

ในการวิเคราะห์ข้อมูลอนุกรมเวลาและการพยากรณ์นั้นมักไม่มีความแม่นยำ ถูกต้อง และน่าเชื่อถือเมื่อมีข้อมูลสูญหายเกิดขึ้น งานวิจัยนี้มีวัตถุประสงค์ในการศึกษาวิธีเพื่อประมาณค่าสูญหายเพื่อใช้ในการวิเคราะห์ตัวแบบ พยากรณ์ โดยทำการเปรียบเทียบวิธีการประมาณค่าสูญหาย 3 วิธีเดี่ยว คือวิธีค่าเฉลี่ย วิธีกำลังสองน้อยสุด วิธีการใส่ค่าหลายค่าแทนข้อมูลที่สูญหายแต่ละค่า และการประมาณค่าสูญหายร่วมด้วยตัวถ่วงน้ำหนัก 3 วิธี คือ วิธีการถ่วงน้ำหนักที่เท่ากัน วิธีถ่วงน้ำหนักโดยค่าสัมบูรณ์ต่ำสุด และวิธีการถ่วงน้ำหนักโดยความแปรปรวนต่ำสุด โดยจำลองข้อมูลด้วยเทคนิคมอนติคาร์โลด้วยโปรแกรม R จำนวน 50,000 ครั้งและกำหนดขนาดตัวอย่าง 30, 50 และ 100 ค่าเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนเท่ากับ 1, 2, 5 และ 10 เปอร์เซนต์การสูญหายเท่ากับ 5%, 10%, 15% และ 20% ขนาดความสัมพันธ์ของตัวแปรอิสระ คือ 0, 0.2 และ 0.5 เกณฑ์ที่ใช้ในการเปรียบเทียบ คือ ค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์

วิธีที่เหมาะสมกับ การถดถอยเชิงเส้นพหุ เมื่อเป็น ข้อมูลภาคตัดขวางคือ วิธีค่าเฉลี่ย วิธีที่เหมาะสมกับ การถดถอยเชิงเส้น เมื่อเป็น ข้อมูลอนุกรมเวลา คือ วิธีกำลังสองน้อยสุด สำหรับทุกกรณีศึกษามีเพียง 3 กรณี วิธีค่าเฉลี่ยเหมาะสมกับข้อมูลดังนี้ ค่าเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนเท่ากับ 5 ขนาดตัวอย่างเท่ากับ 30, ค่าเบี่ยงเบนมาตรฐาน ของความคลาดเคลื่อนเท่ากับ 10 ขนาดตัวอย่างเท่ากับ 30 และค่าเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนเท่ากับ 5 ขนาดตัวอย่างเท่ากับ 50 สำหรับวิธีการประมาณค่าร่วมกันนั้นวิธีถ่วงน้ำหนักโดยค่าสัมบูรณ์ต่ำสุดจะเหมาะสมทุกกรณี

Abstract

228963

An analysis of time series and forecasting is usually inaccurate, invalid and unreliable when there are missing data. The objective of this research is to compare missing value estimation methods and to propose a new method by combining individual methods so called "combined missing value estimation methods". The proposed method combines three individual missing value estimation methods by using weighted average. There are three methods for evaluating weighted averages which are Equally Weight (EW), Least Absolute Value (LAV) and Minimum Variance (MV). The Monte Carlo simulation technique is used in the R program with 50,000 trajectory samples, where sample sizes are 30, 50 and 100, standard deviations of error are 1, 2, 5 and 10, missing percentages are 5%, 10%, 15% and 20% and correlation levels of independent variables are 0, 0.2 and 0.5. The Mean Absolute Percentage Error is used as a criterion for comparison of missing value estimation methods. The proper method for cross – sectional data is the Mean method. For time series data, the Mean method is suitable for three cases; 1) standard deviation of the error is 5 and the sample size is 30 2) standard deviation of the error is 10 and the sample size is 30 3) standard deviation of the error is 10 and the sample size is 50, otherwise, the Ordinary Least Square Method (OLS) is the best estimation method. The weighted of combined estimation method by least absolute value is superior to other methods for all cases studies.