดภัยธรรมชาติอย่างรุนแรง เป็นต้น เรียกข้อมูลที่ได้รับผลกระทบต่างๆ นี้ว่า **ค่านอกเกณฑ์** (Outlier) ผู้วิเคราะห์อาจประสบปัญหาเกี่ยวกับการวิเคราะห์ข้อมูลธุรกิจได้

Fuller (2009, pp. 309-311) ได้กล่าวไว้ว่า ค่านอกเกณฑ์ในข้อมูลอนุกรมเวลา (Outlier in Time Series Data) คือ ข้อมูลที่มีค่าแยกออกจากกลุ่มหรือผิดแผกแตกต่างไปจากข้อมูลค่าอื่น เมื่อข้อมูลนั้นได้ถูกเก็บรวบรวมมาตามลำดับเวลาที่เกิดขึ้นต่อเนื่องไปเรื่อยๆ เช่น ยอดขายรายวันของห้างสรรพสินค้าแห่งหนึ่ง งบประมาณค่าใช้ จ่ายด้านโฆษณารายไตรมาส ยอดส่งออกข้าวสารบรรจุถุงไปต่างประเทศรายปี เป็นต้น

Atsushi, Shunsuke and Tetsuya (2010, pp. 3-4) ได้สรุปไว้ว่า ค่านอกเกณฑ์มีโอกาสเกิดขึ้นได้บนพื้นฐานของเหตุผล 3 ประการ คือ

1) จากความผันแปรที่มีอยู่ในประชากรที่ทำการศึกษา (Inherent Variability)

2) จากการจดบันทึกหรือเก็บข้อมูลมีความคลาดเคลื่อน (Measurement Error or Execution Error)

3) ตัวอย่างที่เก็บรวบรวมข้อมูลมามีความแตกต่างไปจากประชากรกลุ่มจริง

ดังนั้น ค่านอกเกณฑ์จึงมีผลกระทบโดยตรงต่อการประมาณค่าเฉลี่ยประชากร ทำให้ความแปรปรวนของประชากรเกิดค่าอคติ (Bias Variance Population) เป็นผลให้ตัวแบบของการประมาณค่าพารามิเตอร์ต่างๆ ของข้อมูลมีความผิดพลาดสูง จึงเป็นไปได้อย่างยิ่งที่การเก็บรวบรวมข้อมูลธุรกิจของแต่และองค์กร อาจจะเกิดค่านอกเกณฑ์ปลอมปนมาด้วยเสมอ หากข้อมูลธุรกิจที่จะนำมาใช้ในการพยากรณ์ทางธุรกิจเกิดมีค่านอกเกณฑ์เกิดขึ้น ก็จะทำให้การพยากรณ์ข้อมูลทางธุรกิจเกิดความผิดพลาด ส่งผลต่อความถูกต้องและความน่าเชื่อถือของค่าพยากรณ์ วิธีการตรวจหาและปรับแก้ไขค่านอกเกณฑ์ในข้อมูลธุรกิจจึงเป็นเรื่องสำคัญมากเรื่องหนึ่งที่ผู้วิเคราะห์ข้อมูลธุรกิจหลายท่านได้ให้ความสนใจ

แนวคิดและทฤษฎี

การพยากรณ์ทางธุรกิจเริ่มต้นจากที่ผู้ดำเนินธุรกิจ ตระหนักถึงความต้องการผลลัพธ์ที่ได้จากการพยากรณ์ซึ่งจะนำไปใช้ในการตัดสินใจ สามารถแบ่งออกได้เป็น 2 กลุ่มใหญ่และประเภทย่อยๆ ดังนี้

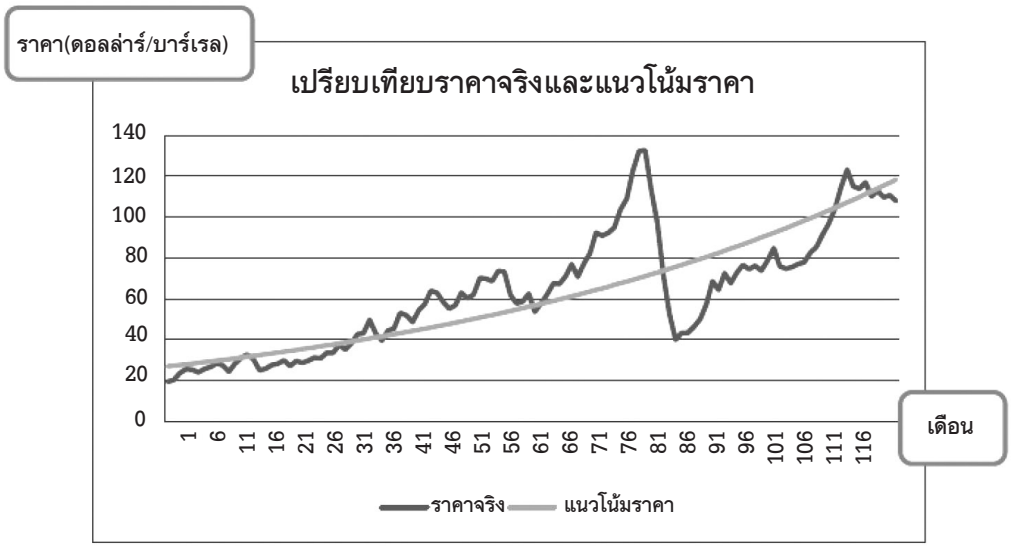
1) การพยากรณ์เชิงคุณภาพ (Qualitative Forecasting)
   - การคาดการณ์ที่ดี (Good Guess)
   - เหตุผลของผู้มีอำนาจจากคณะกรรมการบริหาร (Jury of Executive Opinion)
   - โอกาสในการขายสินค้า (Sale Force Composite)
   - เทคนิคเดลฟาย (Delphi Technique)
2) การพยากรณ์เชิงปริมาณ (Quantitative Forecasting)
   - การพยากรณ์ข้อมูลอนุกรมเวลา (Time Series Data Forecasting)
   - การพยากรณ์อย่างง่าย (Naive Method)
   - ค่าเฉลี่ยเลขคณิตอย่างง่าย (Simple Arithmetic Average)
   - ค่าเฉลี่ยเคลื่อนที่ (Moving Average)
   - ทำให้เรียบแบบ เอ็กซ์โพแนนเชียล (Exponential Smoothing)
   - วิเคราะห์แนวโน้ม (Trend Analysis Method)
   - วิธีแบบคลาสสิค (Classical Decomposition)
   - วิธีบอกซ์–เจนกินส (Box–Jenkins Method)
   - วิเคราะห์ความสัมพันธ์เชิงสาเหตุ (Causal Analysis)
   - วิเคราะห์การถดถอย (Regression Analysis)
   - แบบจำลองทางเศรษฐมิติ (Econometric Model)

การพยากรณ์ที่นิยมใช้ในทางธุรกิจ คือการพยากรณ์เชิงปริมาณ ผู้เขียนจะขอยกมาเพียงบางวิธีที่เป็นที่นิยมคือการพยากรณ์ข้อมูลอนุกรมเวลา โดยจะอธิบายแนวคิดทฤษฎีเพียงบางส่วนเพื่อผู้อ่านจะได้เห็นแนวทางและวิธีการ

การพยากรณ์ข้อมูลอนุกรมเวลา (Time Series Data Forecasting)

ข้ออนุกรมเวลา (Time Series Data) หมายถึง ชุดของข้อมูลที่เก็บรวบรวมตามระยะเวลาเป็นช่วงๆ อย่างต่อ เนื่องกัน ซึ่งอาจเก็บเป็น รายวัน รายเดือน รายไตรมาส หรือรายปี ขึ้นอยู่กับประโยชน์ที่จะนำไปใช้ ลักษณะข้อมูลของอนุกรมเวลา โดยทั่วไปแล้วประกอบด้วย 4 องค์ประกอบด้วยกัน คือ

1) แนวโน้ม (Secular Trend : T) เป็นการเคลื่อนไหวขึ้นหรือลงของข้อมูลอย่างช้าๆ ในช่วงเวลานานๆ ซึ่งอาจมีได้หลายรูปแบบ โอยอาจเป็นเส้นตรง เส้นโค้ง หรือลักษณะอื่นๆ ก็ได้ เช่น ปริมาณการใช้ไฟฟ้าของประเทศไทย หรือ ปริมาณการนำเข้าน้ำมันดิบ เป็นต้น

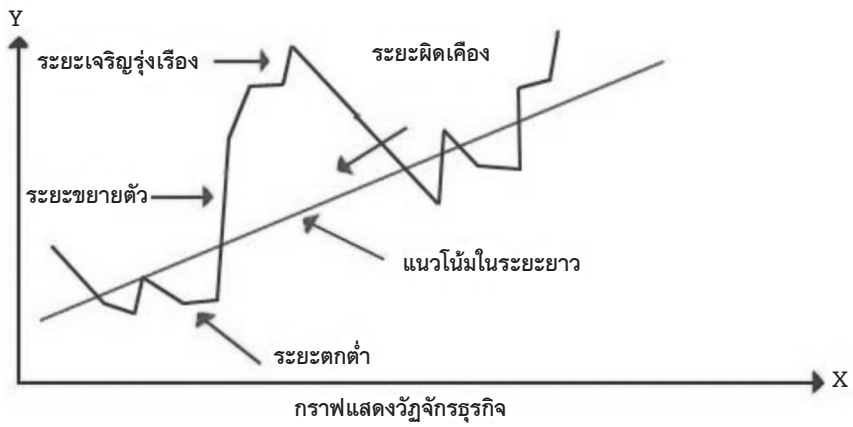


ภาพที่ 1 กราฟอนุกรมเวลาที่แสดงลักษณะของส่วนประกอบแนวโน้ม
ที่มา : ภูริทัต นาคประเสริฐ, 2554

2) การเคลื่อนไหวตามฤดูกาล (Seasonality Movement : S) เป็นความเคลื่อนไหวของข้อมูลที่เกิดขึ้นเนื่องจากอิทธิพลของฤดูกาล ซึ่งจะเคลื่อนไหวขึ้นๆลงๆ ซ้ำกันในช่วงเวลาเดียวกันของแต่ละปี มักพบในข้อมูลช่วงเวลาน้อยกว่า 1 ปี โดยอาจเป็นราย 4 เดือน ราย 3 เดือน รายเดือน หรือรายสัปดาห์ก็ได้ เช่น ภูมิอากาศ หรือจากสภาวะที่มนุษย์สร้างขึ้นเอง เช่น การเกิดอุบัติเหตุในช่วงเทศกาลต่างๆ อัตราการจองห้องของโรงแรม รีสอร์ท ในช่วงเทศกาลเป็นต้น การที่สามารถหาลักษณะของการเปลี่ยนแปลงตามฤดูกาลได้จะเป็นประโยชน์แก่ผู้ควบคุมดูแลในการวางแผนล่วงหน้าเกี่ยวกับการลดจำนวนการเกิดอุบัติเหตุ

ภาพที่ 2 กราฟอนุกรมเวลาที่แสดงลักษณะของอิทธิพลฤดูกาล
ที่มา : ภูริทัต นาคประเสริฐ, 2554

3) การเคลื่อนไหวตามวัฏจักร (Cyclical Movement : C) เป็นการเคลื่อนไหวของข้อมูลที่เกิดขึ้นซ้ำๆกัน คล้ายกับความเคลื่อนไหวตามฤดูกาล เพียงแต่ความเคลื่อนไหวนี้เกิดขึ้นเป็นวัฏจักร (Cycle) ในระยะยาวมากกว่า 1 ปี และในข้อมูลรายปี วัฏจักรเหล่านี้มีแบบแผนไม่แน่นอน จึงทำให้ยากที่จะพยากรณ์ ความเคลื่อนไหวตามวัฏจักรมีสาเหตุเกิดจาก สภาพทางเศรษฐกิจโดยทั่วไป การเปลี่ยนแปลงนโยบายของรัฐ หรือการเปลี่ยนแปลงในรสนิยมของผู้บริโภคและนิสัยการจับจ่ายใช้สอย ซึ่งวัฏจักรโดยทั่วไปจะประกอบด้วยระยะเวลาหนึ่งที่รุ่งเรือง (Prosperity) ตามมาด้วยระยะฝืดเคือง (Recession) ระยะตกต่ำ (Depression) และระยะฟื้นตัว (Recovery) แล้วก็กลับมาสู่ระยะรุ่งเรืองอีกครั้งหนึ่ง หมุนเวียนเป็นวัฏจักรเช่นนี้เป็นเรื่อยๆ เช่น วัฏจักรธุรกิจ (Business Cycle) วัฏจักรเศรษฐกิจ (Economic Cycle) วัฏจักรสภาพอากาศ (Weather Cycle)



ภาพที่ 3 กราฟแสดงอิทธิพลวัฏจักรกับการเคลื่อนไหวของอนุกรมเวลา
ที่มา: สำนักนโยบายการออมและการลงทุน, 2548.

4) ความเคลื่อนไหวผิดปกติ (Irregular Movement : I) เป็นความเคลื่อนไหวที่ไม่แน่นอน ไม่สามารถคาดการณ์ได้ล่วงหน้า โดยข้อมูลในอดีตเกิดขึ้นเนื่องมาจากสาเหตุตามปัจจัย เช่น ภัยธรรมชาติ สงคราม การนัดหยุดงาน การประท้วง ปฏิวัติรัฐประหาร การเลือกตั้ง เป็นต้น การเปลี่ยนแปลงนี้มิได้เกิดขึ้นเป็นประจำ

ภาพที่ 4 กราฟอนุกรมเวลาที่แสดงลักษณะของอิทธิพลเหตุการณ์ผิดปกติ
ที่มา : บริษัท พีทีที โกลบอล เคมิคอล จำกัด (มหาชน), 2555.

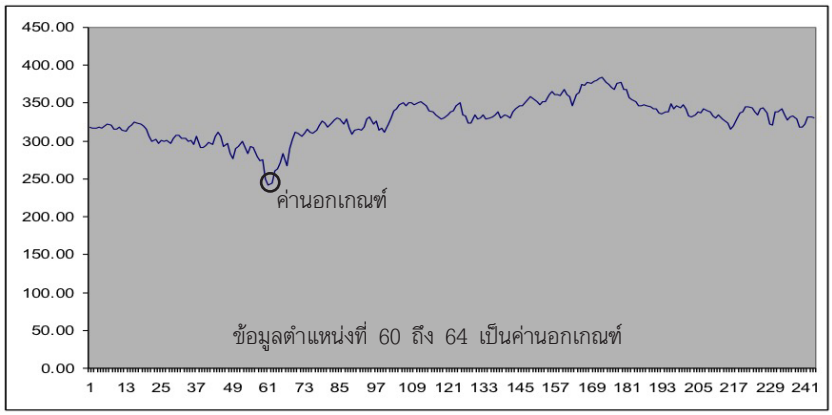
ตัวแบบของอนุกรมเวลาที่นิยมใช้กันนั้นมีอยู่ 2 ตัวแบบ คือ ตัวแบบผลคูณ (Multiplicative Model) และตัวแบบผลบวก (Additive Model) กล่าวคือถ้า Y เป็นอนุกรมเวลาจะได้ว่าตัวแบบผลคูณของ Y คือ Y = T X S X C X I และตัวแบบผลบวกของ Y คือ Y = T + S + C + I สิ่งที่สำคัญมากอย่างหนึ่งของการเลือกวิธีที่ใช้ในการพยากรณ์ข้อมูลอนุกรมเวลาก็คือ ความถูกต้องของการพยากรณ์ คำว่าความถูกต้องในที่นี้หมายถึง ค่าพยากรณ์ที่ได้ เปรียบเทียบกับค่าข้อมูลจริงมีความใกล้เคียงกัน หรือกล่าวอีกนัยหนึ่งว่า ค่าความคลาดเคลื่อนของการพยากรณ์ (Forecast Error) มีค่าน้อยๆ ค่าที่ใช้วัดความถูกต้องของการพยากรณ์มีหลายค่าด้วยกันแต่ที่นิยมใช้ก็มี ค่าเฉลี่ยคลาดเคลื่อนกำลังสองเฉลี่ยของค่าพยากรณ์ (Prediction Mean Square Error : PMSE) กำลังสองของความคลาดเคลื่อน (Mean Squared Error หรือ MSE) หรือค่าเฉลี่ยของค่าสัมบูรณ์ของความเบี่ยงเบน (Mean Absolute Deviation หรือ MAD) ฯลฯ เนื่องจากวิธีการพยากรณ์ทางธุรกิจมีหลากหลายวิธีด้วยกัน ดังนั้นการเลือกวิธีพยากรณ์ให้เหมาะสมกับข้อมูลที่จะพยากรณ์ก็เป็นสิ่งที่สำคัญ ทั้งนี้อาจต้องคำนึงถึงในเรื่องของความถูกต้องของการพยากรณ์ ลักษณะของข้อมูล ค่าใช้จ่าย ระดับความรู้ของผู้พยากรณ์ และหน่วยเวลาที่จะใช้ในการพยากรณ์

จากที่กล่าวมา เป็นการพยากรณ์ที่อาศัยข้อมูลในอดีตมาพิจารณาว่าลักษณะการเปลี่ยนแปลงของข้อมูลเมื่อเวลาเปลี่ยนไปมีลักษณะเป็นอย่างไร มีการเคลื่อนไหวมากน้อยเพียงใด โดยมีข้อสมมติว่าการเคลื่อนไหวของข้อมูลในอนาคตจะไม่แตกต่างกับอดีต ลักษณะข้อมูลที่จะนำมาวิเคราะห์ควรเป็นข้อมูลที่เก็บรวบรวมมานานพอสมควร ข้อมูลต้องมีความสัมพันธ์กับเวลาและระยะเวลาที่ใช้ในการเก็บรวบรวมข้อมูลจะต้องเท่ากัน หากข้อมูลที่เก็บรวบรวมมาได้ถูกกระทบกระเทือนเนื่องจากช่วงเวลาที่ไม่เท่ากันจะต้องปรับแก้ข้อมูลก่อนนำมาวิเคราะห์

หลักการทั้งหมดที่กล่าวมา สามารถนำมาประยุกต์กับการพยากรณ์ทางธุรกิจที่อยู่ภายใต้ข้อมูลเป็นปกติ แต่หากข้อมูลธุรกิจที่นำมาใช้ในวิเคราะห์เกิดมีค่านอกเกณฑ์ ก็จะทำให้การพยากรณ์ธุรกิจเกิดความผิดพลาด ขาดความน่าเชื่อถือ

ค่านอกเกณฑ์ (Outlier)

ค่านอกเกณฑ์ มีผลกระทบโดยตรงต่อตัวแบบการประมาณค่าพารามิเตอร์ต่างๆ ของข้อมูลประชากร ทำให้มีความผิดพลาดสูง เช่น ราคาปิดหุ้น ปตท. ปี 2554 ในตลาดหลังทรัพย์แห่งประเทศไทย



ภาพที่ 5 กราฟอนุกรมเวลาที่แสดงลักษณะของค่านอกเกณฑ์

ที่มา : บริษัท พีทีที โกลบอล เคมิคอล จำกัด (มหาชน), ราคาปิดหุ้น ปตท.

จากภาพที่เห็นพบว่า ข้อมูลในตำแหน่งที่ 60 ถึง 64 เป็นค่านอกเกณฑ์ มีผลกระทบทำให้ค่าเฉลี่ยราคาปิดหุ้น ปตท. ต่ำกว่าเกณฑ์ ปกติ และค่าความแปรปรวนมีค่าสูงกว่าปกติ

จงรักษ์ ศรีทิพย์ (2548, น. 3-5) ได้สรุปและจำแนกค่านอกเกณฑ์ ออกเป็น 5 ประเภท คือ

1) Variance Change (VC) (Tsay, 1988, pp. 1-20) คือ ค่านอกเกณฑ์ที่ส่งผลต่อการเปลี่ยนแปลงของค่าความแปรปรวนของค่าสังเกตในข้อมูลอนุกรมเวลา ณ เวลา T ซึ่งเปลี่ยนแปลงจาก σ_a^2 มาเป็น $(1+ \omega_{VC}) \sigma_a^2$ โดยที่ ω_{VC} คือ ขนาด (Magnitude) ของค่านอกเกณฑ์

2) Level Shift (LS) (Tsay, 1988, pp. 1-20) คือ ค่านอกเกณฑ์ที่ก่อให้เกิดความเปลี่ยนแปลงของค่าเฉลี่ยของข้อมูลอนุกรมเวลาโดยส่งผลกระทบต่อค่าสังเกตที่ตำแหน่ง T และค่าสังเกตในลำดับต่อๆ ไป อันเป็นการเปลี่ยนแปลงที่ถาวร

3) Temporary Change (TC) (Tsay, 1986, pp. 132-141) คือ ค่านอกเกณฑ์ที่ส่งผลกระทบต่อค่าสังเกตที่ทำให้เกิดความเปลี่ยนแปลงอย่างมาก และผลกระทบนั้นส่งผลต่อค่าสังเกตอื่นๆ คือ ส่งผลกระทบต่อค่าสังเกตที่ตำแหน่ง T และค่าสังเกตในลำดับต่อๆ ไป แต่ผลกระทบของค่า Temporary Change (TC) จะไม่อยู่ถาวร แต่จะค่อยลดลงแบบ เอ็กโปแนนเชียล โดยเริ่มต้น ณ เวลาที่ T แล้วค่อยๆ ลดลง

4) Innovational Outlier (IO) (Fox, 1972, pp. 350-363) คือค่านอกเกณฑ์ที่ส่งผลกระทบต่อข้อมูลอนุกรมเวลา ณ ตำแหน่งที่เกิดค่านอกเกณฑ์ หากยังส่งผลกระทบต่อค่าสังเกตอื่นๆ ที่ตามมาอีกด้วย นั่นคือ ค่านอกเกณฑ์ประเภทนี้ จะส่งผลกระทบต่อค่าสังเกตที่ตำแหน่ง T, T+1 , ... เมื่อ T คือตำแหน่งเวลาที่เกิดค่านอกเกณฑ์

5) Additive Outlier (AO) (Fox, 1972, pp. 350-363) คือ ค่านอกเกณฑ์ที่ส่งผลกระทบต่อค่าสังเกตเพียง 1 ค่าเท่านั้น คือ จะส่งผลทำให้ระดับของอนุกรมเวลา ณ เวลาที่เกิดค่านอกเกณฑ์นั้นเกิดความเปลี่ยนแปลงอย่างมาก หลังจากนั้นข้อมูลอนุกรมเวลาจะเข้ากลับสู่เหตุการณ์ปกติ รูปแบบของค่านอกเกณฑ์ AO นี้ จะส่งผลกระทบเฉพาะค่าสังเกตที่ T เท่านั้น เมื่อ T เป็นว่าที่เกิดค่านอกเกณฑ์

ข้อมูลทางธุรกิจหนึ่งๆ อาจจะประกอบด้วยค่านอกเกณฑ์หลายประเภทประปนกันอยู่ในข้อมูลชุดนั้น ทำให้เกิดความยุ่งยากในการวิเคราะห์ Chang, Tiao and Chen (1989, pp. 193-203) ได้กล่าวไว้ว่า ค่านอกเกณฑ์ที่มีอยู่ในข้อมูลธุรกิจส่วนใหญ่มักเป็นค่านอกเกณฑ์ประเภท Additive outlier (AO) และ Innovational Outlier (IO) และ Chen and Lui (1993, pp. 13-35) ได้กล่าวว่า ค่านอกเกณฑ์ที่เกิดขึ้นในข้อมูลจะส่งผลกระทบทำให้การสรุปผลจากการวิเคราะห์ไม่ถูกต้อง และมีอคติ (Bias)

ดังนั้น จึงควรมีวิธีตรวจหาค่านอกเกณฑ์ เมื่อได้มีการตรวจพบค่านอกเกณฑ์แล้ว ควรจะพิจารณาหาสาเหตุของการเกิดค่านอกเกณฑ์ดังกล่าวเสียก่อน ถ้าค่านอกเกณฑ์นั้นเกิดจากความผิดพลาดในการวัด การจดบันทึก ฯลฯ ก็ควรตัดค่านอกเกณฑ์หรือปรับแก้ไขค่านั้นก่อนการวิเคราะห์ การตรวจหาค่านอกเกณฑ์ในข้อมูลจึงมีความสำคัญเป็นอย่างมากต่อการวิเคราะห์ข้อมูล

วิธีการตรวจหาค่านอกเกณฑ์ ที่นิยมใช้

1) การตรวจค่านอกเกณฑ์ด้วยวิธีกราฟเป็นวิธีการตรวจสอบที่ง่ายและสะดวก แต่บางครั้งการใช้วิธีกราฟ อาจให้ผลสรุปที่ไม่แน่นอน ทั้งนี้ขึ้นอยู่กับองค์ประกอบหลายประการรวมทั้งความเชี่ยวชาญและประสบการณ์ของผู้วิเคราะห์ การตรวจสอบด้วยวิธีกราฟมีหลายวิธี วิธีที่ง่ายและนิยมที่สุดคือ บ๊อกพลอต (Box Plot หรือ Box and Whisker Plot)

2) การตรวจสอบค่านอกเกณฑ์ด้วยการทดสอบทางสถิติ มีการศึกษาและพัฒนาตัวสถิติสำหรับการตรวจหาค่านอกเกณฑ์ในหลากหลายสถานการณ์ อาทิ

ก) วิธีการตรวจสอบค่านอกเกณฑ์ด้วยค่าเฉลี่ย (Mean) คือ คำนวณค่าเฉลี่ยของข้อมูลทั้งหมด นำค่าเฉลี่ยที่ได้นั้นไปหักออกจากค่าของข้อมูลชุดนั้น จะได้ข้อมูลชุดใหม่ นำข้อมูลชุดใหม่และชุดเดิมมาหาผลต่างก็จะพบค่านอกเกณฑ์ที่เกิดขึ้น

ข) กระบวนการทำซ้ำ (Iterative Procedure) จงรักษ์ ศรีทิพย์ (2548, น. 20-32) ได้กล่าวไว้ว่า วิธีนี้เหมาะสมกับการตรวจหาค่านอกเกณฑ์ประเภท AO และ IO มีหลักการคือค่อยๆ กำจัดอิทธิพลของค่านอกเกณฑ์ที่ตรวจพบในการทำซ้ำแต่ละรอบ ที่ ส่งผลกระทบต่อค่าประมาณค่าพารามิเตอร์ของตัวแบบให้ลดลงเรื่อยๆ จนกว่าจะไม่สามารถลดอิทธิพลลงต่อไปได้อีกวิธีการนี้แบ่งออกได้เป็น 2 ขั้นตอน คือ

1) ขั้นตรวจหาตำแหน่งและประเภทของค่านอกเกณฑ์ (Outlier-Detection) ซึ่งประกอบด้วยขั้นตอนย่อย 3 ขั้นตอน คือ

1.1) การประมาณค่านอกเกณฑ์ w คือ ขนาดของค่านอกเกณฑ์ที่ส่งผลกระทบต่อค่าสังเกตในข้อมูลอนุกรมเวลา การประมาณ w สามารถใช้วิธีกำลังสองน้อยที่สุด (OLS) ประมาณได้ ตัวแบบของค่านอกเกณฑ์ประเภท AO สามารถเขียนได้เป็น $e_t = \omega_{AO}\pi(B)I_t^{(T)} + a_t$ เมื่อ $\pi(B) = \frac{\phi(B)}{\theta(B)}$

ให้ $\hat{\omega}_{AO,T}$ เป็นตัวประมาณค่าแบบกำลังสองน้อยที่สุดของ ω ของตัวแบบค่านอกเกณฑ์ประเภท AO ณ เวลา T และมี a_t เป็น White-Noise ดังนั้นจะได้ $\hat{\omega}_{AO,T} = \frac{e_t - \sum_{j=1}^{n-T}\pi_j e_{T+j}}{\sum_{j=0}^{n-T}\pi_j^2} = \frac{\pi(F)e_T}{\eta^2}$

โดยที่ $\eta^2 = (1 + \pi_1^2 + \pi_2^2 + ... + \pi_{n-T}^2)$ และ $\pi(F)$ คือ Forward Shift Operator นั่นคือ $F^j e_t = e_{t+j}$ ซึ่งมีความแปรปรวนเป็น $Var(\hat{\omega}_{AO,T}) = \frac{\sigma_a^2}{\eta^2}$ ทำนองเดียวกัน $\hat{\omega}_{IO,T}$ เป็นตัวประมาณค่าแบบกำลังสองน้อยที่สุดของ ω ในตัวแบบค่านอกเกณฑ์ประเภท IO ณ เวลา T มีดังนี้ $\hat{\omega}_{IO,T} = e_T$ และความแปรปรวนของ $\hat{\omega}_{IO,T}$ คือ $Var(\hat{\omega}_{IO,T}) = Var(e_T) = \sigma_a^2$ จะพบว่า $Var(\hat{\omega}_{AO,T}) \le Var(\hat{\omega}_{IO,T}) = \sigma_a^2$

1.2) ตัวสถิติสำหรับการตรวจสอบค่านอกเกณฑ์ เป็นขันตอนการทดสอบสมมติฐานของการตรวจสอบค่านอกเกณฑ์ ให้ H_o : อนุกรมเวลา Y_t ไม่มีค่านอกเกณฑ์ หรือ $\omega = 0$, H_1 : อนุกรมเวลา Y_t มีค่านอกเกณฑ์ประเภท AO หรือ $\omega_{AO} \neq 0$, H_2 : อนุกรมเวลา Y_t มีค่านอกเกณฑ์ประเภท IO หรือ $\omega_{IO} \neq 0$ ตัวสถิติที่ใช้ทดสอบสมมติฐานนี้เป็น Likelihood Ratio (Fox, 1972, pp. 350- 363) ปกติแล้วจะไม่ทราบตำแหน่งของค่านอกเกณฑ์ จึงจำเป็นต้องคำนวณ $\hat{\lambda}_{AO,t}$ และ $\hat{\lambda}_{IO,t}$ (เรียกว่าค่าตัวสถิติทดสอบสมมติฐาน) สำหรับ t = 1, 2, ..., n จะได้ตัวสถิติ

$\tilde{\lambda}_{1,T_A} = \max_{t=1,2,....n} |\tilde{\lambda}_{AO,t}|$ และ $\tilde{\lambda}_{2,T_I} = \max_{t=1,2,....n} |\tilde{\lambda}_{IO,t}|$ และกำหนดให้

$\tilde{\lambda}_T = \max\{\tilde{\lambda}_{1,T_A}, \tilde{\lambda}_{2,T_I}\}$

ต่อมา Chang and Tiao (1983) ได้ทำการพัฒนาวิธีการของ Fox (1972) ให้มีรูปแบบที่เหมาะสมมากขึ้น โดยนำค่าตัวสถิติ $\hat{\lambda}_{AO,t}$ และ $\hat{\lambda}_{IO,t}$ มาเปรียบเทียบกับค่าวิกฤติ C โดยที่ C = 3.0 ใช้สำหรับต้องการให้มีความไวสูง (high sensitivity) C = 3.5 ใช้สำหรับต้องการให้มีความไวปานกลางและ C=4.0 ใช้สำปรับต้องการให้มีความไวต่ำต่อการ ตรวจสอบค่านอกเกณฑ์ นั่นคือ

ถ้า $\hat{\lambda}_T = |\hat{\lambda}_{1,T_A}| > C$ แสดงว่ามีค่านอกเกณฑ์ประเภท AO เกิดขึ้น ณ เวลา t=TA

ถ้า $\hat{\lambda}_T = |\hat{\lambda}_{2,T_I}| > C$ แสดงว่ามีค่านอกเกณฑ์ประเภท IO เกิดขึ้น ณ เวลา t=TI

1.3) การปรับแก้ค่าสังเกตที่ถูกระบุว่าเป็นค่านอกเกณฑ์ เมื่อมีการตรวจพบค่านอกเกณฑ์แล้วจะต้องทำการปรับแก้ค่าสังเกตนั้นๆ เพื่อกำจัดผลกระทบที่จะเกิดขึ้น กรณีที่ค่านอกเกณฑ์เป็นประเภท AO จะทำการปรับแก้ด้วยแทน Y_t ด้วย $\tilde{Y}_t$

$$\tilde{Y}_t = \begin{cases} Y_t & : t \neq T \\ Y_t - \hat{\omega}_{AO} & : t = T \end{cases}$$ และทำการปรับแก้ค่าความคลาดเคลื่อน ณ ตำแหน่ง t ณ T ด้วยตัวสถิติ

$$\tilde{e}_t = \begin{cases} e_t & : t < T \\ e_t - \hat{\omega}_{AO,T}\hat{\pi}_{t-T} & : t \geq T \end{cases}$$ กรณีค่านอกเกณฑ์เป็นประเภท IO จะทำการปรับแก้ด้วยแทน Yt ด้วย นั่นคือ

$$\tilde{Y}_t = \begin{cases} Y_t & : t < T \\ Y_t - \hat{\omega}_{AO} & : t \geq T \end{cases}$$ และทำการปรับแก้ค่าความคลาดเคลื่อน ณ ตำแหน่ง t = T ด้วยตัวสถิติ

$$\tilde{e}_t = \begin{cases} e_t & : t \neq T \\ e_t - \hat{\omega}_{IO,T} & : t = T \end{cases}$$

มีนักสถิติหลายท่านที่ใช้วิธีการตรวจหาค่านอกเกณฑ์ด้วยกระบวนการทำซ้ำ อาทิ Chang and Tiao (1983), Tsay (1986), Chang, Tiao and Chen (1988), Chan and Liu (1993) และ Choy (2001) เป็นต้น นอกจากการทดสอบสมมติฐานด้วยวิธีของ Fox (1972) ดังที่กล่าวแล้ว นิภาพร ปัญญายงค์ (2541, น. 10-16) ได้กล่าวไว้ว่า ยังสามารถทำการทดสอบสมมติฐานค่านอกเกณฑ์ด้วยตัวสถิติ Standardized Residual, Studentized Deleted Residual, Externally Studentized Residual, Externally Studentized Residual และค่า Maximum Normed Residual

2) ขั้นประมาณค่าพารามิเตอร์ (Parameter Estimation) ขั้นตอนนี้จะแบ่งเป็น 2 ขั้นตอนย่อย คือ

2.1) การกำหนดตัวแบบ (Mode: Identification) คือ การกำหนดตัวแบบให้กับข้อมูลอนุกรมเวลาก่อน สามารถทำได้โดยการวิเคราะห์ลักษณะข้อมูลในอดีต พิจารณาจากกราฟข้อมูล (Data Plot) หาฟังก์ชันสหสัมพันธ์ในตัวเอง (Autocorrelation Function: ACF) ฟังก์ชันสหสัมพันธ์

ในตัวเองบางส่วน (Partial Autocorrelation Function: PACF) หรือ ฟังก์ชันขยายของสหสัมพันธ์ในตัวเอง (Extended Autocorrelation Function: EACF) ของค่าสังเกตเทียบกับลักษณะทางทฤษฎีของตัวแบบ

2.2) การประมาณค่าพารามิเตอร์ของตัวแบบอนุกรมเวลาและผลกระทบของค่านอกเกณฑ์ คือ การประมาณค่าพารามิเตอร์ของตัวแบบนั้น วิธีการที่นิยมใช้ในการประมาณค่าพารามิเตอร์ มี 2 วิธี คือ

1) วิธีตัวประมาณแบบภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation Method: MLE) เป็นวิธีที่มีประสิทธิภาพสูง และวิธีกำลังสองน้อยที่สุด (OLS) (Ansley, อ้างถึงใน Vandaele 1983, p.115) และตามแนวคิดของ Chang, Tiao and Chen (1988) ต่อมา Chen and Liu (1993) ได้พัฒนาวิธีการประมาณค่าพารามิเตอร์ร่วมกับการประมาณค่าผลกระทบของค่านอกเกณฑ์

2) ใช้กระบวนทำซ้ำเพื่อตรวจหาค่านอก เกณฑ์และปรับแก้ไขตัวประมาณค่าพารามิเตอร์โดยใช้ตัวประมาณภาวะน่าจะเป็นสูงสุด เป็นกระบวนการที่นิยมใช้มาถึงปัจจุบัน

ค) กระบวนการตรวจสอบแบบตัดออก (Deletion Diagnostics) จงรักษ์ ศรีทิพย์ (2548, น. 20-32) ได้กล่าวไว้ว่า วิธีนี้เหมาะสมกับการตรวจหาค่านอกเกณฑ์ประเภท AO และ IO เช่นกัน โดยมีหลักการ คือ ตัดค่านอกเกณฑ์ออกจากชุดข้อมูล โดยถือว่าค่าสังเกตที่ถูกตัดออกไปนั้นเป็นค่าสูญหาย (Missing Observation) แล้วทำการประมาณค่าสูญหายและหาค่าประมาณพารามิเตอร์โดยใช้ข้อมูลชุดใหม่ ซึ่งได้จากการประมาณค่าสูญหายนี้ เพื่อวัดความเปลี่ยนแปลงของค่าประมาณพารามิเตอร์ที่ได้ เทียบกับค่าประมาณพารามิเตอร์ที่ใช้ข้อมูลชุดเดิม ถ้าพบว่าค่าสังเกตที่ถูกตัดออกจากชุดข้อมูลนั้นส่งผลทำให้ค่าประมาณพารามิเตอร์เปลี่ยนแปลงมาก แสดงว่าค่าสังเกตที่เป็นค่าสูญหายนี้เป็นค่านอกเกณฑ์ Bruce and Martin (1989, pp. 363-424) ได้กล่าวว่า วิธีการตรวจสอบแบบตัดออก k ค่า มีตัวสถิติทดสอบ 2 ตัว คือ DC (Diagnostics for Coefficients) และ DV (Diagnostics for the Innovational Variance) โดยถือว่าค่านอกเกณฑ์เป็นค่าสูญหายตัวหนึ่งนั่นเอง วิธีการประมาณค่าสูญหายของอนุกรมเวลามีหลายวิธี อาทิ เช่น การประมาณค่าสูญหายโดยใช้ค่าเฉลี่ยของข้อมูล (Series Mean) การประมาณค่าสูญหายโดยใช้ค่ามัธยฐานของค่าสังเกตที่อยู่ใกล้เคียง (Median of near by points) และวิธีการของ Jones (1980, pp. 389-395) เป็นต้น

ง) การตรวจสอบค่านอกเกณฑ์โดยใช้วิธี Grubbs' test จันทรัตน์ วรสรรพวิทย์ (2554, น. 1-5) ได้ศึกษาเกี่ยวกับการทดสอบค่านอกเกณฑ์โดยใช้วิธี Grubbs' test เป็นการตรวจสอบค่านอกเกณฑ์ โดยการหาอัตราส่วนค่าความแตกต่างระหว่างค่าที่สงสัยกับค่าเฉลี่ยของตัวอย่างกับค่าส่วนเบี่ยงเบนมาตรฐานที่ยังไม่ตัดค่าที่สงสัยออก นำค่าที่ได้ไปเทียบกับค่าวิกฤตที่กำหนดในตารางสถิติของ Grubb วิธีนี้สามารถนำมาใช้ทดสอบค่าที่สงสัยได้ครั้งละ 1 ข้อมูล หรือ 2 ข้อมูล มีหลักการ

1) สงสัยมีค่านอกเกณฑ์ 1 ค่า มีขั้นตอนดังนี้

- เรียงข้อมูลจากน้อยไปหามาก
- ตั้งสมมติฐานของการทดสอบ

H_0 : ค่าที่สงสัย 1 ค่า (ค่าน้อยที่สุด หรือค่ามากที่สุด) ไม่แตกต่างจากค่าอื่น

H_1 : ค่าที่สงสัย 1 ค่า (ค่าน้อยที่สุด หรือค่ามากที่สุด) แตกต่างจากค่าอื่น

- กำหนดระดับนัยสำคัญ (α) กำหนด α = 0.01 สำหรับการทดสอบ ค่านอกเกณฑ์ และ กำหนด

α = 0.05 สำหรับการทดสอบ Straggle ซึ่งเป็นเกณฑ์การเตือนก่อนข้อมูลเป็นค่านอกเกณฑ์
- คำนวณค่าเฉลี่ย ($\overline{X}$) และค่าส่วนเบี่ยงเบนมาตรฐาน (S) เมื่อ n คือ จำนวนข้อมูลที่ศึกษา
- คำนวณค่าสถิติทดสอบ (Gexp) จาก

กรณีทดสอบข้อมูลที่มีค่าน้อยที่สุด $G\exp = G_1 = \frac{(\overline{x} - x_1)}{s}$

กรณีทดสอบข้อมูลที่มีค่ามากที่สุด $G\exp = G_p = \frac{(x_p - \overline{x})}{s}$

- กำหนดค่าวิกฤต (Gcrit) จากตาราง Grubbs'test two-tailed ในตารางที่ 1 พิจารณาค่าวิกฤตในช่อง one largest or one smallest ที่ upper 1% สำหรับ α = 0.01 และ upper 5% สำหรับ α = 0.05

- สรุปผล ถ้าค่าสถิติจากการคำนวณ (Gexp) มากกว่าค่าวิกฤต (Gcrit) แสดงว่าค่าที่สงสัยนั่น เป็นค่านอกเกณฑ์

2) กรณีค่าที่สงสัยมี 2 ค่า มีขั้นตอนดังนี้

- เรียงข้อมูลจากน้อยไปหามาก - ตั้งสมมติฐานของการทดสอบ

H_0 : ค่าที่สงสัย 2 ค่า (ค่ามากที่ติดกันหรือค่าน้อยที่ติดกัน) ไม่แตกต่างจากค่าอื่น

H_1 : ค่าที่สงสัย 2 ค่า (ค่ามากที่ติดกันหรือค่าน้อยที่ติดกัน) แตกต่างจากค่าอื่น

- กำหนดระดับนัยสำคัญ α = 0.01 สำหรับการทดสอบค่านอกเกณฑ์ และ α = 0.05 สำหรับการทดสอบ Straggle

- คำนวณค่าสถิติทดสอบ

จากกรณีทดสอบค่าที่สงสัยมีค่ามาก สูตร $G\exp = \frac{s^2_{n-1,n}}{s^2_0}$ โดยที่ $s^2_0 = \sum_{i=1}^{n}(x_i - \overline{x})^2$

และ $s^2_{n-1,n} = \sum_{i=1}^{n-2}(x_i - \overline{x}_{n-1,n})^2$ และ $\overline{x}_{n-1,n} = \frac{1}{n-2}\sum_{i=1}^{n-2} x_i$

กรณีทดสอบค่าที่สงสัยมีค่าน้อย สูตร $G\exp = \frac{s^2_{1,2}}{s^2_0}$ โดยที่ $s^2_{1,2} = \sum_{i=3}^{p}(x_i - \overline{x}_{1,2})^2$

และ $\overline{x}_{1,2} = \frac{1}{n-2}\sum_{i=3}^{n} x_i$

- กำหนดค่าวิกฤต (Gcrit) จากตาราง Grubbs'test two-tailed ในตารางที่ 1 พิจารณาค่าวิกฤตในช่อง two largest or two smallest ที่ lower 1% สำหรับ α = 0.01 และ lower 5% สำหรับ α = 0.05

- สรุปผล ถ้าค่าสถิติจากการคำนวณ (Gexp) น้อยกว่าค่าวิกฤต (Gcrit) แสดงว่าค่าที่สงสัยเป็นค่านอกเกณฑ์ (outlier)

ตารางที่ 1 แสดงค่าวิกฤตของการทดสอบ Grubb (Grubbs' test two-tailed)

P	One largest or One smallest		Two largest or One smallest	
	Upper 1%	Upper 1%	Lower 1%	Lower 5%
7	2.139	2.020	0.0308	0.0708
8	2.274	2.126	0.0563	0.1101
9	2.387	2.215	0.0851	0.1492

n คือ จำนวนข้อมูล

วิธีการพยากรณ์ข้อมูลธุรกิจเมื่อมีค่านอกเกณฑ์

วิธีการพยากรณ์ข้อมูลทางธุรกิจเมื่อพบค่านอกเกณฑ์ปะปนอยู่ มีด้วยกันหลายวิธี อาทิ เช่น วิธีกำลังสองน้อยที่สุด (Ordinary Least Square Estimation Method) วิธีตัวประมาณภาวะน่าจะเป็นเมื่อปรับปรุงข้อมูล (Maximum Likeli- hood Estimation Method with Iterative for Adjusted Series) วิธีกำลังสองน้อยที่สุดเมื่อปรับปรุงข้อมูล (Ordinary Least Square Estimation Method with Iterative for Adjusted Series) วิธีกำลังสองน้อยสุดแบบมีเงื่อนไข (Condi-tional Least Squares Method : CLS) วิธีประมาณร่วมพารามิเตอร์ตัวแบบและผลกระทบของข้อมูลผิดปกติ (Joint Estimation of Model Parameters and Outliers Effect Method : JEMPOE) วิธีกำลังสองน้อยสุดแบบถ่วงน้ำหนักด้วยบูทสแทร็พ (Bootstrap Weighted Least Squares Method : BWLS) อีกทั้ง

วราฤทธิ์ พานิชกิจโกศลกุล และคณะ (2552, น. 54-66) ได้ศึกษาเกี่ยววิธีการพยากรณ์ทางธุรกิจเมื่อมีค่านอกเกณฑ์ไว้หลายวิธี คือ วิธีการพยากรณ์โดยใช้วิธีกำลังสองน้อยที่สุดแบบค่าเฉลี่ยเวียนเกิด (Recursive Mean Ordinary Least Squares Method : RM) วิธีการพยากรณ์โดยใช้วิธีกำลังสองน้อยที่สุดแบบมัธยฐานเวียนเกิด (Recursive Median Ordinary Least Squares Method : RMD) และวิธีการพยากรณ์โดยใช้วิธีกำลังสองน้อยที่สุดแบบมัธยฐานเวียนเกิดปรับปรุง (Improved Recursive Median Ordinary Least Squares Method : IRMD) พบว่า วิธี RM และวิธี IRMD ให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของค่าพยากรณ์ (Predict Mean Square Error : PMSE) ต่ำที่สุด แสดงว่าให้ค่าพยากรณ์ที่แม่นยำกว่า วิธี RMD และวิธี RM เป็นต้น ซึ่งแต่ละวิธีของการวิเคราะห์นั้นแตกต่างกันออกไป

สรุป

Hawkins (1980) ได้กล่าวว่า ค่านอกเกณฑ์ เป็นค่าสังเกตที่เบี่ยงเบนไปจากค่าสังเกตอื่นมาก จนทำให้สงสัยว่าค่าสังเกตนั้นได้มาจากวิธีอื่น ต่อมา Beckman and Cook (1983, p. 121) ได้ให้นิยามและกำหนดลักษณะของค่านอกเกณฑ์ไว้ 2 ลักษณะ คือ 1) ค่าสังเกตที่มีค่าสูงหรือต่ำมาก หรือเป็นค่าที่เบี่ยงเบนไปจากค่าสังเกตส่วนใหญ่ เรียกค่านอกเกณฑ์นี้ว่า Discordant Observations 2) ค่าสังเกตที่มีลักษณะการแจกแจงแตกต่างจากลักษณะการแจกแจงของประชากรที่สนใจศึกษา

เรียกค่านอกเกณฑ์นี้ว่า Contaminate Observations กรณีที่ค่านอกเกณฑ์เป็นค่าสังเกตที่มีความน่าสนใจในตัวเอง ถ้าตัดค่าสังเกตส่วนนี้ออกก่อนที่จะทำการวิเคราะห์ข้อมูล อาจทำให้ขาดสารสนเทศบางอย่างไป แต่ถ้ากรณีที่ค่านอกเกณฑ์เกิดขึ้นเนื่องจากความคลาดเคลื่อนในการวัด หรือเกิดจากการปฏิบัติการที่มีความคลาดเคลื่อน หรือหน่วยตัวอย่างที่เก็บรวบรวมได้ไม่ใช่หน่วยตัวอย่างในประชากรเป้าหมาย ก็สามารถตัดค่าสังเกตที่มีความผิดปกติออกไปได้ เพื่อให้ผลที่ได้จากการวิเคราะห์มีความเชื่อถือได้ การตรวจสอบค่านอกเกณฑ์จึงถือเป็นเรื่องสำคัญที่ควรกระทำก่อนที่จะลงมือวิเคราะห์ข้อมูล ซึ่งจะทำให้ผู้วิเคราะห์ข้อมูลเห็นลักษณะโครงสร้างข้อมูลและหาวิธีการแก้ปัญหาด้วยการเลือกวิธีที่ได้กล่าวไว้แล้วในตอนต้นอย่างใดอย่างหนึ่งที่เหมาะสม

ค่านอกเกณฑ์ที่มีค่าผิดปกติมากซึ่งอาจเรียกค่านอกเกณฑ์ลักษณะนี้เรียกว่าค่าสุดขีด (Extreme Value) สามารถสังเกตและหาแนวทางการตรวจสอบได้ง่าย แต่ค่านอกเกณฑ์ที่มีค่าผิดปกติเล็กน้อย (Mild Value) ควรตรวจสอบด้วยวิธีการทางสถิติที่เหมาะสม เช่น การตรวจสอบค่านอกเกณฑ์ด้วยวิธีกราฟ หรือการทดสอบสมมติฐาน เป็นต้น

อภิปราย

วิธีการตรวจหาค่านอกเกณฑ์ด้วยกระบวนการทำซ้ำ เป็นวิธีการที่ได้รับการพัฒนาจากนักวิจัยหลายท่านอย่าง ต่อเนื่อง อาทิ Fox (1972) เป็นผู้ริเริ่มพัฒนาการทดสอบ Likelihood Ratio Test ใช้ในการตรวจสอบค่านอกเกณฑ์ในข้อมูลอนุกรมเวลาภายใต้ตัวแบบ AR(p) (Autoregressive Model : AR(p)) ต่อมา Chang and Tiao (1983) ได้พัฒนากระบวนการตรวจสอบโดยใช้ Likelihood Ratio Test ของ Fox ภายใต้ตัวแบบ ARMA (p,q) (Auto Regressive–moving-average Model : ARMA (p,q)) ต่อมา Tsay (1986) พบว่า ค่านอกเกณฑ์ที่ตรวจพบตามแนวทางของ Chang and Tiao มีผลกระทบต่อตัวสถิติที่ใช้ในการกำหนดตัวแบบ เช่น ฟังก์ชันสหสัมพันธ์ในตัวเอง (Autocorrelation Function : ACF) ฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วน (Partial Autocorrelation Function : PACF) จึงได้ใช้ฟังก์ชันขยายของสหสัมพันธ์ในตัวเอง (Extended Autocorrelation Function : EACF) ในการกำหนดตัวแบบ ARMA (p,q) และประมาณค่าพารามิเตอร์ด้วยวิธี OLS ต่อจากนั้น Chang, Tiao and Chen (1988) ได้ร่วมกันพัฒนาสร้างค่าวิกฤติสำหรับตรวจ สอบค่านอกเกณฑ์ แต่ยังด้อยประสิทธิภาพ Chen and Liu (1993) ได้นำแนวทางของ Chang, Tiao and Chen (1988) มาพัฒนาเพิ่มขึ้น ได้ใช้ตัวประมาณภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Ratio Test : MLE) ร่วมกับการประมาณค่าผลกระทบของค่านอกเกณฑ์ภายใต้ตัวแบบ ARIMA (p,d,q) และใช้เกณฑ์ RMSE (Root Mean Square Error: RMSE) มาพิจารณาความถูกต้องของตัวประมาณพารามิเตอร์ที่ปรับแก้ไขค่านอกเกณฑ์แล้ว ต่อมา Choy (2001) ได้ปรับปรุงกระบวนการทำซ้ำ โดยใช้วิธีการประมาณค่าแบบ Whittle-Type มาประมาณค่าพารามิเตอร์แทนวิธี MLE หรือ OLS เรียกวิธีใหม่นี้ว่า Spectrum-Based Outlier Detection Algorithm (SODA) ใช้ได้ดีสำหรับตัวอย่างขนาดใหญ่ วิธีการตรวจสอบค่านอกเกณฑ์โดยใช้กระบวนการทำซ้ำ สามารถตรวจหาค่านอกเกณฑ์ได้ ทั้งประเภท AO และ IO แต่มีบางกรณีข้อมูลอาจจะมีค่านอกเกณฑ์มากกว่า 1 ค่า และค่านอกเกณฑ์เหล่านี้อยู่ติดกัน (วิธีการนี้อาจตรวจไม่พบ) วิธีการตรวจสอบแบบตัดออกสามารถตรวจพบค่านอกเกณฑ์ในกรณีดังกล่าวได้ดีกว่า แต่วิธีการตรวจสอบแบบตัดออกไม่ได้คำนึงถึงประเภทของค่านอกเกณฑ์และมีขั้นตอนค่อนข้างยุ่งยากและ

ใช้เวลาในการคำนวณมาก ประสิทธิภาพของวิธีนี้ยังขึ้นอยู่กับการประมาณค่าสูญหายว่าได้ค่าใกล้เคียงกับค่าที่แท้จริงเพียงใดด้วย สำหรับวิธี Grubbs' test เป็นวิธีที่ไม่ยุ่งยากใช้เวลาน้อยเหมาะสมกับการใช้ในห้องปฏิบัติการเกี่ยวกับการควบคุมคุณ ภาพมากกว่า

ข้อเสนอแนะ

สำหรับในบทความวิชาการนี้ ได้นำเสนอวิธีการตรวจหาและปรับแก้ไขค่านอกเกณฑ์ ตลอดจนกระบวนการสร้างตัวแบบเพื่อการพยากรณ์ที่เหมาะสมแล้วนั้น ยังมีวิธีการตรวจหาค่านอกเกณฑ์ในข้อมูลธุรกิจที่เก็บรวบรวมข้อมูลตามช่วงระยะเวลาที่น่าสนใจ อาทิ Franses and Ghijsel (1999, pp. 1-9) ได้พัฒนาวิธีการตรวจหา AO และ IO โดยใช้ตัวแบบ GARCH และ Charles and Darne (2005, pp. 347-352) ได้ขยายแนวคิดและปรับปรุงวิธีของ Franses and Ghijsel (1999) ให้มีความทันสมัยจนถึงปัจจุบัน

เอกสารอ้างอิง

จงรักษ์ ศรีทิพย์. (2548). *การตรวจหาค่านอกกลุ่มในอนุกรมเวลาโดยกระบวนการทำซ้ำและการตรวจสอบแบบตัดออก k ค่า.* สืบค้นเมื่อ 28 ส.ค.2558, จาก http://www.thapra.lib.su.ac.th/thesis/listprogram_th.asp?program=0703.

จันทรัตน์ วรสรรพวิทย์. (2554). การทดสอบค่าสุดต่างโดยใช้ Grubbs' test. *วารสารออนไลน์สำนักบริหารและรับรองห้องปฏิบัติการ, 7*(19), 1-5.

นิภาพร ปัญญายงค์. (2541). *การศึกษาเปรียบเทียบวิธีการตรวจสอบค่าผิดปกติหนึ่งค่าจากการถดถอยเชิงเส้นอย่างง่าย.* (วิทยานิพนธ์ปริญญามหาบัณฑิต). กรุงเทพฯ: มหาวิทยาลัยเกษตรศาสตร์.

พีทีที โกลบอล เคมิคอล จำกัด (มหาชน). (2554). รายงานกราฟราคาหุ้นย้อนหลัง 5 ปี. (ข้อมูลราคาหลักทรัพย์, พีทีที โกลบอล เคมิคอล จำกัด (มหาชน), กรุงเทพฯ) สืบค้นเมื่อ 28 ส.ค. 2558, จาก http://pttgc-h.listedcompany.com/stock_chart_history.html.

ภูริทัต นาคประเสริฐ. (2554). พยากรณ์แนวโน้มราคาและปริมาณการนำเข้าน้ำมันดิบของประเทศไทย. (การค้นคว้าแบบอิสระ, มหาวิทยาลัยเชียงใหม่, เชียงใหม่). สืบค้นจาก http://www.library.cmu.ac.th/faculty/econ/Exer 751409/2554/Exer2554_no 192.

มนตรี สิงหะวาระ. (2549). *เศรษฐศาสตร์การจัดการ.* สืบค้นเมื่อ 25 กรกฎาคม 2557, จาก http://www.coursewares.mju.ac.th:81/e-learning49/EC373/content5.html

วราฤทธิ์ พานิชกิจโกศลกุล และคณะ. (2552). การเปรียบเทียบวิธีการพยากรณ์ข้อมูลล่วงหน้าหนึ่งคาบเวลาสำหรับตัวแบบอัตตสหสัมพันธ์อันดับที่หนึ่ง เมื่อข้อมูลมีค่าผิดปกติ. *วารสารวิชาการ*, มหาวิทยาลัยหอการค้าไทย, 29(4), 54 - 65.

วราฤทธิ์ พานิชกิจโกศลกุล. (2554). การเปรียบเทียบวิธีการพยากรณ์ข้อมูลล่วงหน้าหนึ่งคาบเวลาสำหรับตัวแบบอัตตสหสัมพันธ์อันดับที่หนึ่งเมื่อข้อมูลมีค่าผิดปกติและค่าสูญหาย. *วารสารวิทยาศาสตร์,* มหาวิทยาลัยอุบลราชธานี, ฉบับพิเศษ, 1(2554), 1-9.

สำนักนโยบายการออมและการลงทุน. (2548, 27 พฤษภาคม). เศรษฐศาสตร์น่ารู้: อนุกรมเวลา. สำนักงานเศรษฐกิจการคลัง, กรุงเทพฯ) สืบค้นเมื่อ 28 ส.ค. 2558, จาก http://www.fpo.go.th/S-I/Source/ECO/ECO24.htm.

Abrahom, B. & Chuang, A. (1989). Outlier Detection and Time Series Modeling. *Technomertrics,* 31, 241-248.

Atsushi I., Shunsuke E. & Tetsuya S. (2010). Treatment of Outliers in Business Surveys: The Case of Short-term Economic Survey of Enterprises in Japan (Tankan). *Bank of Japan Working Paper Series.* September, 3-4.

Beckman, R. J. & Cook, R. D. (1983). Outliers. *Technomertrics. 2*(25), 119-149.

Beckman, R. J. & Cook, R. D. (1983). Outliers: *Response. Technomertrics. 2*(25), 161-163.

Bruce, A.G. & Martin, R.D. (1989). Leave-k-out Diagnostics for Time Series. *Journal of the royal Statistical Society Ser B.* 5, 363-424.

Chang, I. and Tiao, G. C. (1983). Estimation of time series parameters in the presence of outliers. *Technical Report 8,* University of Chicago, Statistical Research Center, USA.

Chang, I., Tiao, G.C. & Chen, C. (1989). Estimation of Time Series Parameters in the Presence of Outliers. *Technomertrics.* 88, 193-203.

Charles, A. & Darne, O. (2005). Outliers and GARCH models in financial data. *Economics Letters.* 86, 347-352.

Chen, C. & Lui, L.M. (1993). Joint Estimation of Model Parameter and Outliers Effects in Time Series. *Journal of The American Statistical Association.* 88, 284-297.

Chen, C. & Lui, L.M. (1993). Forecasting Time series with Outliers Series. *Journal of Forecasting.* 12, 13-35.

Choy, K. (2001). Outliers Detection for stationary time series. *Journal of statistical planning and inference.* 99, 111-127.

Fox, A.J. (1972). Outliers in time series. *Journal of the royal Statistical Society Ser.B,* 34, pp. 350-363.

Franses, P. & Ghijsels, H. (1999). Additive Outliers, GARCH and Forecasting Volatility. *International Journal of Forecasting.* 15, 1–9.

Fuller, W. A. (2009). *Sampling Statistics.* New York: Wiley Online Library, 309-311.

Hawkins, D.M. (1978). Analysis of three tests for one or two outliers. *Statistica Neerlandica,* 32, 137-148.

Hawkins, D.M. (1980). *Identification of Outliers.* New York: Chapman & Hall.

Jones, R.H. (1980). Maximum likelihood Fitting of ARMA Models to Time Series. *Technometrics.* 22, 389-395.

Lohr, S.L. (1999). *Sampling : Design and Analysis*. Pacific Grove, CA : Duxbury, 330-332.

Vaage, K. (2000). Detection of outliers and level shifts in time series: An evaluation of two alternative procedures. *Journal of Forecasting.* 19, 23-37.

Trivez, F.J. (1993). Level Shifts, Temporary Changes and Forecasting. *Journal of Forecasting.* 14, 543-550.

Tsay, R.S. (1986). Time Series Model Specification in the Presence of Outliers. *Journal of The American Statistical Association.* 81, 32-141.

Tsay, R.S. (1988). Outlier, level shift, and variance changes in time series. *Journal of Forecasting.* 7, 1-20.

Vandaele. W. (1983). *Applied time series and Box-Jenkins models.* St. Louis, Missouri: Academic Press.