การเปรียบเทียบวิธีการประมาณค่าสูญหายในการวิเคราะห์การถดถอยเชิงเส้น

นางสาว เพียงออ ยีสา

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาสถิติศาสตรมหาบัณฑิต

สาขาวิชาสถิติ ภาควิชาสถิติ

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2551

COMPARISON OF THE ESTIMATION METHODS FOR MISSING DATA IN MULTIPLE LINEAR REGRESSION

Miss Peang – or Yeesa

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Statistics
Department of Statistics
Faculty of Commerce and Accountancy
Chulalongkorn University
Academic Year 2008

หัวข้อวิทยานิพนธ์	การเปรียบเทียบวิธีการประมาณค่าสูญหายในการวิเคราะห์การถดถอยเชิงเส้น
โดย	นางสาว เพียงออ ยีสา
สาขาวิชา	สถิติ
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	รองศาสตราจารย์ ดร.กัลยา วานิชย์บัญชา

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

....คณบดีคณะพาณิชยศาสตร์และการบัญชี
(รองศาสตราจารย์ ดร.อรรณพ ต้นละมัย)

คณะกรรมการสอบวิทยานิพนธ์

....ประธานกรรมการ
(รองศาสตราจารย์ ดร.ธีระพร วีระถาวร)

.... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(รองศาสตราจารย์ ดร.กัลยา วานิชย์บัญชา)

.... กรรมการ
(อาจารย์ ดร.อรุณี กำลัง)

.... กรรมการภายนอกมหาวิทยาลัย
(รองศาสตราจารย์ ชูศักดิ์ อุดมศรี)

เพียงออ ยีสา: การเปรียบเทียบวิธีการประมาณค่าสูญหายในการวิเคราะห์การถดถอยเชิงเส้น. (Comparison of the Estimation Methods for Missing Data in Multiple Linear Regression)
อ.ที่ปรึกษาวิทยานิพนธ์หลัก: รศ. ดร. กัลยา วานิชย์บัญชา, 106 หน้า.

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบวิธีการประมาณค่าสูญหายของตัวแปรตามในการวิเคราะห์การถดถอยเชิงเส้นพหุเพื่อการพยากรณ์โดยพิจารณาข้อมูล 2 ลักษณะ คือ ข้อมูลภาคตัดขวาง และข้อมูลอนุกรมเวลาที่มีปัจจัยด้านแนวโน้ม และปัจจัยฤดูกาลเข้ามาเกี่ยวข้องโดยทำการประมาณค่าสูญหายด้วยวิธี Regression Imputation (RI) วิธี Nearest Neighbor Imputation (NNI) วิธี Weighted Nearest Neighbor and Regression Imputation (WNR) และวิธี EM algorithm (EM) โดยจำลองข้อมูลซึ่งกำหนดขนาดตัวอย่าง 50 , 100 และ 200 ส่วนเบี่ยงเบนมาตรฐาน 5,10,15,20 และ 25 ร้อยละการสูญหายของตัวแปรเป็น 5 , 10 และ 20 ตามลำดับ ซึ่งกำหนดให้ตัวแปรอิสระมีการแจกแจงแบบปกติ เกณฑ์ในการเปรียบเทียบประสิทธิภาพในการประมาณค่าสูญหายจะใช้ค่า MAPE

ผลการวิจัยสรุปได้ดังนี้ สำหรับข้อมูลภาคตัดขวาง กรณีที่ค่าสหสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระทั้ง 2 ตัวสูง เมื่อส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับต่ำถึงปานกลางวิธี RI และ EM มีค่า MAPE ต่ำกว่าวิธีอื่น เมื่อส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับสูงวิธี WNR ให้ผลดีกว่าวิธีอื่นๆ ที่นำมาเปรียบเทียบ กรณีที่ค่าสหสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระตัวหนึ่งสูงมากและอีกตัวหนึ่งปานกลาง วิธี RI และ EM จะให้ผลดีกว่าวิธีอื่นๆ สำหรับข้อมูลอนุกรมเวลา วิธี WNR มีค่า MAPE ต่ำกว่าวิธีการประมาณค่าอื่นๆในกรณีที่ข้อมูลที่มีอิทธิพลของฤดูกาลสูง และวิธี NNI จะให้ผลดีเมื่อส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับสูง สำหรับข้อมูลที่มีอิทธิพลจากปัจจัยแนวโน้มสูงวิธี RI และ EM เป็นวิธีที่ให้ผลดีกว่าวิธีอื่นๆที่นำมาเปรียบเทียบ กรณีที่ข้อมูลมีอิทธิพลจากปัจจัยแนวโน้มและปัจจัยฤดูกาลระดับปานกลาง เมื่อส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับต่ำ วิธี RI และ EM มีค่า MAPE ต่ำกว่าวิธีการประมาณค่าอื่นๆ เมื่อส่วนเบี่ยงเบนมาตรฐานเพิ่มสูงขึ้น วิธี WNR เป็นวิธีที่มีค่า MAPE ต่ำกว่าวิธีการประมาณค่าอื่นๆที่นำมาเปรียบเทียบ

ภาควิชา	สถิติ	ลายมือชื่อนิสิต....เพียงออ ยีสา....
สาขาวิชา	สถิติ	ลายมือชื่ออ.ที่ปรึกษาวิทยานิพนธ์หลัก..........
ปีการศึกษา	2551	

4982205826: MAJOR STATISTICS

KEYWORDS: MISSING DATA / REGRESSION IMPUTATION / NEAREST NEIGHBOR IMPUTATION / WEIGHTED NEAREST NEIGHBOR REGRESSION IMPUTATION / EM ALGORITHM

PEANG-OR YEESA: COMPARISON OF THE ESTIMATION METHODS FOR MISSING DATA IN MULTIPLE LINEAR REGRESSION. ADVISOR: ASSOC. PROF. KANLAYA VANICHBUNCHA, Ph.D, 106 pp.

The purpose of this research is to study and compare the estimation methods for missing data of the dependent variable in multiple linear regression equation for forecasting. In this study, we focus on cross-sectional data and time series data, which involves trend and seasonal factor. The methods used to estimate missing data are Regression Imputation (RI), Nearest Neighbor Imputation (NNI), Weighted Nearest Neighbor and Regression Imputation (WNR) and EM algorithm method. The comparison was done under condition of sample size of 50, 100 and 200; standard deviations of error of 5,10,15,20 and 25; percentage of missing data of 5%,10% and 20% and the distribution of independent variables are normal. The criterion of determination is Mean Absolute Percentage Error (MAPE). The result for cross-sectional data shows that when correlation between dependent variable and two independent variables is high, the MAPE of RI and EM are lower than the MAPE of other methods which standard deviation is from low to medium level. The MAPE of WNR is the lowest when standard deviation is high. In the case when correlation between dependent and independent variable is high and medium, it is found that RI and EM are better than other methods. In case of time series data, WNR is better than other methods for high influence from seasonal factor. However, if standard deviation increases, NNI is more suitable. The MAPE of RI and EM are lower than MAPE of other methods for high influence from trend factor. For medium influence from trend and seasonal factor, the MAPE of RI and EM are lower than other methods when standard deviation is low. If standard deviation increases, the MAPE of WNR is the lowest.

Department: Statistics — Student's signature Peang-or Yeesa

Field of Study: Statistics — Advisor's signature Kanlaya Vanichbuncha

Academic Year: 2008

กิตติกรรมประกาศ

งานวิจัยฉบับนี้สามารถสำเร็จลุล่วง ได้ด้วยความกรุณาและความช่วยเหลืออย่างดียิ่งจาก รองศาสตราจารย์ ดร.กัลยา วานิชย์บัญชา อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ให้คำแนะนำปรึกษา ตลอดจนช่วยเหลือแก้ไขข้อบกพร่องต่างๆ จนงานวิจัยฉบับนี้เสร็จสมบูรณ์ ผู้วิจัยจึงขอกราบขอบพระคุณเป็นอย่างสูง

ผู้วิจัยขอกราบขอบพระคุณ รอง ศาสตราจารย์ ดร.ธีระพร วีระถาวร ในฐานะประธานกรรมการ อาจารย์ ดร.อรุณี กำลัง ในฐานะกรรมการสอบวิทยานิพนธ์ และรองศาสตราจารย์ชูศักดิ์ อุดมศรี ในฐานะกรรมการภายนอกมหาวิทยาลัย ที่กรุณาตรวจสอบวิทยานิพนธ์ฉบับนี้ให้มีความสมบูรณ์ยิ่งขึ้น ขอกราบขอบพระคุณคณาจารย์ประจำภาควิชาสถิติที่ให้โอกาสทางการศึกษาและประสิทธิ์ประสาทความรู้จนกระทั่งสำเร็จการศึกษา

ผู้วิจัยขอกราบขอบพระคุณ บิดา – มารดา ซึ่งสนับสนุนด้านการ เรียน ให้กำลังใจตลอดจนให้คำแนะนำต่างๆ แก่ผู้วิจัยเสมอมาจนสำเร็จการศึกษา และขอขอบคุณเพื่อนๆที่คอยช่วยเหลือผู้วิจัย รวมทั้งให้กำลังใจในเรื่องต่างๆ มา ณ โอกาสนี้ด้วย

สารบัญ

หน้า

บทที่ หน้า

สารบัญตาราง

หน้า

หน้า

สารบัญภาพ

หน้า

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันการพยากรณ์เป็นเทคนิคหนึ่งที่นิยมใช้กันทั่วไปในวงการธุรกิจ เนื่องจากสภาพสังคม และ สิ่งแวดล้อมมีความซับซ้อน และ มีการเปลี่ยนแปลงที่รวดเร็วความผิดพลาดที่เกิดจากการวางแผนและการตัดสินใจในการดำเนินงานอาจเกิดขึ้นได้ง่าย เพื่อลดความเสี่ยงจากความไม่แน่นอนของเหตุการณ์ที่เกิดขึ้นจึงจำเป็นต้องใช้เทคนิคการพยากรณ์ ซึ่งช่วยในการคาดการณ์เกี่ยวกับภาวะต่างๆทางธุรกิจได้ล่วงหน้า ทำให้สามารถวางแผนงานและตัดสินใจสำหรับการดำเนินงานต่างๆได้ดียิ่งขึ้น

การพยากรณ์เชิงปริมาณเป็นการพยากรณ์ที่อาศัยความรู้ทางด้านคณิตศาสตร์และสถิติไปสร้างรูปแบบหรือสมการพยากรณ์โดยใช้ข้อมูลในอดีต ซึ่งสามารถแบ่งออกเป็น 2 ประเภทใหญ่ๆ คือ ประเภทที่ 1 การพยากรณ์เชิงอธิบาย (Explanatory Forecasting) หรือการพยากรณ์แบบเป็นเหตุเป็นผล เป็นการพยากรณ์โดยมีแนวคิดว่าพฤติกรรมของสิ่งที่จะพยากรณ์ถูกกำหนดหรือถูกอธิบายโดยสิ่งอื่นๆ ซึ่งมีความสัมพันธ์บางลักษณะกับสิ่งที่จะพยากรณ์ สำหรับการพยากรณ์ด้วยการหาความสัมพันธ์แบบนี้นั้น สามารถจะใช้พยากรณ์ได้ทุกช่วงเวลา และจำเป็นต้องเก็บรวบรวมข้อมูลจำนวนมาก แต่ให้ความแม่นยำของการพยากรณ์ค่อนข้างสูง ได้แก่เทคนิคการวิเคราะห์การถดถอย (Regression Analysis) และการจำลองเศรษฐมิติ (Econometric Model) และประเภทที่ 2 การพยากรณ์แบบอนุกรมเวลา (Time Series Forecasting) เป็นการพยากรณ์โดยมีแนวความคิดว่า พฤติกรรมของสิ่งที่จะพยากรณ์ควรจะเพียงพอที่จะพยากรณ์พฤติกรรมในอนาคตได้ ได้แก่ การวิเคราะห์อนุกรมเวลาแบบแยกส่วน(Decomposition or Classical Method) เทคนิคการทำให้เรียบ (Smoothing Technique) การพยากรณ์แบบการกรองแบบปรับได้ (Adaptive Filtering) และการวิเคราะห์อนุกรมเวลาแบบ Box-jenkins เป็นต้น

ข้อมูลสูญหายเป็นปัญหาหนึ่งที่พบได้เสมอในการนำข้อมูลไปใช้วิเคราะห์โดยเทคนิคต่างๆทางสถิติ ซึ่งข้อมูลสูญหาย(Missing Data) นี้เป็นค่าสังเกตที่ต้องการทราบค่าแต่ไม่สามารถทราบค่าได้ โดยผู้วิจัยจะต้องหาทางแก้ไขปัญหาในส่วนนี้เพื่อทำให้การวิเคราะห์ข้อมูลมีความสมบูรณ์มากที่สุดและเบี่ยงเบนน้อยที่สุด

ดังนั้นในการพยากรณ์ เมื่อเกิดการสูญหายของข้อมูลจะส่งผลให้เกิดความคลาดเคลื่อนซึ่งการสูญหายไปนี้อาจเกิดขึ้น จากการไม่ตอบของหน่วยตัวอย่างบางหน่วย (Unit Non-response) หรือเกิดจากการไม่ตอบเฉพาะบางคำถามหรือเฉพาะบางตัวแปร (Item Non-response)

การจัดการกับข้อมูลสูญหายมีหลายวิธีการให้เลือกใช้ การเลือกใช้วิธีการใดวิธีการหนึ่งขึ้นอยู่กับลักษณะข้อมูลสูญหายที่เกิดขึ้นหากเลือกวิธีการที่ไม่เหมาะสมอาจเป็นการเพิ่มค่าความคลาดเคลื่อนและมีผลกระทบต่อการวิเคราะห์ข้อมูล ซึ่งวิธีการจัดการกับข้อมูลสูญหายกรณีข้อมูลสูญหายเกิดจากการไม่ตอบเฉพาะบางคำถามหรือบางตัวแปร (Item Non-response) แบ่งออกเป็น 2 กลุ่มหลัก (Laaksonen,2000) คือ

1. Model-donor Imputation คือ การประมาณค่าที่ได้มาจากตัวแบบ (Model) โดยตรง ได้แก่ Mean Imputation , Regression Imputation และ Ratio Imputation เป็นต้น
2. Real-donor Imputation คือ การประมาณค่าที่ได้จากเซตข้อมูลของค่าที่สังเกตได้ ได้แก่ Cold Deck Imputation , Hot Deck Imputation และ Nearest Neighbor Imputation เป็นต้น

ชุติมา (2533) ได้ศึกษาเปรียบเทียบวิธีการประมาณค่าสูญหายในการวิเคราะห์การถดถอยเชิงซ้อนเมื่อประมาณข้อมูลของตัวแปรอิสระที่สูญหายด้วยวิธีวิเคราะห์ความถดถอย วิธี Maximum Likelihood วิธีค่าเฉลี่ย และวิธีมัธยฐานโดยใช้ค่าคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของสมการถดถอยเป็นเกณฑ์ในการเปรียบเทียบ ซึ่งศึกษาจำแนกตามสถานการณ์ต่างๆโดยใช้การกระจายของข้อมูล ขนาดตัวอย่าง จำนวนตัวแปรอิสระ ส่วนเบี่ยงเบนมาตรฐาน และเปอร์เซ็นต์การสูญหายของข้อมูลซึ่งเท่ากันในทุกตัวแปรเป็นตัวกำหนด พบว่าวิธีการประมาณค่าสูญหายในการวิเคราะห์การถดถอยเชิงซ้อนทั้ง 4 วิธีให้ผลแตกต่างกันตามสถานการณ์ซึ่งโดยส่วนใหญ่วิธีค่าเฉลี่ยให้ผลดีที่สุด ยกเว้นเมื่อมีขนาดตัวอย่างน้อยและจำนวนตัวแปรอิสระมาก วิธีวิเคราะห์ความถดถอยจะให้ผลดีที่สุด แต่ถ้าขนาดตัวอย่างใหญ่และจำนวนตัวแปรอิสระน้อยการตัดชุดของข้อมูลสูญหายทิ้งจะไม่มีผลกระทบต่อการวิเคราะห์การถดถอยด้วยวิธีกำลังสองน้อยที่สุด

วารุณี (2537) ได้ศึกษาเปรียบเทียบการประมาณค่าสูญหายของตัวแปรตามในสมการถดถอยเชิงเส้นพหุเพื่อการพยากรณ์ โดยทำการประมาณค่าสูญหายของตัวแปรตามด้วยวิธีสูญหาย วิธีค่าเฉลี่ย วิธีสมการถดถอย วิธีอีเอ็ม และวิธีการของฮันท์ การเปรียบเทียบกระทำภายใต้สถานการณ์ของขนาดตัวอย่าง ค่าเบี่ยงเบนมาตรฐานของความคลาดเคลื่อน สัดส่วนการสูญหายของตัวแปรตาม และลักษณะของตัวแปรอิสระที่แตกต่างกัน และหาค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) ของค่าพยากรณ์จากวิธีการทั้ง 5 พบว่าในสถานการณ์ที่ขนาดตัวอย่างมีขนาดเล็ก (10-20) เมื่อค่าเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนมีขนาดไม่ใหญ่นักและสัดส่วนของการสูญหายของตัวแปรตามมีจำนวนมาก (60% – 70%) วิธีการของฮันท์จะให้ค่า RMSE ของค่าพยากรณ์ต่ำกว่าวิธีการอื่นๆ แต่ถ้าค่าเบี่ยงเบนมาตรฐานของความ

คลาดเคลื่อนมีขนาดเพิ่มขึ้น วิธีค่าเฉลี่ยจะให้ค่า RMSE ต่ำกว่าวิธีอื่นๆ ในทุกสัดส่วนของการสูญหายของตัวแปรตาม ส่วนในสถานการณ์ที่ขนาดตัวอย่างมีขนาดปานกลางถึงใหญ่ (30–70) วิธีสูญหายจะเหมาะสมเกือบทุกกรณี นั่นคือ ถ้าขนาดตัวอย่างใหญ่พอ การตัดชุดข้อมูลสูญหายทิ้งจะมีผลกระทบน้อยมากกับผลการวิเคราะห์การถดถอยเชิงเส้นพหุ โดยใช้วิธีกำลังสองน้อยที่สุด

Watchareeporn & Prachoom (2005) ในงานวิจัยนี้ต้องการเปรียบเทียบวิธีการประมาณค่าสูญหายที่เคยพัฒนาไว้ในงานวิจัยก่อนหน้านี้ คือ วิธี Weighted Nearest Neighbor and Regression Imputation (WNR) ในการวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย กับวิธีการประมาณค่าสูญหายอีก 3 วิธีเมื่อมีการสูญหายของข้อมูลตัวแปรตาม คือ วิธีค่าเฉลี่ย วิธี Hot Deck Imputation วิธี Random Regression Imputation เมื่อใช้ตัวอย่างสุ่มที่ได้จากประชากรขนาด $n = 100$ พบว่า เมื่อพิจารณาค่า R^2 วิธี WNR เป็นวิธีที่ดีกว่าวีธีอื่นๆ สำหรับทุกระดับการสูญหายของข้อมูล และวิธีนี้ยังมีค่า Mean Absolute Deviation (MAD) น้อยที่สุด สำหรับทุกๆระดับการสูญหายของข้อมูล เมื่อพิจารณาช่วงความเชื่อมั่นของสัมประสิทธิ์การถดถอย $\hat{\beta}_0, \hat{\beta}_1$ วิธี WNR ให้ช่วงความเชื่อมั่นดีกว่าวิธีอื่นๆสำหรับการสูญหาย 10%-20% ดังนั้นสำหรับข้อมูลที่สูญหาย 10%-20% จึงแนะนำให้ใช้วิธี WNR ในการประมาณค่าสูญหาย

ในการวิจัยครั้งนี้ผู้วิจัยต้องการศึกษาความแตกต่างสำหรับวิธีการที่ใช้ในการประมาณค่าสูญหาย ในการวิเคราะห์การถดถอยเชิงเส้น (Linear Regression) สำหรับข้อมูลที่แตกต่างกัน 2 ลักษณะ คือ 1) ข้อมูลภาคตัดขวาง (Cross – sectional Data) 2) ข้อมูลอนุกรมเวลา(Time – Series Data) เมื่อมีปัจจัยแนวโน้ม และ ปัจจัยฤดูกาลเข้ามาเกี่ยวข้อง ซึ่งการวิเคราะห์การถดถอยเชิงเส้นเป็นเทคนิคหนึ่งที่นำมาใช้ในการพยากรณ์กันอย่างแพร่หลาย โดยการใช้ค่าของตัวแปรอิสระ (Independent Variable) มาพยากรณ์ค่าของตัวแปรตาม (Dependent Variable) ซึ่งตัวแปรทั้งสองประเภทนี้มีความสัมพันธ์กันในลักษณะใดลักษณะหนึ่ง และใช้วิธีกำลังสองน้อยสุดในการหาสัมประสิทธิ์การถดถอยเพื่อหาสมการถดถอยเชิงเส้นในการพยากรณ์ ซึ่งวิธีการประมาณค่าสูญหายที่สนใจคือ

1. วิธีการประมาณค่าสูญหายแบบ Regression Imputation (RI)
2. วิธีการประมาณค่าสูญหายแบบ Nearest Neighbor Imputation (NNI)
3. วิธีการประมาณค่าสูญหายแบบ Weighted Nearest Neighbor and Regression Imputation (WNR)
4. วิธีการประมาณค่าสูญหายแบบด้วย EM algorithm (Expectation Maximization)

1.2 วัตถุประสงค์ของการวิจัย

เพื่อศึกษาและเปรียบเทียบวิธีการประมาณค่าสูญหายของตัวแปรตามในการวิเคราะห์การถดถอยเชิงเส้นสำหรับข้อมูลที่แตกต่างกัน 2 ลักษณะ คือ

1. ความถดถอยเชิงเส้นพหุ เมื่อเป็นข้อมูลภาคตัดขวาง (Cross – sectional Data)
2. ความถดถอยเชิงเส้นพหุ เมื่อเป็นข้อมูลอนุกรมเวลา(Time – Series Data) ที่มีปัจจัยแนวโน้ม และปัจจัยฤดูกาลเข้ามาเกี่ยวข้อง

1.3 ข้อตกลงเบื้องต้น

1. ตัวแปรที่สนใจศึกษา y และ x มีความสัมพันธ์ภายใต้การถดถอยเชิงเส้น(Linear Regression) โดยแบ่งเป็น 2 กรณี คือ

กรณีที่1 ความถดถอยเชิงเส้นพหุ เมื่อเป็นข้อมูลภาคตัดขวาง

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \qquad ; i = 1,2,\ldots,r,r+1,\ldots,n$$

เมื่อ y_i เป็นค่าสังเกตของตัวแปรตามของข้อมูลตัวที่ i หรือคาบเวลาที่ i

x_{1i} เป็นค่าสังเกตของตัวแปรอิสระตัวที่ 1ของข้อมูลตัวที่ i หรือคาบเวลาที่ i

x_{2i} เป็นค่าสังเกตของตัวแปรอิสระตัวที่ 2ของข้อมูลตัวที่ i หรือคาบเวลาที่ i

β_k เป็นสัมประสิทธิ์ความถดถอย $;k =$ 0 , 1 , 2

ε_i เป็นค่าความคลาดเคลื่อน

n เป็นจำนวนค่าสังเกตทั้งหมด

r เป็นจำนวนค่าสังเกตที่ทราบค่า

$n-r$ เป็นจำนวนค่าสังเกตที่สูญหาย

กรณีที่2 ความถดถอยเชิงเส้นพหุ เมื่อเป็นข้อมูลอนุกรมเวลา ที่ประกอบด้วยปัจจัยแนวโน้ม และปัจจัยฤดูกาล

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 t + \beta_3 Q_{1i} + \beta_4 Q_{2i} + \beta_5 Q_{3i} + \varepsilon_i \qquad ; i = 1,2,\ldots,r,r+1,\ldots,n$$

เมื่อ y_i เป็นค่าสังเกตของตัวแปรตามของข้อมูลตัวที่ i หรือคาบเวลาที่ i

x_i เป็นค่าสังเกตของตัวแปรอิสระของข้อมูลตัวที่ i หรือคาบเวลาที่ i

β_k เป็นสัมประสิทธิ์ความถดถอย $;k =$ 0 , 1 ,…,5

ε_i เป็นค่าความคลาดเคลื่อน

n เป็นจำนวนค่าสังเกตทั้งหมด

r เป็นจำนวนค่าสังเกตที่ทราบค่า

$n-r$ เป็นจำนวนค่าสังเกตที่สูญหาย

t เป็นค่าแนวโน้มของอนุกรมเวลา

Q_{li} เป็นตัวแปรบ่งชี้ของข้อมูลรายไตรมาส (มีทั้งหมด 4 ฤดูกาล) $l = 1,2,3$

2. ความคลาดเคลื่อนเป็นตัวแปรสุ่มที่มี $E(\varepsilon_i) = 0$, $Var(\varepsilon_i) = \sigma^2$ และสำหรับข้อมูลอนุกรมเวลา $Cov(\varepsilon_i, \varepsilon_j) = 0$ เมื่อ $i \neq j$
3. ความคลาดเคลื่อนมีการแจกแจงแบบปกติ ($\varepsilon_i \sim N(0,\sigma^2)$)
4. การสูญหายเกิดขึ้นกับตัวแปรตามเท่านั้น และเป็นการสูญหายแบบสุ่ม (Missing at Random (MAR))

1.4 ขอบเขตของการวิจัย

1. กำหนดการสูญหายของข้อมูลเกิดขึ้นกับตัวแปรตาม และเป็นการสูญหายแบบสุ่ม
2. ลักษณะของตัวแปรอิสระที่นำมาศึกษามีรูปแบบการแจกแจงปกติ (Normal Distribution)

   ฟังก์ชันการแจกแจงคือ

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad , -\infty < x < \infty$$

   ค่าคาดหวัง $E(X) = \mu$ ความแปรปรวน $V(X) = \sigma^2$

   โดยกำหนดให้ $X \sim N\ (60,100)$

3. ตัวแปรที่สนใจศึกษา y และ x มีความสัมพันธ์ด้วยรูปแบบการถดถอยเชิงเส้น(Linear Regression) โดยแบ่งเป็น 2 กรณีภายใต้ขอบเขตของการวิจัยแตกต่างกันดังนี้

กรณีที่1 ความถดถอยเชิงเส้นพหุ เมื่อเป็นข้อมูลภาคตัดขวาง

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \qquad ; i = 1,2,...,r,r+1,...,n$$

กำหนดค่าสหสัมพันธ์ (Correlation) ระหว่างตัวแปรดังนี้

1) ค่าสหสัมพันธ์ (Correlation) ระหว่าง y_i กับ x_{1i} และ y_i กับ x_{2i} เท่ากัน คือ 0.7

2) ค่าสหสัมพันธ์ (Correlation) ระหว่าง y_i กับ x_{1i} และ y_i กับ x_{2i} คือ 0.9 และ 0.4 ตามลำดับ

โดยกำหนดให้ตัวแปรอิสระไม่มีความสัมพันธ์กัน คือ กำหนดให้ค่าสหสัมพันธ์ (Correlation) มีค่าเท่ากับ 0

กรณีที่2 ความถดถอยเชิงเส้นพหุ เมื่อเป็นข้อมูลอนุกรมเวลา ที่ประกอบด้วยปัจจัยแนวโน้มและปัจจัยฤดูกาล

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 t + \beta_3 Q_{1i} + \beta_4 Q_{2i} + \beta_5 Q_{3i} + \varepsilon_i \qquad ; i = 1,2,...,r,r+1,...,n$$

$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ และ β_5 เป็นพารามิเตอร์ที่ไม่ทราบค่า โดยในงานวิจัยนี้กำหนดให้ค่าสัมประสิทธิ์การถดถอยแตกต่างกัน 3 ลักษณะ คือ

1) $\beta_0 = 0.5, \beta_1 = 1, \beta_2 = 0.3, \beta_3 = -0.6, \beta_4 = 0.6$ และ $\beta_5 = -0.8$
2) $\beta_0 = 0.5, \beta_1 = 1, \beta_2 = 0.8, \beta_3 = 0.1, \beta_4 = -0.1$ และ $\beta_5 = -0.2$
3) $\beta_0 = 0.5, \beta_1 = 1, \beta_2 = 0.5, \beta_3 = 0.5, \beta_4 = -0.1$ และ $\beta_5 = -0.3$

4. ค่าความคลาดเคลื่อนของข้อมูลมีการแจกแจงแบบปกติ $(\varepsilon_i \sim N(0,\sigma^2))$ กำหนดให้ $\sigma =$5 ,10 ,15 ,20 และ 25
5. ขนาดตัวอย่างที่ใช้ในการศึกษาเท่ากับ 50 , 100 และ 200
6. ข้อมูลตัวแปรตามที่สูญหายคิดเป็นร้อยละ 5 , 10 และ 20
7. การวิจัยครั้งนี้ทำการจำลองข้อมูลให้มีสถานการณ์ที่แตกต่างกันตามข้อกำหนดข้างต้น โดยใช้เทคนิคการจำลองแบบมอนติคาร์โล (Monte carlo Simulation Technique) ทำการจำลองในแต่ละสถานการณ์ 1,000 รอบ

1.5 เกณฑ์ที่ใช้ในการตัดสินใจ

เกณฑ์ที่ใช้ในการตัดสินใจว่าการประมาณค่าสูญหายด้วยวิธีใดใช้ได้ดีกว่าจากการเปรียบเทียบค่าความคลาดเคลื่อนระหว่างค่าพยากรณ์ของตัวแปรตามกับค่าจริง ในรูปแบบ Mean absolute percentage error (MAPE) ว่าวิธีการใดให้ค่า MAPE ต่ำกว่า แสดงว่าเป็นวิธีการประมาณที่ดีกว่า โดยการคำนวณจากสูตร

$$MAPE = \frac{\sum_{i=1}^{n}\left|\frac{(y_i - \hat{y}_i)}{y_i}\right|}{n} \times 100$$

เมื่อ y_i คือ ค่าจริงของข้อมูลตัวแปรตามตัวที่ i หรือคาบเวลาที่ i

$\hat{y}_i$ คือ ค่าประมาณของข้อมูลตัวแปรตามตัวที่ i หรือคาบเวลาที่ i

n คือ ขนาดตัวอย่าง

1.6 ประโยชน์ที่คาดว่าจะได้รับ

เพื่อเป็นแนวทางในการตัดสินใจเลือกวิธีการประมาณค่าสูญหายของตัวแปรตามในสมการถดถอยเชิงเส้นเพื่อใช้ในการพยากรณ์ และเป็นแนวทางในการศึกษา เปรียบเทียบวิธีการการประมาณค่าสูญหายในการวิเคราะห์ข้อมูลในรูปแบบอื่นๆต่อไป

บทที่ 2

แนวคิดและทฤษฎี

วิธีการประมาณค่าสูญหายของตัวแปรตาม ในการวิเคราะห์การถดถอยเชิงเส้น เมื่อใช้วิธีกำลังสองน้อยที่สุดในการสัมประสิทธิ์การถดถอย เพื่อหาสมการถดถอยในการพยากรณ์ ที่พิจารณามี 4 วิธีดังนี้

1. วิธีการประมาณค่าสูญหายแบบ Regression Imputation (RI)
2. วิธีการประมาณค่าสูญหายแบบ Nearest Neighbor Imputation (NNI)
3. วิธีการประมาณค่าสูญหายแบบ Weighted Nearest Neighbor and Regression Imputation (WNR)
4. วิธีการประมาณค่าสูญหายแบบด้วย EM algorithm (Expectation Maximization)

2.1 <u>**วิธีกำลังสองของความคลาดเคลื่อนน้อยที่สุดแบบทั่วไป**</u> (Ordinary Least Squares Method: OLS Method)

ในการประมาณค่าพารามิเตอร์ของตัวแบบความถดถอยเชิงเส้นนั้นมีหลายวิธี ซึ่งวิธีหนึ่งที่นิยมใช้นั้น คือ วิธีกำลังสองของความคลาดเคลื่อนน้อยที่สุดแบบทั่วไป (Ordinary Least Squares Method : OLS Method) นั่นคือ หาค่าสัมประสิทธิ์ความถดถอยที่ทำให้ผลบวกกำลังสองของความคลาดเคลื่อน(Sum Square of Errors : SSE) มีค่าน้อยที่สุด

กรณีที่1 ความถดถอยเชิงเส้นพหุ เมื่อเป็นข้อมูลภาคตัดขวางตัวแบบความถดถอยคือ

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \qquad ; i = 1,2,\ldots,r,r+1,\ldots,n$$

หรือ

$$\tilde{y} = X\tilde{\beta} + \tilde{\varepsilon}$$

$$\tilde{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ . \\ . \\ . \\ y_n \end{bmatrix} , \; X = \begin{bmatrix} 1 & x_{11} & x_{21} \\ 1 & x_{12} & x_{22} \\ \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} \end{bmatrix} , \tilde{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} , \; \tilde{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ . \\ . \\ . \\ \varepsilon_n \end{bmatrix}$$

กรณีที่2 ความถดถอยเชิงเส้นพหุ เมื่อเป็นข้อมูลอนุกรมเวลาที่มีปัจจัยแนวโน้มและปัจจัยฤดูกาลตัวแบบความถดถอยคือ

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 t + \beta_3 Q_{1i} + \beta_4 Q_{2i} + \beta_5 Q_{3i} + \varepsilon_i \qquad ; i = 1,2,...,r,r+1,...,n$$

หรือ

$$\tilde{y} = X\tilde{\beta} + \tilde{\varepsilon}$$

$$\tilde{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ . \\ . \\ . \\ y_n \end{bmatrix}, \; X = \begin{bmatrix} 1 & x_1 & 1 & 1 & 0 & 0 \\ 1 & x_2 & 2 & 0 & 1 & 0 \\ 1 & x_3 & 3 & 0 & 0 & 1 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 1 & x_n & n & 0 & 0 & 0 \end{bmatrix}, \tilde{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix}, \; \tilde{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ . \\ . \\ . \\ \varepsilon_n \end{bmatrix}$$

ด้วยวิธี OLS นั่นคือ หาค่าสัมประสิทธิ์ความถดถอยที่ทำให้ผลบวกกำลังสองของความคลาดเคลื่อน (Sum Square of Errors: SSE) มีค่าน้อยที่สุด

กำหนด ตัวประมาณ $\tilde{\beta}$ คือ $\tilde{b}$

จาก $SSE = \tilde{\varepsilon}^T \tilde{\varepsilon}$

$$= (\tilde{y} - X\tilde{b})^T (\tilde{y} - X\tilde{b})$$

$$= \tilde{y}^T \tilde{y} - \tilde{y}^T X\tilde{b} - \tilde{b}^T X^T \tilde{y} + \tilde{b}^T X^T X\tilde{b}$$

$$= \tilde{y}^T \tilde{y} - 2\tilde{b}^T X^T \tilde{y} + \tilde{b}^T X^T X\tilde{b}$$

นั่นคือ $\frac{\partial}{\partial \tilde{b}} (\tilde{y}^T \tilde{y} - 2\tilde{b}^T X^T \tilde{y} + \tilde{b}^T X^T X\tilde{b}) \overset{SET}{=} \tilde{0}$

$$-2X^T \tilde{y} + 2X^T X\tilde{b} = \tilde{0}$$

$$(X^T X)\tilde{b} = X^T \tilde{y}$$

$$\tilde{b} = (X^T X)^{-1} X^T \tilde{y}$$

การประมาณค่าความแปรปรวน

$$\hat{\sigma}^2 = s^2 = MSE = \frac{\sum_{i=1}^{n} e_i^2}{n-p-1} = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-p-1} = \frac{SSE}{n-p-1}$$

เรียก MSE ว่า ค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (Mean Square of Errors) เรียก SSE ว่า ค่ากำลังสองของความคลาดเคลื่อน (Sum Square of Errors) และค่า $n-p-1$ คือค่าองศาความเป็นอิสระของค่ากำลังสองของความคลาดเคลื่อน โดยที่ n คือ จำนวนข้อมูลค่าสังเกต และ $p+1$ คือ จำนวนพารามิเตอร์ในตัวแบบความถดถอยเชิงเส้นอย่างง่าย โดย $E(MSE)=\sigma^2$ นั่นคือ MSE เป็นตัวประมาณที่ไม่เอนเอียงของพารามิเตอร์ σ^2

2.2 **<u>วิธีการประมาณค่าสูญหายแบบ Regression Imputation</u>** (RI)

การประมาณค่าสูญหายแบบ Regression Imputation จัดเป็นเทคนิคขั้นพื้นฐานที่ใช้ในการประมาณค่าสูญหายโดยใช้การถดถอยจากชุดข้อมูล $(x_i, y_i)\ ; i=1,\ldots,r$ ที่ทราบค่า เมื่อตัวแปรตามส่วนหนึ่งมีค่าสูญหาย $(y_j)\ ; j=r+1,\ldots,n$ จะพยากรณ์ค่าสูญหายด้วยสมการถดถอยโดยคำนวณค่าสัมประสิทธิ์การถดถอยจากชุดข้อมูล (x_i, y_i) ที่เหลืออยู่ นั่นคือ

กรณีที่1 ความถดถอยเชิงเส้นพหุ เมื่อเป็นข้อมูลภาคตัดขวาง

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1j} + \hat{\beta}_2 x_{2j} \qquad ; j=r+1,\ldots,n$$

กรณีที่2 ความถดถอยเชิงเส้นพหุ เมื่อเป็นข้อมูลอนุกรมเวลา ที่ประกอบด้วยปัจจัยแนวโน้มและปัจจัยฤดูกาล

กำหนดให้ $$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_j + \hat{\beta}_2 t + \hat{\beta}_3 Q_{1j} + \hat{\beta}_4 Q_{2j} + \hat{\beta}_5 Q_{3j} \qquad ; j=r+1,\ldots,n$$

เมื่อ $\tilde{b}_{obs} = (\mathbf{X}_{obs}^T \mathbf{X}_{obs})^{-1} \mathbf{X}_{obs}^T \tilde{y}_{obs}$

2.3 **<u>วิธีการประมาณค่าสูญหายแบบ Nearest Neighbor Imputation</u>** (NNI)

Nearest Neighbor Imputation เป็นวิธีการหนึ่งที่จัดอยู่ในวิธี Hot Deck (Hot Deck Method) คือเป็นการแทนข้อมูลที่สูญหายด้วยข้อมูลที่ทราบค่าโดยวิธีนี้เป็นวิธีที่มีประสิทธิภาพมากกว่าวิธี Hot Deck อื่นๆ (เช่น Sequence Hot – decking , Random Hot – decking) (Chen & Shao ,2001) ซึ่งวิธี Nearest Neighbor Imputation จะประมาณค่าโดยใช้ค่าใกล้สุด โดยพิจารณาเลือกหน่วยตัวอย่างจากชุดข้อมูล (x_i, y_i) ที่ทราบค่าที่มีลักษณะคล้ายคลึงกับหน่วยตัวอย่างที่เกิดค่าสูญหายมากที่สุด จากนั้นแทนค่าข้อมูลสูญหายด้วยค่าของหน่วยตัวอย่างที่คล้ายคลึงกัน

กำหนดให้ $y_j^* \, ; j = r+1,...,n$ คือ ค่าที่ได้จากการประมาณค่าสูญหายแบบ NNI

เมื่อ $\left|x_i - x_j\right| = \min_{1 \le i \le r} \left|x_i - x_j\right|$ สำหรับ $i, 1 \le i \le r$ และ $j = r+1,...,n$

จะได้ว่า $y_j^* = y_i$

เมื่อเป็นความถดถอยเชิงเส้นพหุจะพิจารณาความคล้ายจากระยะทางยุคลิด (Euclidean Distance) ดังนี้

กำหนดให้ $D_{ij} = \sqrt{\sum_{k=1}^{2} \left(x_{ki} - x_{kj}\right)^2} = \min_{1 \le i \le r} \sqrt{\sum_{k=1}^{2} \left(x_{ki} - x_{kj}\right)^2}$

สำหรับ $i, 1 \le i \le r$ และ $j = r+1,...,n$

จะได้ว่า $y_j^* = y_i$

2.4 วิธีการประมาณค่าสูญหายแบบ Weighted Nearest Neighbor and Regression Imputation (WNR)

วิธี Weighted Nearest Neighbor and Regression Imputation เป็นวิธีการประมาณค่าสูญหายประเภท Composite Method คือเป็นการพัฒนาวิธีการประมาณค่าสูญหายด้วยการรวมแนวคิดของสองวิธีเข้าด้วยกันโดยรวมวิธีการประมาณค่าสูญหายแบบ Nearest Neighbor Imputation กับ Regression Imputation แล้วถ่วงน้ำหนัก (Chaimonkol and Suwattee ,2004)

กำหนดให้ $\hat{y}_j^*$ $; j = r+1,...,n$ เป็นค่าประมาณที่ได้จากวิธีการประมาณค่าสูญหายแบบ WNR

$$\hat{y}_j^* = w_j y_j^* + \left(1 - w_j\right)\hat{y}_j \quad \text{สำหรับ } j = r+1,...,n \text{ ที่ } 0 < w_j < 1$$

โดยที่ $w_j = \dfrac{V(\hat{y}_j)}{V(y_j^*) + V(\hat{y}_j)} \, ; j = r+1,...,n$ เป็นค่าที่ใช้ในการถ่วงน้ำหนักระหว่าง NNI และ RI

เมื่อความแปรปรวนของตัวประมาณ RI และ NNI ประมาณได้ดังนี้

- ความแปรปรวนของตัวประมาณ Regression Imputation

จาก $\tilde{y} = \mathbf{X}\tilde{b}$

$$Var(\tilde{y}) = Var(\mathbf{X}^T\tilde{b})$$
$$= \mathbf{X}^T Var(\tilde{b})\mathbf{X}$$
$$= \mathbf{X}^T(\sigma^2(\mathbf{X}^T\mathbf{X})^{-1})\mathbf{X}$$

- ความแปรปรวนของตัวประมาณ Nearest Neighbor Imputation

จาก y_j^* ซึ่งเป็นค่าประมาณที่ได้มาจากวิธี NNI โดยมีตัวแปรที่สนใจ y และ x ที่มีความสัมพันธ์ภายใต้รูปแบบ linear model ที่ $\varepsilon_j \sim N(0,\sigma^2)$ และ $Cov(\varepsilon_i,\varepsilon_j) = 0$ สำหรับ $i \neq j = 1,2,\ldots,n$

ดังนั้น $V(y_j^*) = \sigma^2$

2.5 <u>วิธีการประมาณค่าสูญหายโดยวิธี EM algorithm (Expectation Maximization)</u>

การประมาณค่าสูญหายโดยวิธี EM algorithm (Expectation Maximization) เป็นทางเลือกหนึ่ง ซึ่งเสนอโดยเด็มเสตอร์ ลายด์ และรูบิน (Dempster Laird and Rubin) ค่าที่ประมาณขึ้นเป็นค่าที่มาจากกระบวนการวนซ้ำเพื่อค้นหาค่าประมาณ Maximum Likelihood ของค่าพารามิเตอร์ EM algorithm นี้แบ่งได้เป็น 2 ขั้นคือ ขั้นหาค่าคาดหวัง E Step และขั้นหาค่ามากที่สุด M Step

Little and Rubin (1987) ได้เสนอวิธีการEMในการประมาณค่าสูญหายในการวิเคราะห์การถดถอยเชิงเส้น ซึ่งลักษณะที่สนใจคือประมาณค่าสูญหายของตัวแปรตามในการวิเคราะห์การถดถอยเชิงเส้น โดยมีขั้นตอนดังนี้

1. จัดข้อมูลให้อยู่ในรูป

$$\begin{bmatrix}\tilde{y}_1\\ \tilde{y}_2\end{bmatrix} = \begin{bmatrix}\mathbf{X}_1\\ \mathbf{X}_2\end{bmatrix}\tilde{\beta} + \tilde{\varepsilon}$$

เมื่อ

$\tilde{y}_1$ คือ เวกเตอร์ของ y ที่ไม่สูญหายขนาด $r \times 1$

$\tilde{y}_2$ คือ เวกเตอร์ของ y ที่สูญหายขนาด $(n-r) \times 1$

$\mathbf{X}_1$ คือ เมทริกซ์ตัวแปรอิสระ x ของข้อมูลที่ชุดตัวแปรตามไม่มีค่าสูญหายขนาด $r \times (p+1)$

$\mathbf{X}_2$ คือ เมทริกซ์ตัวแปรอิสระ x ของข้อมูลที่ชุดตัวแปรตามมีค่าสูญหายขนาด $(n-r) \times (p+1)$

2. ประมาณพารามิเตอร์ β_i จากชุดข้อมูลที่ไม่สูญหายโดยวิธี OLS ซึ่งถือเป็นพารามิเตอร์เริ่มต้น $\hat{\beta}^0$
3. ในรอบที่ 1 ประมาณค่าที่สูญหายของข้อมูลตัวแปรตามภายใต้เงื่อนไขของชุดข้อมูลที่ไม่สูญหายและพารามิเตอร์ตัวปัจจุบันขั้นตอนนี้เรียกว่า E Stepโดยในนี้ใช้พารามิเตอร์เริ่มต้น $\hat{\beta}^0$

$$E(y_i \mid \mathbf{X}, \tilde{y}_{obs}, \hat{\beta}^0) = \begin{cases} y_i & ; i = 1,2,\ldots,r \\ \mathbf{X}_2 \hat{\beta}^0 & ; i = r+1, r+2,\ldots,n \end{cases}$$

4. แทนค่าข้อมูลสูญหายจากค่าประมาณที่ได้ แล้วหาพารามิเตอร์ปัจจุบันใหม่ $\hat{\beta}^1$ โดยวิธี OLS ซึ่งในขั้นนี้เรียกว่า M Step
5. หาค่าสัมบูรณ์ของผลต่างระหว่างค่าสัมประสิทธิ์การถดถอยเริ่มต้น ($\hat{\beta}^0$) กับค่าสัมประสิทธิ์การถดถอยในรอบที่ 1($\hat{\beta}^1$)
6. เปรียบเทียบค่าในขั้นที่ 5 ถ้าไม่มากกว่า 0.0001 ค่าที่ได้ในรอบที่ 1 จะเป็นค่าประมาณค่าสูญหาย แต่ถ้ามากกว่า 0.0001 ให้ทำขั้นต่อไป
7. ขั้นตอนนี้เรียกว่า E Stepในการทำซ้ำรอบที่ t ; $t = 2,3$

$$E(y_i \mid \mathbf{X}, \tilde{y}_{obs}, \hat{\beta}^t) = \begin{cases} y_i & ; i = 1,2,\ldots,r \\ \mathbf{X}_2 \hat{\beta}^{(t)} & ; i = r+1, r+2,\ldots,n \end{cases}$$

8. แทนค่าข้อมูลสูญหายจากค่าประมาณที่ได้ แล้วหาพารามิเตอร์ปัจจุบันใหม่ $\hat{\beta}^t$ โดยวิธี OLS ในขั้น M Step
9. หาค่าสัมบูรณ์ของผลต่างระหว่างค่าสัมประสิทธิ์การถดถอยรอบที่ $t-1$ ($\hat{\beta}^{t-1}$) กับค่าสัมประสิทธิ์การถดถอยในรอบที่ t ($\hat{\beta}^t$)
10. เปรียบเทียบค่าในขั้นที่ 9 ถ้าไม่มากกว่า 0.0001 ค่าที่ได้ในรอบที่ $t-1$ จะเป็นค่าประมาณค่าสูญหาย แต่ถ้ามากกว่า 0.0001 ให้ทำขั้นที่ 7 ถึง ขั้นที่ 9 ทำซ้ำจนกระทั่งพารามิเตอร์ปัจจุบันคงที่(จนกว่าจะผ่านเงื่อนไข) จะได้ค่าประมาณค่าที่สูญหาย คือ $E(y_i \mid \mathbf{X}, \tilde{y}_{obs}, \hat{\beta}^*)$ ซึ่งตัวประมาณที่ได้เป็นตัวประมาณภาวะน่าจะเป็นสูงสุด

11. แทนค่าประมาณค่าสูญหายที่ได้แล้วทำการประมาณค่าสัมประสิทธ์การถดถอยโดยวิธี OLS จะได้สมการถดถอยใหม่มาใช้ในการพยากรณ์

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยในครั้งนี้เป็นการวิจัยเชิงทดลอง เพื่อเปรียบเทียบวิธีการประมาณค่าสูญหายของตัวแปรตามในสมการถดถอยเชิงเส้นพหุเพื่อการพยากรณ์ เมื่อการสูญหายของข้อมูลตัวแปรตามเป็นไปอย่างสุ่ม โดยทำการประมาณค่าสูญหายของตัวแปรตาม 4 วิธีดังนี้

1. วิธีการประมาณค่าสูญหายแบบ Regression Imputation (RI)
2. วิธีการประมาณค่าสูญหายแบบ Nearest Neighbor Imputation (NNI)
3. วิธีการประมาณค่าสูญหายแบบ Weighted Nearest Neighbor and Regression Imputation (WNR)
4. วิธีการประมาณค่าสูญหายแบบด้วย EM algorithm (Expectation Maximization)

การเปรียบเทียบจะเปรียบเทียบความคลาดเคลื่อนของแต่ละวิธีด้วยค่าความคลาดเคลื่อนระหว่างค่าพยากรณ์ของตัวแปรตามกับค่าจริง ในรูปแบบ Mean absolute percentage error (MAPE) ด้วยวิธีการต่างๆ ที่ขนาดตัวอย่าง 3 ระดับ ค่าความคลาดเคลื่อนของข้อมูล 5 ระดับ ร้อยละการสูญหายของข้อมูลตัวแปรตาม 3 ระดับ และลักษณะของข้อมูลที่แตกต่างกัน 2 กรณี โดยกรณีที่ 1 มี 2 รูปแบบ กรณีที่ 2 มี 3 รูปแบบ โดยที่ในแต่ละสถานการณ์มีการทำซ้ำ 1,000 ครั้ง

ในการวิจัยครั้งนี้ ผู้วิจัยใช้เทคนิคการจำลองแบบมอนติคาร์โล (Monte carlo Simulation Technique) ทำการจำลองในแต่ละสถานการณ์ ดังนั้นในส่วนแรกจะกล่าวถึงวิธีการจำลองโดยใช้เทคนิคการจำลองแบบมอนติคาร์โล และแสดงรายละเอียดของขั้นตอนการวิจัยในส่วนถัดไป

3.1 เทคนิคการจำลองแบบมอนติคาร์โล

เทคนิคที่ใช้แก้ปัญหาในการคำนวณทางสถิตินั้นมีอยู่หลายวิธี เทคนิคการจำลองแบบมอนติคาร์โลเป็นเทคนิคหนึ่งที่นิยมใช้ในการแก้ปัญหา ซึ่งหลักของการจำลองโดยใช้เทคนิคนี้ จะใช้เลขสุ่ม (Random Numbers) ในการหาคำตอบของปัญหาที่ต้องการศึกษา

ขั้นตอนการจำลองด้วยเทคนิคการจำลองแบบมอนติคาร์โลที่ใช้ในปัจจุบัน แบ่งได้เป็น 3 ขั้นตอน คือ

1. การสร้างตัวเลขสุ่ม การใช้ตัวเลขสุ่มเป็นส่วนสำคัญสำหรับเทคนิคนี้ เพราะหลักการการจำลองแบบมอนติคาร์โลนั้น จะใช้ตัวเลขสุ่มมาช่วยในการหาคำตอบของปัญหา โดยลักษณะของตัวเลขสุ่มที่นำมาใช้จะมีการแจกแจงแบบสม่ำเสมอในช่วง (0 , 1) ตัวเลขสุ่มแต่ละตัวเป็นอิสระต่อกันและมีช่วงยาวก่อนจะเกิดการสุ่มซ้ำ (มีวัฏจักรยาว)

2. การนำตัวเลขสุ่มมาประยุกต์ใช้กับปัญหาที่ต้องการศึกษา ซึ่งขั้นตอนนี้ขึ้นอยู่กับลักษณะของปัญหา บางปัญหาอาจไม่ใช้ตัวเลขสุ่มโดยตรง แต่จะนำไปสร้างเลขสุ่มที่มีการแจกแจงแบบอื่นๆต่อไป
3. การทดลองกระทำ เมื่อนำตัวเลขสุ่มมาประยุกต์ให้เข้ากับปัญหาที่ต้องการศึกษาได้แล้ว ขั้นต่อไปคือ การทดลองโดยใช้กระบวนการของการสุ่ม (Random Process) มากระทำในลักษณะซ้ำๆกันหลายๆครั้ง เพื่อหาคำตอบที่ต้องการ

3.2 แผนการดำเนินการวิจัย

การวิจัยครั้งนี้ต้องการประมาณค่าสูญหายของตัวแปรตาม เพื่อการพยากรณ์ในการวิเคราะห์การถดถอยเชิงเส้นพหุ เมื่อข้อมูลตัวแปรตามมีการสูญหายแบบสุ่มโดยทำการประมาณค่าสูญหายของตัวแปรตามทั้ง 4 วิธี ที่ขนาดตัวอย่าง 3 ระดับ ค่าความคลาดเคลื่อนของข้อมูล 5 ระดับ ร้อยละการสูญหายของข้อมูลตัวแปรตาม 3 ระดับ และลักษณะของข้อมูลที่แตกต่างกัน 2 กรณี และใช้วิธีกำลังสองน้อยที่สุดหาสัมประสิทธิ์การถดถอยเพื่อนำมาพยากรณ์ค่าสังเกตของตัวแปรตาม จากนั้นทำการเปรียบเทียบความคลาดเคลื่อนของแต่ละวิธีด้วยค่าความคลาดเคลื่อนระหว่างค่าพยากรณ์ของตัวแปรตามกับค่าจริง ในรูปแบบ Mean absolute percentage error (MAPE) ของค่าพยากรณ์ที่ได้จากการประมาณค่าสูญหายทั้ง 4 วิธี เพื่อหาวิธีที่เหมาะสมในแต่ละสถานการณ์

3.3 ขั้นตอนการดำเนินการวิจัย

ในการวิจัยครั้งนี้ ได้ทำการจำลองการทดลองตามสถานการณ์ต่างๆโดยการสร้างโปรแกรมคอมพิวเตอร์ ด้วยโปรแกรมแมทแล็บ (Matlab) โดยอาศัยเทคนิคมอนติคาร์โล เพื่อสร้างข้อมูลให้เป็นไปตามการวิจัยโดยจะกระทำซ้ำ ๆ กัน 1,000 ครั้ง ในแต่ละสถานการณ์ตามขั้นตอนดังนี้

3.3.1 การสร้างข้อมูล

ในการวิจัยครั้งนี้ จะสร้างข้อมูลตัวแปรที่สนใจศึกษา y และ x ให้มีความสัมพันธ์ภายใต้การถดถอยเชิงเส้นพหุ (Multiple Linear Regression) โดยแบ่งเป็น 2 กรณี คือ

กรณีที่1 ความถดถอยเชิงเส้นพหุ เมื่อเป็นข้อมูลภาคตัดขวาง

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \qquad ; i = 1,2,\ldots,r,r+1,\ldots,n$$

เมื่อ y_i เป็นค่าสังเกตของตัวแปรตามของข้อมูลตัวที่ i หรือคาบเวลาที่ i

x_{1i} เป็นค่าสังเกตของตัวแปรอิสระตัวที่ 1 ของข้อมูลตัวที่ i หรือคาบเวลาที่ i

x_{2i} เป็นค่าสังเกตของตัวแปรอิสระตัวที่ 2 ของข้อมูลตัวที่ i หรือคาบเวลาที่ i

β_k เป็นสัมประสิทธิ์ความถดถอย ; $k = 0, 1, 2$

ε_i เป็นค่าความคลาดเคลื่อน

n เป็นจำนวนค่าสังเกตทั้งหมด

r เป็นจำนวนค่าสังเกตที่ทราบค่า

$n-r$ เป็นจำนวนค่าสังเกตที่สูญหาย

จำลองข้อมูลตัวแปรอิสระ x_{1i} , x_{2i} และตัวแปรตาม y_i ตามค่าสหสัมพันธ์และรูปแบบการแจกแจงที่กำหนด โดยใช้หลักการจากวิธี โชเลสกี (Cholesky) ในการสร้างข้อมูลแจกแจงแบบปกติที่ไม่เป็นอิสระต่อกัน โดยกำหนดค่าสหสัมพันธ์ (Correlation) ระหว่างตัวแปรดังนี้

3) ค่าสหสัมพันธ์ (Correlation) ระหว่าง y_i กับ x_{1i} และ y_i กับ x_{2i} เท่ากัน คือ 0.7

4) ค่าสหสัมพันธ์ (Correlation) ระหว่าง y_i กับ x_{1i} และ y_i กับ x_{2i} คือ 0.9 และ 0.4

โดยจำลองข้อมูลให้มีสถานการณ์ที่ค่าความคลาดเคลื่อน ขนาดตัวอย่างที่ใช้แตกต่างกัน

กรณีที่2 ความถดถอยเชิงเส้นพหุ เมื่อเป็นข้อมูลอนุกรมเวลา ที่ประกอบด้วยปัจจัยแนวโน้มและปัจจัยฤดูกาล

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 t + \beta_3 Q_{1i} + \beta_4 Q_{2i} + \beta_5 Q_{3i} + \varepsilon_i \qquad ; i = 1,2,...,r,r+1,...,n$$

เมื่อ y_i เป็นค่าสังเกตของตัวแปรตามของข้อมูลตัวที่ i หรือคาบเวลาที่ i

x_i เป็นค่าสังเกตของตัวแปรอิสระของข้อมูลตัวที่ i หรือคาบเวลาที่ i

β_k เป็นสัมประสิทธิ์ความถดถอย ; $k = 0, 1, \ldots, 5$

ε_i เป็นค่าความคลาดเคลื่อน

n เป็นจำนวนค่าสังเกตทั้งหมด

r เป็นจำนวนค่าสังเกตที่ทราบค่า

$n-r$ เป็นจำนวนค่าสังเกตที่สูญหาย

t เป็นค่าแนวโน้มของอนุกรมเวลา

Q_{li} เป็นตัวแปรบ่งชี้ของข้อมูลรายไตรมาส (มีทั้งหมด 4 ฤดูกาล) $l = 1,2,3$

ในกรณีที่ 2 ขั้นตอนในการสร้างข้อมูลมีรายละเอียดดังนี้

1. จำลองข้อมูลตัวแปรอิสระ x_i ให้มีการแจกแจงแบบปกติ โดยสร้างจากฟังก์ชั่น Normal
2. สร้างค่าความคลาดเคลื่อน ε_i จากการแจกแจงแบบปกติมีค่าเฉลี่ยเป็น 0 และค่าเบี่ยงเบนมาตรฐานตามที่กำหนดโดยสร้างจากฟังก์ชั่น เช่นเดียวกันกับ ตัวแปรอิสระ x_i
3. สร้างข้อมูลตัวแปรตาม y_i ตามรูปแบบความสัมพันธ์ในการวิเคราะห์การถดถอยเชิงเส้น (Linear Regression) ดังที่แสดงไว้ตามค่าพารามิเตอร์ β_k ที่กำหนดให้แตกต่างกัน 3 ลักษณะ คือ
   1) $\beta_0 = 0.5, \beta_1 = 1, \beta_2 = 0.3, \beta_3 = -0.6, \beta_4 = 0.6$ และ $\beta_5 = -0.8$
   2) $\beta_0 = 0.5, \beta_1 = 1, \beta_2 = 0.8, \beta_3 = 0.1, \beta_4 = -0.1$ และ $\beta_5 = -0.2$
   3) $\beta_0 = 0.5, \beta_1 = 1, \beta_2 = 0.5, \beta_3 = 0.5, \beta_4 = -0.1$ และ $\beta_5 = -0.3$

โดยจำลองข้อมูลให้มีสถานการณ์ที่ค่าความคลาดเคลื่อน ขนาดตัวอย่างที่ใช้แตกต่างกัน

3.3.2 สุ่มตำแหน่งการสูญหายของข้อมูล

ทำการสุ่มตำแหน่งที่สูญหายตามร้อยละการสูญหายที่กำหนดโดยใช้การวนซ้ำในการหาต่ำแหน่งการสูญหาย

3.3.3 ประมาณค่าสูญหายด้วยวิธีการทั้ง 4 วิธี คือ

1. วิธีการประมาณค่าสูญหายแบบ Regression Imputation (RI) เป็นวิธีการประมาณค่าสูญหายจากสมการถดถอยเชิงเส้นพหุของข้อมูลที่ไม่สูญหาย
2. วิธีการประมาณค่าสูญหายแบบ Nearest Neighbor Imputation (NNI) เป็นวิธีการประมาณค่าโดยพิจารณาเลือกหน่วยตัวอย่างจากชุดข้อมูล (x_i, y_i) ที่ทราบค่าที่มีลักษณะคล้ายคลึงกับหน่วยตัวอย่างที่เกิดค่าสูญหายมากที่สุด จากนั้นแทนค่าข้อมูลสูญหายด้วยค่าของหน่วยตัวอย่างที่คล้ายคลึงกัน
3. วิธีการประมาณค่าสูญหายแบบ Weighted Nearest Neighbor and Regression Imputation (WNR) เป็นวิธีการประมาณค่าสูญหายที่รวมวิธีการประมาณค่าสูญหายแบบ Nearest Neighbor Imputation กับ Regression Imputation เข้าด้วยกันแล้วถ่วงน้ำหนัก (Chaimonkol and Suwattee ,2004)
4. วิธีการประมาณค่าสูญหายแบบด้วย EM algorithm (Expectation Maximization) มีขั้นตอนดังนี้

1) ประมาณพารามิเตอร์ β_i จากชุดข้อมูลที่ไม่สูญหาย ซึ่งถือเป็นพารามิเตอร์เริ่มต้น $\hat{\beta}^0$

2) ประมาณข้อมูลสูญหายด้วยค่าคาดหวังของค่าที่สูญหายภายใต้เงื่อนไขของชุดข้อมูลที่ไม่สูญหายและพารามิเตอร์ตัวปัจจุบันซึ่งในขั้นตอนนี้เรียกว่า E Stepโดยในรอบแรกใช้พารามิเตอร์เริ่มต้น

$$E(y_i \mid \mathbf{X}, \tilde{y}_{obs}, \hat{\beta}^t) = \begin{cases} y_i & ; i = 1,2,\dots,r \\ \mathbf{X}_2\hat{\beta}^{(t)} & ; i = r+1, r+2,\dots,n \end{cases}$$

เมื่อข้อมูลอยู่ในรูป

$$\begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \tilde{\beta} + \tilde{\varepsilon}$$

เมื่อ

$\tilde{y}_1$ คือ เวกเตอร์ของ y ที่ไม่สูญหายขนาด $r \times 1$

$\tilde{y}_2$ คือ เวกเตอร์ของ y ที่สูญหายขนาด $(n-r) \times 1$

$\mathbf{X}_1$ คือ เมทริกซ์ตัวแปรอิสระ x ของข้อมูลที่ชุดตัวแปรตามไม่มีค่าสูญหายขนาด $r \times (p+1)$

$\mathbf{X}_2$ คือ เมทริกซ์ตัวแปรอิสระ x ของข้อมูลที่ชุดตัวแปรตามมีค่าสูญหายขนาด $(n-r) \times (p+1)$

3) แทนค่าข้อมูลสูญหายจากค่าประมาณที่ได้ แล้วหาพารามิเตอร์ปัจจุบันใหม่ $\hat{\beta}^t$ โดยวิธี OLS ซึ่งในขั้นนี้เรียกว่า M Step

4) ทำซ้ำจนกระทั่งพารามิเตอร์ปัจจุบันคงที่จะได้ค่าประมาณค่าที่สูญหาย คือ $E(y_i \mid \mathbf{X}, \tilde{y}_{obs}, \hat{\beta}^*)$ ซึ่งตัวประมาณที่ได้เป็นตัวประมาณภาวะน่าจะเป็นสูงสุด

3.3.4 ประมาณพารามิเตอร์ใหม่

จากวิธีการประมาณค่าสูญหาย เมื่อแทนค่าข้อมูลสูญหายจากค่าประมาณที่ได้แล้วจะทำการประมาณค่าสัมประสิทธิ์การถดถอยใหม่ด้วยวิธีกำลังสองน้อยสุด (OLS) เพื่อหาสมการถดถอยเชิงเส้นพหุในการพยากรณ์จากวิธีการประมาณค่าสูญหายทั้ง 4 วิธี และกรณีที่ไม่มีข้อมูลสูญหาย

3.3.5 หาค่าความคลาดเคลื่อนระหว่างค่าพยากรณ์ของตัวแปรตามกับค่าจริง

หาค่าความคลาดเคลื่อนระหว่างค่าพยากรณ์ของตัวแปรตามกับค่าจริง ในรูปแบบ Mean absolute percentage error (MAPE) โดยมีขั้นตอนดังนี้

1) ประมาณค่าพยากรณ์ $\hat{y}$ ด้วยพารามิเตอร์ใหม่ที่ประมาณโดยสมการถดถอยจากวิธีการประมาณค่าสูญหายทั้ง 4 วิธี และกรณีที่ไม่มีข้อมูลสูญหาย

2) นำมาเปรียบเทียบกับ y ที่สร้างไว้ในตอนต้น โดยใช้ Mean absolute percentage error (MAPE) โดยมีสูตรการคำนวณดังนี้

$$MAPE_l = \frac{\sum_{i=1}^{n}\left|\frac{(y_i - \hat{y}_i)}{y_i}\right|}{n} \times 100$$

เมื่อ y_i คือ ค่าจริงของข้อมูลตัวแปรตามตัวที่ i หรือคาบเวลาที่ i

$\hat{y}_i$ คือ ค่าประมาณของข้อมูลตัวแปรตามตัวที่ i หรือคาบเวลาที่ i

n คือ ขนาดตัวอย่าง

3) หาค่า Mean absolute percentage error (MAPE) เฉลี่ยจากการจำลองข้อมูลตามสถานการณ์ต่างๆ โดยใช้เทคนิคมอลติคาร์โลซิมมูเลชั่น 1,000 รอบ

$$MAPE = \frac{\sum_{l=1}^{1000} MAPE_l}{1000}$$

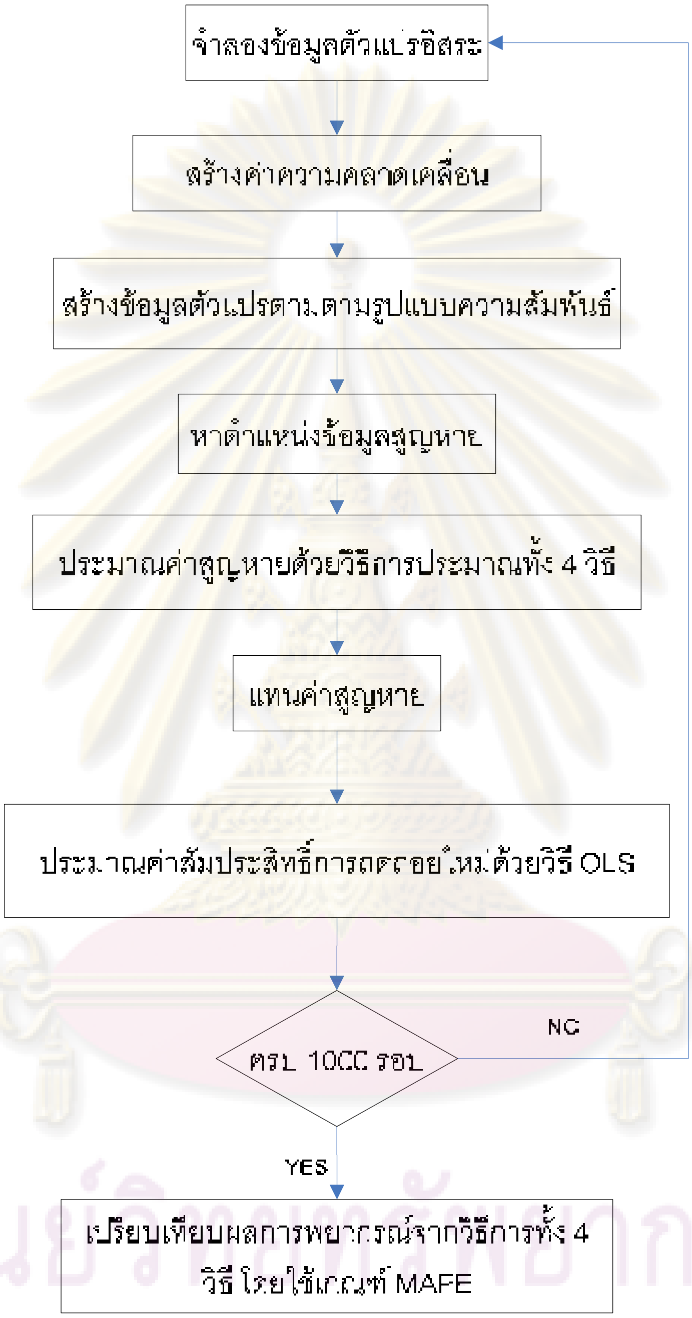
4) เปรียบเทียบค่า Mean absolute percentage error (MAPE) จากการประมาณค่าสูญหายแต่ละวิธี และสรุปผลการเปรียบเทียบ

โดยการทำตามขั้นตอนต่างๆจะเปลี่ยน ค่าความคลาดเคลื่อนของข้อมูล 5 ระดับ ที่ขนาดตัวอย่าง 3 ระดับ และร้อยละการสูญหายของข้อมูลตัวแปรตาม 3 ระดับ จากลักษณะของข้อมูลที่แตกต่างกัน 2 กรณี โดยที่ในแต่ละสถานการณ์มีการทำซ้ำ 1000 รอบ จนครบทุกสถานการณ์ ซึ่งขั้นตอนการวิจัยดังกล่าวนี้ สามารถสรุปเป็นผังงานได้ดังรูป

กรณีที่ 1


จำลองข้อมูลตัวแปรอิสระและข้อมูลตัวแปรตาม ตามรูปแบบความสัมพันธ์
หาตำแหน่งข้อมูลสูญหาย
ประมาณค่าสูญหายด้วยวิธีการประมาณทั้ง 4 วิธี
แทนค่าสูญหาย
ประมาณค่าสัมประสิทธิ์การถดถอยใหม่ด้วยวิธี OLS
ครบ 1000 รอบ
NO
YES
เปรียบเทียบผลการพยากรณ์จากวิธีการทั้ง 4 วิธี โดยใช้เกณฑ์ MAFE

กรณีที่ 2


จำลองข้อมูลตัวแปรอิสระ
สร้างค่าความคลาดเคลื่อน
สร้างข้อมูลตัวแปรตามตามรูปแบบความสัมพันธ์
หาตำแหน่งข้อมูลสูญหาย
ประมาณค่าสูญหายด้วยวิธีการประมาณทั้ง 4 วิธี
แทนค่าสูญหาย
ประมาณค่าสัมประสิทธิ์การถดถอยใหม่ด้วยวิธี OLS
ครบ 1000 รอบ
NO
YES
เปรียบเทียบผลการพยากรณ์จากวิธีการทั้ง 4 วิธี โดยใช้เกณฑ์ MAPE

บทที่ 4

ผลการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบวิธีการประมาณค่าสูญหายของตัวแปรตามในสมการถดถอยเชิงเส้นพหุเพื่อการพยากรณ์ เมื่อการสูญหายของข้อมูลตัวแปรตามเป็นไปอย่างสุ่ม โดยทำการประมาณค่าสูญหายของตัวแปรตาม 4 วิธี ซึ่งได้แก่ วิธีการประมาณค่าสูญหายแบบ Regression Imputation (RI) วิธีการประมาณค่าสูญหายแบบ Nearest Neighbor Imputation (NNI) วิธีการประมาณค่าสูญหายแบบ Weighted Nearest Neighbor and Regression Imputation (WNR) และวิธีการประมาณค่าสูญหายแบบด้วย EM algorithm (Expectation Maximization) โดยใช้เกณฑ์เปรียบเทียบความคลาดเคลื่อนของแต่ละวิธีด้วยค่าความคลาดเคลื่อนระหว่างค่าพยากรณ์ของตัวแปรตามกับค่าจริง ในรูปแบบ Mean absolute percentage error (MAPE)

การเสนอผลการวิเคราะห์ข้อมูลในบทนี้ เป็นการแสดงผลการวิเคราะห์ตามขั้นตอนการวิเคราะห์ข้อมูลที่ได้กล่าวไว้ในบทที่ 3 โดยผู้วิจัยได้เสนอผลการวิจัยโดยแบ่งออกเป็น 2 ส่วน คือ

ส่วนที่ 1 ผลการเปรียบวิธีประมาณค่าสูญหายของตัวแปรตามในสมการถดถอยเชิงเส้นพหุกรณีที่เป็นข้อมูลภาคตัดขวาง

ส่วนที่ 2 ผลการเปรียบวิธีประมาณค่าสูญหายของตัวแปรตามในสมการถดถอยเชิงเส้นพหุกรณีที่เป็นข้อมูลอนุกรมเวลา ที่ประกอบด้วยปัจจัยแนวโน้มและปัจจัยฤดูกาล

สำหรับการนำเสนอผลการวิจัยจะนำเสนอในรูปแบบตารางเพื่อความสะดวกในการอธิบายจึงใช้สัญลักษณ์ต่อไปนี้เพื่อแทนความหมายต่างๆ

σ	หมายถึง	ส่วนเบี่ยงเบนมาตรฐานของค่าความคลาดเคลื่อนของข้อมูล
n	หมายถึง	ขนาดตัวอย่าง
pm	หมายถึง	ร้อยละของการสูญหายของข้อมูล
MAPE	หมายถึง	ค่าความคลาดเคลื่อนระหว่างค่าพยากรณ์ของตัวแปรตามกับค่าจริงในรูปแบบ Mean absolute percentage error
Complete	หมายถึง	กรณีไม่มีข้อมูลสูญหาย
RI	หมายถึง	วิธี Regression Imputation
NNI	หมายถึง	วิธี Nearest Neighbor Imputation

WNR	หมายถึง	วิธีWeighted Nearest Neighbor and Regression Imputation
EM	หมายถึง	วิธี EM algorithm (Expectation Maximization)
*	หมายถึง	ค่าต่ำที่สุดของ MAPE ในการเปรียบเทียบ

4.1 การเปรียบวิธีประมาณค่าสูญหายของตัวแปรตามในสมการถดถอยเชิงเส้นพหุกรณีที่เป็นข้อมูลภาคตัดขวาง

การวิจัยครั้งนี้ผู้วิจัยทำการศึกษาเมื่อความคลาดเคลื่อนมีการแจกแจงแบบปกติมีพารามิเตอร์ $\mu = 0$ และ $\sigma = 5$, 10 , 15 , 20 และ 25 ตามลำดับ

โดยสร้างข้อมูลตัวแปรอิสระและตัวแปรตามที่มีรูปแบบการแจกแจงปกติ (Normal Distribution) และ $\beta_0, \beta_1, \beta_2$ เป็นสัมประสิทธิ์ความถดถอย ซึ่งเป็นพารามิเตอร์ที่ไม่ทราบค่า เมื่อกำหนดค่าสหสัมพันธ์ (Correlation) ระหว่างตัวแปรดังนี้

1) ค่าสหสัมพันธ์ (Correlation) ระหว่าง ค่าสังเกตของตัวแปรตาม(y_i)กับค่าสังเกตของตัวแปรอิสระตัวที่ 1(x_{1i})และค่าสังเกตของตัวแปรตาม(y_i)กับค่าสังเกตของตัวแปรอิสระตัวที่ 2 (x_{2i}) เท่ากัน คือ 0.7 ซึ่งผลการวิจัยส่วนนี้ได้นำเสนอในตารางที่ 4.1-4.2 โดยแสดงค่า MAPE เมื่อขนาดตัวอย่างเท่ากับ 50 , 100 และ 200 ที่ร้อยละของการสูญหายเป็น 5 , 10 และ 15 ตามลำดับ

2) ค่าสหสัมพันธ์ (Correlation) ระหว่าง ค่าสังเกตของตัวแปรตาม(y_i)กับค่าสังเกตของตัวแปรอิสระตัวที่ 1(x_{1i})และค่าสังเกตของตัวแปรตาม(y_i)กับค่าสังเกตของตัวแปรอิสระตัวที่ 2 (x_{2i})คือ 0.9 และ 0.4 ซึ่งผลการวิจัยส่วนนี้ได้นำเสนอในตารางที่ 4.3-4.4 โดยแสดงค่า MAPE เมื่อขนาดตัวอย่างเท่ากับ 50 , 100 และ 200 ที่ร้อยละของการสูญหายเป็น 5 , 10 และ 15 ตามลำดับ

ตารางที่ 4.1 แสดงการเปรียบเทียบวิธีการประมาณค่าสูญหาย โดยพิจารณาจาก ค่า MAPE สำหรับข้อมูลภาคตัดขวางเมื่อค่าสหสัมพันธ์ (Correlation) ระหว่าง ค่าสังเกตของตัวแปรตาม (y_i) กับค่าสังเกตของตัวแปรอิสระตัวที่ 1(x_{1i}) และ ตัวที่ 2 (x_{2i}) เท่ากับ 0.7 และ 0.7

σ	n	pm	MAPE				
			Complete	RI	NNI	WNR	EM
5	50	5	0.9409	0.9420*	0.9508	0.9422	0.9420*
		10	0.9329	0.9362*	0.9672	0.9365	0.9362*
		20	0.9368	0.9447*	1.0184	0.9452	0.9447*
	100	5	0.9335	0.9292*	0.9331	0.9292*	0.9292*
		10	0.9321	0.9335*	0.9436	0.9335*	0.9335*
		20	0.9301	0.9359*	0.9630	0.9360	0.9359*
	200	5	0.9199	0.9203*	0.9220	0.9203*	0.9203*
		10	0.9190	0.9199*	0.9235	0.9199*	0.9199*
		20	0.9208	0.9225*	0.9330	0.9225*	0.9225*
10	50	5	1.9220	1.9247*	1.9511	1.9252	1.9247*
		10	1.9202	1.9270*	1.9888	1.9276	1.9270*
		20	1.9155	1.9308*	2.0772	1.9324	1.9308*
	100	5	1.9107	1.9118*	1.9217	1.9118*	1.9118*
		10	1.9039	1.9075*	1.9309	1.9076	1.9075*
		20	1.9101	1.9176*	1.9788	1.9177	1.9176*
	200	5	1.8952	1.8959*	1.9003	1.8959*	1.8959*
		10	1.8865	1.8882*	1.8958	1.8882*	1.8882*
		20	1.8928	1.8965*	1.9178	1.8965*	1.8965*
15	50	5	3.0497	3.0573*	3.1060	3.0585	3.0573*
		10	3.0361	3.0466*	3.1301	3.0468	3.0466*
		20	3.0853	3.1069*	3.3708	3.1115	3.1069*
	100	5	3.0081	3.0111*	3.0308	3.0111*	3.0111*
		10	3.0109	3.0161*	3.0529	3.0165	3.0161*
		20	3.0191	3.0330*	3.1356	3.0339	3.0330*
	200	5	2.9696	2.9681*	2.9743	2.9681*	2.9681*
		10	2.9511	2.9533*	2.9725	2.9533*	2.9533*
		20	2.9884	2.9945*	3.0143	2.9945*	2.9945*

ตารางที่ 4.1(ต่อ) แสดงการเปรียบเทียบวิธีการประมาณค่าสูญหาย โดยพิจารณาจาก ค่า MAPE สำหรับข้อมูลภาคตัดขวางเมื่อค่าสหสัมพันธ์ (Correlation) ระหว่าง ค่าสังเกตของตัวแปรตาม (y_i) กับค่าสังเกตของตัวแปรอิสระตัวที่ 1(x_{1i}) และ ตัวที่ 2 (x_{2i}) เท่ากับ 0.7 และ 0.7

σ	n	pm	MAPE				
			Complete	RI	NNI	WNR	EM
20	50	5	5.7181	5.7251*	5.8353	5.7275	5.7251*
		10	5.6675	5.7140*	5.9862	5.7180	5.7140*
		20	5.5888	5.6619*	6.2511	6.6753	5.6619*
	100	5	4.8702	4.8694	4.8814	4.8686*	4.8694
		10	4.9172	4.9439*	5.0643	4.9464	4.9439*
		20	4.8833	4.9247*	5.2122	4.9298	4.9247*
	200	5	4.7509	4.7567	4.7927	4.7560*	4.7567
		10	4.7971	4.8034*	4.8637	4.8034*	4.8034*
		20	4.8067	4.8161*	5.0399	4.8161*	4.8161*
25	50	5	8.8367	8.8341*	9.1463	8.8414	8.8341*
		10	8.7314	8.7192*	9.2396	8.7278	8.7192*
		20	8.6884	8.7441*	9.7417	8.7452	8.7441*
	100	5	8.2197	8.2076	8.2208	8.2039*	8.2076
		10	8.3297	8.3407	8.5550	8.3369*	8.3407
		20	8.2794	8.3560	8.8760	8.3554*	8.3560
	200	5	8.1537	8.1487	8.2015	8.1482*	8.1487
		10	8.2007	8.2168	8.2624	8.2151*	8.2168
		20	8.1886	8.2086*	8.6287	8.2086*	8.2086*

หมายเหตุ

* หมายถึง วิธีการประมาณค่าสูญหายที่มีค่า MAPE ต่ำที่สุด

ตารางที่ 4.2 ผลการวิเคราะห์การทดสอบความแตกต่างของค่าเฉลี่ยของค่า MAPE ของวิธีการประมาณค่าสูญหายทั้ง 4 วิธี เปรียบเทียบกับกรณีที่ไม่มีข้อมูลสูญหายเมื่อค่าสหสัมพันธ์ระหว่างค่าสังเกตของตัวแปรตาม (y_i) กับค่าสังเกตของตัวแปรอิสระตัวที่ 1(x_{1i}) และ ตัวที่ 2 (x_{2i}) เท่ากับ 0.7 และ 0.7

ส่วนเบี่ยงเบนมาตรฐาน	ขนาดตัวอย่าง	ร้อยละการสูญหาย	วิธีการประมาณค่าสูญหายที่ไม่แตกต่าง จากกรณีที่ไม่มีข้อมูลสูญหาย
5	50	5	RI NNI WNR EM
		10	RI WNR EM
		20	RI WNR EM
	100	5	RI NNI WNR EM
		10	RI WNR EM
		20	RI WNR EM
	200	5	RI NNI WNR EM
		10	RI NNI WNR EM
		20	RI WNR EM
10	50	5	RI NNI WNR EM
		10	RI WNR EM
		20	RI WNR EM
	100	5	RI NNI WNR EM
		10	RI WNR EM
		20	RI WNR EM
	200	5	RI NNI WNR EM
		10	RI NNI WNR EM
		20	RI WNR EM

ตารางที่ 4.2 (ต่อ) ผลการวิเคราะห์การทดสอบความแตกต่างของค่าเฉลี่ยของค่า MAPE ของวิธีการประมาณค่าสูญหายทั้ง 4 วิธี เปรียบเทียบกับกรณีที่ไม่มีข้อมูลสูญหายเมื่อค่าสหสัมพันธ์ (Correlation) ระหว่าง ค่าสังเกตของตัวแปรตาม (y_i) กับค่าสังเกตของตัวแปรอิสระตัวที่ 1(x_{1i}) และ ตัวที่ 2 (x_{2i}) เท่ากับ 0.7 และ 0.7

ส่วนเบี่ยงเบนมาตรฐาน	**ขนาดตัวอย่าง**	**ร้อยละการสูญหาย**	**วิธีการประมาณค่าสูญหายที่ไม่แตกต่าง จากกรณีที่ไม่มีข้อมูลสูญหาย**
15	50	5	RI NNI WNR EM
		10	RI WNR EM
		20	RI WNR EM
	100	5	RI NNI WNR EM
		10	RI WNR EM
		20	RI WNR EM
	200	5	RI NNI WNR EM
		10	RI NNI WNR EM
		20	RI NNI WNR EM
20	50	5	RI NNI WNR EM
		10	RI WNR EM
		20	RI WNR EM
	100	5	RI NNI WNR EM
		10	RI NNI WNR EM
		20	RI NNI WNR EM
	200	5	RI NNI WNR EM
		10	RI NNI WNR EM
		20	RI NNI WNR EM

ตารางที่ 4.2 (ต่อ) ผลการวิเคราะห์การทดสอบความแตกต่างของค่าเฉลี่ยของค่า MAPE ของวิธีการประมาณค่าสูญหายทั้ง 4 วิธี เปรียบเทียบกับกรณีที่ไม่มีข้อมูลสูญหายเมื่อค่าสหสัมพันธ์ (Correlation) ระหว่าง ค่าสังเกตของตัวแปรตาม (y_i) กับค่าสังเกตของตัวแปรอิสระตัวที่ 1(x_{1i}) และ ตัวที่ 2 (x_{2i}) เท่ากับ 0.7 และ 0.7

ส่วนเบี่ยงเบนมาตรฐาน	ขนาดตัวอย่าง	ร้อยละการสูญหาย	วิธีการประมาณค่าสูญหายที่ไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย
25	50	5	RI NNI WNR EM
		10	RI NNI WNR EM
		20	RI WNR EM
	100	5	RI NNI WNR EM
		10	RI NNI WNR EM
		20	RI NNI WNR EM
	200	5	RI NNI WNR EM
		10	RI NNI WNR EM
		20	RI NNI WNR EM

หมายเหตุ

RI หมายถึง จากการทดสอบความแตกต่างของค่าเฉลี่ย MAPE วิธี Regression Imputation ให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย

NNI หมายถึง จากการทดสอบความแตกต่างของค่าเฉลี่ย MAPE วิธี Nearest Neighbor Imputationให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย

WNR หมายถึง จากการทดสอบความแตกต่างของค่าเฉลี่ย MAPE วิธี Weighted Nearest Neighbor and Regression Imputation ให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย

EM หมายถึง จากการทดสอบความแตกต่างของค่าเฉลี่ย MAPE วิธี EM algorithm (Expectation Maximization) ให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย

เมื่อพิจารณาเปรียบเทียบค่า MAPE กรณีที่ค่าสหสัมพันธ์ (Correlation) ระหว่าง ค่าสังเกตของตัวแปรตาม (y_i) กับค่าสังเกตของตัวแปรอิสระตัวที่ 1(x_{1i}) และ ตัวที่ 2 (x_{2i}) เท่ากันคือ 0.7 จากวิธีการประมาณค่าสูญหายทั้ง 4 วิธีและกรณีที่ไม่มีข้อมูลสูญหายเปรียบเทียบกัน พบว่าค่า MAPE จะเพิ่มขึ้นเมื่อข้อมูลมีความแปรปรวนเพิ่มขึ้น แต่มีค่าลดลงเมื่อขนาดตัวอย่างใหญ่ขึ้น และมีผลต่างกับกรณีที่ไม่มีข้อมูลสูญหาย (ค่าจริง) มากขึ้นเมื่อร้อยละการสูญหายของข้อมูลเพิ่มขึ้น

เมื่อส่วนเบี่ยงมาตรฐานอยู่ในระดับต่ำ ($\sigma =$ 5) และขนาดตัวอย่างมีขนาดเล็ก (n = 50) วิธีที่มีค่า MAPE ต่ำสุด 2 วิธี คือ RI และ EM เมื่อขนาดตัวอย่างเพิ่มขึ้นเป็นขนาดปานกลางและขนาดใหญ่ (n = 100 และ 200) วิธีที่มีค่า MAPE ต่ำสุดมี 3 วิธี คือ RI, WNR และ EM

จากการทดสอบความแตกต่างของค่าเฉลี่ยค่า MAPE พบว่า เมื่อตัวอย่างมีขนาดเล็กและปานกลาง (n = 50 และ 100) วิธีการประมาณค่าสูญหายทั้ง 4 วิธีให้ผลไม่แตกต่างกันเมื่อร้อยละของการสูญหายอยู่ในระดับต่ำ(5) และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย และเมื่อร้อยละของการสูญหายเพิ่มขึ้นเป็นระดับปานกลางและสูง (10 และ 20) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ให้ผลไม่แตกต่างกันและไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ที่ระดับนัยสำคัญ 0.05 เมื่อตัวอย่างมีขนาดใหญ่ (n = 200) วิธีการประมาณสูญหายทั้ง 4 วิธีให้ผลไม่แตกต่างกันเมื่อร้อยละการสูญหายอยู่ในระดับต่ำและปานกลาง (5 และ 10) และเมื่อร้อยละของการสูญหายเพิ่มขึ้นสูง (20) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ที่ระดับนัยสำคัญ 0.05

เมื่อส่วนเบี่ยงมาตรฐานอยู่ในระดับที่ค่อนข้างต่ำและระดับปานกลาง ($\sigma =$ 10 และ15) ที่ตัวอย่างมีขนาดเล็ก (n = 50) วิธีการประมาณค่าสูญหายที่มีค่า MAPE ต่ำสุด 2 วิธีคือ วิธี RI และ EM เมื่อตัวอย่างมีขนาดปานกลาง (n = 100) ที่ร้อยละการสูญหายของข้อมูลอยู่ในระดับต่ำ (5) 3 วิธี ที่ค่า MAPE ต่ำสุด คือ วิธี RI, WNR และ EM เมื่อร้อยละการสูญหายของข้อมูลเพิ่มขึ้นเป็นระดับปานกลางและระดับสูง (10 และ 20) วิธีที่มีค่า MAPE ต่ำสุดมี 2 วิธี คือ วิธี RI และ EM เมื่อขนาดตัวอย่างเพิ่มขึ้นเป็นขนาดใหญ่ (n = 200) วิธีที่มีค่า MAPE ต่ำสุดเท่ากัน 3 วิธี คือ RI, WNR และ EM

จากการทดสอบความแตกต่างของค่าเฉลี่ยค่า MAPE พบว่า เมื่อตัวอย่างมีขนาดเล็กและปานกลาง (n = 50 และ 100) ที่ร้อยละของการสูญหายอยู่ในระดับต่ำ(5) วิธีการประมาณค่าสูญหายทั้ง 4 วิธีให้ผลไม่แตกต่างกัน และเมื่อร้อยละของการสูญหายเพิ่มขึ้นเป็นระดับปานกลางและสูง (10 และ 20) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ที่ให้ผลไม่แตกต่างกัน และไม่แตกต่างจาก

กรณีที่ไม่มีข้อมูลสูญหาย ที่ระดับนัยสำคัญ 0.05 เมื่อตัวอย่างมีขนาดใหญ่ (n = 200) ที่ส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับที่ค่อนข้างต่ำ (10) วิธีการประมาณสูญหายทั้ง 4 วิธีให้ผลไม่แตกต่างกันเมื่อร้อยละการสูญหายอยู่ในระดับต่ำและปานกลาง (5 และ 10) และเมื่อร้อยละของการสูญหายอยู่ในระดับสูง (20) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ที่ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ที่ระดับนัยสำคัญ 0.05 และที่ส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับปานกลาง (15) วิธีการประมาณค่าสูญหายทั้ง 4 วิธี ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ทุกระดับร้อยละการสูญหาย

เมื่อส่วนเบี่ยงมาตรฐานอยู่ในระดับที่ค่อนข้างสูง (σ = 20) และตัวอย่างมีขนาดเล็ก (n = 50) วิธีที่มีค่า MAPE ต่ำสุดมี 2 วิธี คือ วิธี RI และ EM เมื่อตัวอย่างมีขนาดปานกลาง (n =100) ที่ร้อยละการสูญหายของข้อมูลอยู่ในระดับต่ำ (5) วิธีที่มีค่า MAPE ต่ำสุด คือวิธี WNR กรณีที่ร้อยละการสูญหายของข้อมูลเพิ่มขึ้นเป็นระดับปานกลางและระดับสูง (10 และ 20) วิธีที่มีค่า MAPE ต่ำสุดมี 2 วิธี คือ วิธี RI และ EM เมื่อตัวอย่างมีขนาดใหญ่ (n = 200) ที่ร้อยละการสูญหายของข้อมูลอยู่ในระดับต่ำ (5) วิธี WNR มีค่า MAPE ต่ำสุด เมื่อร้อยละการสูญหายของข้อมูลเพิ่มขึ้นเป็น 10 และ 20 พบว่ามี 3 วิธีที่มีค่า MAPE ต่ำสุดคือ วิธี RI, WNR และ EM

จากการทดสอบความแตกต่างของค่าเฉลี่ยค่า MAPE พบว่า เมื่อตัวอย่างมีขนาดเล็ก (n = 50) วิธีการประมาณค่าสูญหายทั้ง 4 วิธีให้ผลไม่แตกต่างกันและไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย เมื่อร้อยละของการสูญหายอยู่ในระดับต่ำ (5) และเมื่อร้อยละของการสูญหายเพิ่มขึ้นเป็นระดับปานกลางและสูง (10 และ 20) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ที่ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหายที่ระดับนัยสำคัญ 0.05 เมื่อตัวอย่างมีขนาดปานกลางและขนาดใหญ่ (n = 100 และ 200) วิธีการประมาณค่าสูญหายทั้ง 4 วิธีให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ที่ระดับนัยสำคัญ 0.05 ทุกระดับร้อยละการสูญหาย

เมื่อส่วนเบี่ยงมาตรฐานอยู่ในระดับสูง(σ = 25) และขนาดตัวอย่างมีขนาดเล็ก (n = 50) วิธีที่มีค่า MAPE ต่ำสุดมี 2 วิธี คือ RI และ EM เมื่อขนาดตัวอย่างเพิ่มขึ้นเป็นขนาดปานกลางและขนาดใหญ่ (n = 100 และ 200) วิธีที่มีค่า MAPE ต่ำสุดมีคือวิธีWNR

จากการทดสอบความแตกต่างของค่าเฉลี่ยค่า MAPE พบว่า วิธีการประมาณค่าสูญหายทั้ง 4 วิธีให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ยกเว้นเมื่อตัวอย่างมีขนาดเล็ก (n = 50) ที่ร้อยละของการสูญหายอยู่ในระดับสูง (20) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ที่ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ที่ระดับนัยสำคัญ 0.05

ข้อมูลภาคตัดขวางที่มีค่าสหสัมพันธ์ (Correlation) ระหว่าง ค่าสังเกตของตัวแปรตาม (y_i) กับค่าสังเกตของตัวแปรอิสระตัวที่ 1(x_{1i}) และ ตัวที่ 2 (x_{2i}) เท่ากัน คือ 0.7 ซึ่งจัดอยู่ในระดับสูง เมื่อส่วนเบี่ยงเบนมาตรฐานเพิ่มขึ้นการกระจายของข้อมูลเพิ่มขึ้น ทำให้การแทนค่าสูญหายด้วยวิธี WNR คือ แทนค่าสูญหายจากวิธี NNI ร่วมกับการใช้สมการพยากรณ์จากวิธี RI ให้ผลดีกว่าวิธีอื่นๆ เนื่องจากการพิจารณาหน่วยตัวอย่างที่คล้ายคลึงกันของตัวแปรอิสระเกิดความสอดคล้องทำให้มีความแม่นยำในการประมาณค่าสูญหายเพิ่มขึ้น เมื่อสมการถดถอยคลาดเคลื่อนมากขึ้น

ตารางที่ 4.3 แสดงการเปรียบเทียบวิธีการประมาณค่าสูญหาย โดยพิจารณาจาก ค่า MAPE สำหรับข้อมูลภาคตัดขวางเมื่อค่าสหสัมพันธ์ (Correlation) ระหว่าง ค่าสังเกตของตัวแปรตาม (y_i) กับค่าสังเกตของตัวแปรอิสระตัวที่ 1(x_{1i}) และ ตัวที่ 2 (x_{2i}) เท่ากับ 0.9 และ 0.4

σ	n	pm	MAPE				
			Complete	RI	NNI	WNR	EM
5	50	5	1.1504	1.1516*	1.1608	1.1518	1.1516*
		10	1.1477	1.1516*	1.1765	1.1519	1.1516*
		20	1.1515	1.1605*	1.2193	1.1612	1.1605*
	100	5	1.1419	1.1428*	1.1470	1.1428*	1.1428*
		10	1.1454	1.1471*	1.1568	1.1471*	1.1471*
		20	1.1434	1.1480*	1.1725	1.1482	1.1480*
	200	5	1.1315	1.1319*	1.1337	1.1319*	1.1319*
		10	1.1263	1.1272*	1.1303	1.1272*	1.1272*
		20	1.1311	1.1334*	1.1415	1.1334*	1.1334*
10	50	5	2.3661	2.3692*	2.3915	2.3695	2.3692*
		10	2.3675	2.3760*	2.4312	2.3765	2.3760*
		20	2.3520	2.3698*	2.5119	2.3708	2.3698*
	100	5	2.3406	2.3428*	2.3523	2.3429	2.3428*
		10	2.3372	2.3414*	2.3614	2.3415	2.3414*
		20	2.3309	2.3408*	2.3941	2.3409	2.3408*
	200	5	2.3068	2.3079*	2.3111	2.3079*	2.3079*
		10	2.3058	2.3081*	2.3154	2.3081*	2.3081*
		20	2.3063	2.3107*	2.3297	2.3108	2.3107*
15	50	5	3.7384	3.7436*	3.7881	3.7458	3.7436*
		10	3.7455	3.7639*	3.8898	3.7673	3.7639*
		20	3.7203	3.7513*	3.9713	3.7539	3.7513*
	100	5	3.6769	3.6800*	3.6950	3.6803	3.6800*
		10	3.6911	3.6990*	3.7395	3.6994	3.6990*
		20	3.6892	3.7067*	3.8164	3.7079	3.7067*
	200	5	3.6599	3.6609*	3.6717	3.6610	3.6609*
		10	3.6381	3.6410*	3.6613	3.6411	3.6410*
		20	3.6196	3.6262*	3.6626	3.6263	3.6262*

ตารางที่ 4.3 (ต่อ) แสดงการเปรียบเทียบวิธีการประมาณค่าสูญหาย โดยพิจารณาจาก ค่า MAPE สำหรับข้อมูลภาคตัดขวางเมื่อค่าสหสัมพันธ์ (Correlation) ระหว่าง ค่าสังเกตของตัวแปรตาม (y_i) กับค่าสังเกตของตัวแปรอิสระตัวที่ 1(x_{1i}) และ ตัวที่ 2 (x_{2i}) เท่ากับ 0.9 และ 0.4

σ	n	pm	MAPE				
			Complete	RI	NNI	WNR	EM
20	50	5	5.8880	5.8978*	5.9903	5.9033	5.8978*
		10	5.8965	5.9097*	6.2624	5.9162	5.9097*
		20	5.9195	5.9484*	6.5772	5.9736	5.9484*
	100	5	5.8592	5.8688*	5.8987	5.8696	5.8688*
		10	5.8678	5.8838*	5.9561	5.8839	5.8838*
		20	5.9000	5.9347*	6.2499	5.9389	5.9347*
	200	5	5.7894	5.7904*	5.8313	5.7907	5.7904*
		10	5.8440	5.8600*	5.9349	5.8611	5.8600*
		20	5.7847	5.7867*	5.9380	5.7881	5.7867*
25	50	5	9.4079	9.4668*	9.6720	9.4715	9.4668*
		10	9.8266	9.9339*	10.2028	9.9340	9.9339*
		20	9.6419	9.8262*	11.0470	9.8692	9.8262*
	100	5	9.1805	9.2622*	9.2981	9.2622*	9.2622*
		10	9.5221	9.6357*	9.9113	9.6409	9.6357*
		20	9.0902	9.2588*	9.6318	9.2715	9.2588*
	200	5	9.0553	9.0420*	9.0645	9.0422	9.0420*
		10	9.1754	9.2060*	9.3424	9.2067	9.2060*
		20	9.2688	9.3058*	9.5747	9.3072	9.3058*

หมายเหตุ

*   หมายถึง วิธีการประมาณค่าสูญหายที่มีค่า MAPE ต่ำที่สุด

ตารางที่ 4.4 ผลการวิเคราะห์การทดสอบความแตกต่างของค่าเฉลี่ยของค่า MAPE ของวิธีการประมาณค่าสูญหายทั้ง 4 วิธี เปรียบเทียบกับกรณีที่ไม่มีข้อมูลสูญหายเมื่อค่าสหสัมพันธ์ (Correlation) ระหว่าง ค่าสังเกตของตัวแปรตาม (y_i) กับค่าสังเกตของตัวแปรอิสระตัวที่ 1(x_{1i}) และ ตัวที่ 2 (x_{2i}) เท่ากับ 0.9 และ 0.4

ส่วนเบี่ยงเบนมาตรฐาน	ขนาดตัวอย่าง	ร้อยละการสูญหาย	วิธีการประมาณค่าสูญหายที่ไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย
5	50	5	RI NNI WNR EM
		10	RI WNR EM
		20	RI WNR EM
	100	5	RI NNI WNR EM
		10	RI WNR EM
		20	RI WNR EM
	200	5	RI NNI WNR EM
		10	RI NNI WNR EM
		20	RI WNR EM
10	50	5	RI NNI WNR EM
		10	RI WNR EM
		20	RI WNR EM
	100	5	RI NNI WNR EM
		10	RI WNR EM
		20	RI WNR EM
	200	5	RI NNI WNR EM
		10	RI NNI WNR EM
		20	RI WNR EM

ตารางที่ 4.4 (ต่อ) ผลการวิเคราะห์การทดสอบความแตกต่างของค่าเฉลี่ยของค่า MAPE ของวิธีการประมาณค่าสูญหายทั้ง 4 วิธี เปรียบเทียบกับกรณีที่ไม่มีข้อมูลสูญหายเมื่อค่าสหสัมพันธ์ (Correlation) ระหว่าง ค่าสังเกตของตัวแปรตาม (y_i) กับค่าสังเกตของตัวแปรอิสระตัวที่ 1(x_{1i}) และ ตัวที่ 2 (x_{2i}) เท่ากับ 0.9 และ 0.4

ส่วนเบี่ยงเบนมาตรฐาน	ขนาดตัวอย่าง	ร้อยละการสูญหาย	วิธีการประมาณค่าสูญหายที่ไม่แตกต่าง จากกรณีที่ไม่มีข้อมูลสูญหาย
15	50	5	RI NNI WNR EM
		10	RI WNR EM
		20	RI WNR EM
	100	5	RI NNI WNR EM
		10	RI WNR EM
		20	RI WNR EM
	200	5	RI NNI WNR EM
		10	RI NNI WNR EM
		20	RI WNR EM
20	50	5	RI NNI WNR EM
		10	RI WNR EM
		20	RI WNR EM
	100	5	RI NNI WNR EM
		10	RI NNI WNR EM
		20	RI NNI WNR EM
	200	5	RI NNI WNR EM
		10	RI NNI WNR EM
		20	RI NNI WNR EM

ตารางที่ 4.4 (ต่อ) ผลการวิเคราะห์การทดสอบความแตกต่างของค่าเฉลี่ยของค่า MAPE ของวิธีการประมาณค่าสูญหายทั้ง 4 วิธี เปรียบเทียบกับกรณีที่ไม่มีข้อมูลสูญหายเมื่อค่าสหสัมพันธ์ (Correlation) ระหว่าง ค่าสังเกตของตัวแปรตาม (y_i) กับค่าสังเกตของตัวแปรอิสระตัวที่ 1(x_{1i}) และ ตัวที่ 2 (x_{2i}) เท่ากับ 0.9 และ 0.4

ส่วนเบี่ยงเบนมาตรฐาน	ขนาดตัวอย่าง	ร้อยละการสูญหาย	วิธีการประมาณค่าสูญหายที่ไม่แตกต่าง จากกรณีที่ไม่มีข้อมูลสูญหาย
25	50	5	RI NNI WNR EM
		10	RI NNI WNR EM
		20	RI WNR EM
	100	5	RI NNI WNR EM
		10	RI NNI WNR EM
		20	RI NNI WNR EM
	200	5	RI NNI WNR EM
		10	RI NNI WNR EM
		20	RI NNI WNR EM

หมายเหตุ

RI หมายถึง จากการทดสอบความแตกต่างของค่าเฉลี่ย MAPE วิธี Regression Imputation ให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย

NNI หมายถึง จากการทดสอบความแตกต่างของค่าเฉลี่ย MAPE วิธี Nearest Neighbor Imputationให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย

WNR หมายถึง จากการทดสอบความแตกต่างของค่าเฉลี่ย MAPE วิธี Weighted Nearest Neighbor and Regression Imputation ให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย

EM หมายถึง จากการทดสอบความแตกต่างของค่าเฉลี่ย MAPE วิธี EM algorithm (Expectation Maximization) ให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย

เมื่อพิจารณาเปรียบเทียบค่า MAPE กรณีที่ค่าสหสัมพันธ์ (Correlation) ระหว่าง ค่าสังเกตของตัวแปรตาม (y_i) กับค่าสังเกตของตัวแปรอิสระตัวที่ 1(x_{1i}) และ ตัวที่ 2 (x_{2i}) เท่ากับ 0.9 และ 0.4 ตามลำดับ จากวิธีการประมาณค่าสูญหายทั้ง 4 วิธีและกรณีที่ไม่มีข้อมูลสูญหายเปรียบเทียบกัน พบว่าค่า MAPE มีค่าเพิ่มขึ้นเมื่อข้อมูลมีความแปรปรวนเพิ่มขึ้น แต่มีค่าลดลงเมื่อขนาดตัวอย่างใหญ่ และมีผลต่างกับกรณีที่ไม่มีข้อมูลสูญหาย (ค่าจริง) มากขึ้นเมื่อร้อยละการสูญหายของข้อมูลเพิ่มขึ้น

เมื่อส่วนเบี่ยงมาตรฐานอยู่ในระดับต่ำ ($\sigma =$ 5) และขนาดตัวอย่างมีขนาดเล็ก (n = 50) วิธีที่มีค่า MAPE ต่ำสุดมี 2 วิธี คือ RI และ EM เมื่อขนาดตัวอย่างเพิ่มขึ้นเป็นขนาดปานกลาง (n = 100) ที่ร้อยละการสูญหายอยู่ในระดับต่ำและปานกลาง (5 และ 10) วิธีที่มีค่า MAPE ต่ำสุดมี 3 วิธี คือ RI, WNR และ EM และเมื่อร้อยละการสูญหายอยู่ในระดับสูง (20) วิธีที่มีค่า MAPE ต่ำสุดมี 2 วิธี คือ RI และ EM และเมื่อตัวอย่างมีขนาดใหญ่ (n = 200) วิธีที่มีค่า MAPE ต่ำสุดมี 3 วิธี คือ RI, WNR และ EM ทุกระดับร้อยละของการสูญหาย

เมื่อส่วนเบี่ยงมาตรฐานอยู่ในระดับที่ค่อนข้างต่ำ ($\sigma =$ 10) เมื่อขนาดตัวอย่างมีขนาดเล็กและปานกลาง (n = 50 และ 100) วิธีที่มีค่า MAPE ต่ำสุดมี 2 วิธี คือ RI และ EM และที่ตัวอย่างมีขนาดปานกลาง วิธี RI และ EM มีค่า MAPE ใกล้เคียงกับวิธี WNR อย่างเห็นได้ชัด เมื่อตัวอย่างมีขนาดใหญ่ (200) วิธี RI และ EM มีค่า MAPE ใกล้เคียงกับวิธี WNR เช่นกัน และที่ร้อยละการสูญหายของข้อมูลอยู่ในระดับต่ำและปานกลาง (5 และ10) วิธีที่มีค่า MAPE ต่ำสุดมี 3 วิธี คือ RI, WNR และ EM เมื่อร้อยละการสูญหายของข้อมูลอยู่ในระดับสูง (20) วิธีที่มีค่า MAPE ต่ำสุดมี 2 วิธี คือ RI และ EM

เมื่อส่วนเบี่ยงมาตรฐานอยู่ในระดับปานกลาง ($\sigma =$ 15) ทุกระดับของขนาดตัวอย่างและร้อยละการสูญหายของข้อมูลตัวแปรตามวิธีที่มีค่า MAPE ต่ำสุดมี 2 วิธี คือ RI และ EM

จากการทดสอบความแตกต่างของค่าเฉลี่ยค่า MAPE สำหรับกรณีที่ส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับต่ำ ค่อนข้างต่ำ และปานกลาง ($\sigma =$ 5 , 10 และ15) ให้ผลตรงกันคือ เมื่อตัวอย่างมีขนาดเล็กและปานกลาง (n = 50 และ 100) ที่ร้อยละของการสูญหายอยู่ในระดับต่ำ (5) วิธีการประมาณค่าสูญหายทั้ง 4 วิธีให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย และเมื่อร้อยละของการสูญหายเพิ่มขึ้นเป็นระดับปานกลางและระดับสูง (10 และ 20) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ที่ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ที่ระดับนัยสำคัญ 0.05 เมื่อตัวอย่างมีขนาดใหญ่ (n = 200) ที่ร้อยละการสูญหายอยู่ในระดับ

ต่ำและปานกลาง (5 และ 10)วิธีการประมาณค่าสูญหายทั้ง 4 วิธีให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย เมื่อร้อยละการสูญหายอยู่ในระดับสูง (20) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ที่ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ที่ระดับนัยสำคัญ 0.05

เมื่อส่วนเบี่ยงมาตรฐานอยู่ในระดับที่ค่อนข้างสูง ($\sigma =$ 20) ทุกระดับของขนาดตัวอย่าง และร้อยละการสูญหายของข้อมูลตัวแปรตามวิธีที่มีค่า MAPE ต่ำสุดมี 2 วิธี คือ RI และ EM

จากการทดสอบความแตกต่างของค่าเฉลี่ยค่า MAPE เมื่อตัวอย่างมีขนาดเล็ก (n = 50) ที่ร้อยละของการสูญหายอยู่ในระดับต่ำ (5) วิธีการประมาณค่าสูญหายทั้ง 4 วิธีให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย และเมื่อร้อยละของการสูญหายเพิ่มขึ้นเป็นระดับปานกลางและระดับสูง (10 และ 20) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ที่ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ที่ระดับนัยสำคัญ 0.05 เมื่อตัวอย่างมีขนาดปานกลางและขนาดใหญ่ (n = 100 และ 200) วิธีการประมาณค่าสูญหายทั้ง 4 วิธีให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ที่ระดับนัยสำคัญ 0.05 ทุกระดับร้อยละของการสูญหาย

เมื่อส่วนเบี่ยงมาตรฐานอยู่ในระดับสูง($\sigma =$ 25) วิธีที่มีค่า MAPE ต่ำสุดมี 2 วิธี คือ วิธีRI และ EM ยกเว้นเมื่อขนาดตัวอย่างมีขนาดปานกลาง (n = 100) กรณีที่ร้อยละของการสูญหายอยู่ในระดับต่ำ (5) วิธี WNR มีค่า MAPE ต่ำสุดเท่ากันกับวิธี RI และ EM

จากการทดสอบความแตกต่างของค่าเฉลี่ยค่า MAPE วิธีการประมาณค่าสูญหายทั้ง 4 วิธีให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ยกเว้นเมื่อร้อยละของการสูญหายเพิ่มขึ้นเป็นระดับสูง (20) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ที่ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ที่ระดับนัยสำคัญ 0.05

ข้อมูลภาคตัดขวางที่มีค่าสหสัมพันธ์ (Correlation) ระหว่าง ค่าสังเกตของตัวแปรตาม (y_i) กับค่าสังเกตของตัวแปรอิสระตัวที่ 1(x_{1i}) และ ตัวที่ 2 (x_{2i}) คือ 0.9 และ 0.4 ตามลำดับ เนื่องจากในกรณีนี้ค่าสหสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระตัวหนึ่งสูงอีกตัวหนึ่งปานกลาง ส่งผลให้วิธี NNI มีประสิทธิภาพในการประมาณค่าสูญหายลดลง การใช้สมการถดถดถอยในการประมาณค่าสูญหายเพียงวิธีเดียวจึงดีกว่าการใช้วิธี WNR แม้ว่าส่วนเบี่ยงเบนมาตรฐานจะสูง

4.2 การเปรียบวิธีประมาณค่าสูญหายของตัวแปรตามในสมการถดถอยเชิงเส้นพหุกรณีที่เป็นข้อมูลอนุกรมเวลา ที่ประกอบด้วยปัจจัยแนวโน้มและปัจจัยฤดูกาล

การวิจัยครั้งนี้ผู้วิจัยทำการศึกษาเมื่อความคลาดเคลื่อนมีการแจกแจงแบบปกติมีพารามิเตอร์ $\mu = 0$ และ $\sigma = 5, 10, 15, 20$ และ 25 ตามลำดับ

โดยได้ทำการศึกษาในกรณีต่างๆดังนี้

4.2.1 กรณีมีอิทธิพลจากปัจจัยแนวโน้มต่ำแต่ปัจจัยฤดูกาลสูงคือ กรณีที่กำหนดให้สัมประสิทธิ์การถดถอยเป็นดังนี้ $\beta_0 = 0.5, \beta_1 = 1, \beta_2 = 0.3, \beta_3 = -0.6, \beta_4 = 0.6$ และ $\beta_5 = -0.8$ ซึ่งผลการวิจัยส่วนนี้ได้นำเสนอในตารางที่4.5-4.6 โดยแสดงค่า MAPE เมื่อขนาดตัวอย่างเท่ากับ 50 , 100 และ 200 ที่ร้อยละของการสูญหายเป็น 5 , 10 และ 15 ตามลำดับ

4.2.2 กรณีมีอิทธิพลจากปัจจัยแนวโน้มสูงแต่ปัจจัยฤดูกาลต่ำ คือ กรณีที่กำหนดให้สัมประสิทธิ์การถดถอยเป็นดังนี้ $\beta_0 = 0.5, \beta_1 = 1, \beta_2 = 0.8, \beta_3 = 0.1, \beta_4 = -0.1$ และ $\beta_5 = -0.2$ ซึ่งผลการวิจัยส่วนนี้ได้นำเสนอในตารางที่4.7-4.8 โดยแสดงค่า MAPE เมื่อขนาดตัวอย่างเท่ากับ 50 , 100 และ 200 ที่ร้อยละของการสูญหายเป็น 5 , 10 และ 15 ตามลำดับ

4.2.3 กรณีมีอิทธิพลจากปัจจัยแนวโน้มและปัจจัยฤดูกาลระดับปานกลาง คือ กรณีที่กำหนดให้สัมประสิทธิ์การถดถอยเป็นดังนี้ $\beta_0 = 0.5, \beta_1 = 1, \beta_2 = 0.5, \beta_3 = 0.5, \beta_4 = -0.1$ และ $\beta_5 = -0.3$ ซึ่งผลการวิจัยส่วนนี้ได้นำเสนอในตารางที่4.9-4.10 โดยแสดงค่า MAPE เมื่อขนาดตัวอย่างเท่ากับ 50 , 100 และ 200 ที่ร้อยละของการสูญหายเป็น 5 , 10 และ 15 ตามลำดับ

ตารางที่ 4.5 แสดงการเปรียบเทียบวิธีการประมาณค่าสูญหาย โดยพิจารณาจากค่า MAPE สำหรับข้อมูลอนุกรมเวลากรณีมีอิทธิพลจากปัจจัยแนวโน้มต่ำแต่ปัจจัยฤดูกาลสูง ($\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 0.3$, $\beta_3 = -0.6$, $\beta_4 = 0.6$ และ $\beta_5 = -0.8$)

σ	n	pm	MAPE				
			Complete	RI	NNI	WNR	EM
5	50	5	5.8517	5.8661	5.9125	5.8649*	5.8661
		10	5.6870	5.7433	5.9503	5.7384*	5.7433
		20	5.7074	5.8190	6.8286	5.8042*	5.8190
	100	5	5.2925	5.2993	5.4383	5.2988*	5.2993
		10	5.2938	5.3123*	5.7364	5.3124	5.3123*
		20	5.3918	5.4438*	7.0498	5.4504	5.4438*
	200	5	4.5874	4.5897*	5.1608	4.5903	4.5897*
		10	4.5807	4.5871*	6.4567	4.5899	4.5871*
		20	4.6146	4.6323*	10.4796	4.6584	4.6323*
10	50	5	11.5355	11.5767	11.5903	11.5708*	11.5767
		10	11.8903	11.9593	12.0726	11.9463*	11.9593
		20	11.7649	11.9980	12.2608	11.9499*	11.9980
	100	5	10.8442	10.8575	10.8931	10.8549*	10.8575
		10	10.9542	10.9831	11.2091	10.9729*	10.9831
		20	11.0082	11.0984	11.7400	11.0790*	11.0984
	200	5	9.4932	9.4979	9.9326	9.4966*	9.4979
		10	9.5214	9.5327	10.7860	9.5296*	9.5327
		20	9.3482	9.3833*	13.2218	9.3844	9.3833*
15	50	5	20.0014	20.0070	20.0862	20.0017*	20.0070
		10	20.3791	20.4788	20.5686	20.4625*	20.4788
		20	21.1033	21.5277	21.5906	21.3810*	21.5277
	100	5	17.7553	17.7812	17.7771	17.7706*	17.7812
		10	18.0268	18.0820	18.2074	18.0598*	18.0820
		20	17.9567	18.5188	19.1362	18.4661*	18.5188
	200	5	15.0909	15.0950	15.1739	15.0945*	15.0950
		10	15.3198	15.3315	15.8372	15.3281*	15.3315
		20	15.4024	15.4336	17.6803	15.4259*	15.4336

ตารางที่ 4.5 (ต่อ) แสดงการเปรียบเทียบวิธีการประมาณค่าสูญหาย โดยพิจารณาจาก ค่า MAPE สำหรับข้อมูลอนุกรมเวลากรณีมีอิทธิพลจากปัจจัยแนวโน้มต่ำแต่ปัจจัยฤดูกาลสูง ($\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 0.3$, $\beta_3 = -0.6$, $\beta_4 = 0.6$ และ $\beta_5 = -0.8$)

σ	n	pm	MAPE				
			Complete	RI	NNI	WNR	EM
20	50	5	31.6216	31.6779	31.7441	31.6778*	31.6779
		10	33.2662	33.3322	33.5591	33.3289*	33.3322
		20	30.9168	31.1766	31.4693	31.0817*	31.1766
	100	5	27.4952	27.5335	27.5856	27.5334*	27.5335
		10	29.3787	29.4156	29.4792	29.4075*	29.4156
		20	28.8935	28.9421	29.2613	28.8735*	28.9421
	200	5	23.2063	23.2056	23.4375	23.2050*	23.2056
		10	22.7819	22.7947	23.2751	22.7899*	22.7947
		20	23.8309	23.8668	25.5654	23.8496*	23.8668
25	50	5	48.4863	48.4712	48.4710*	48.4727	48.4712
		10	47.0652	47.3584	47.3554*	47.3854	47.3584
		20	46.1120	46.4835	46.4712	46.3743*	46.4835
	100	5	41.7283	41.7330	41.6376*	41.7207	41.7330
		10	42.3219	42.3290	42.3226*	42.3269	42.3290
		20	43.9321	44.0010	43.9321	43.9215*	44.0010
	200	5	34.7386	34.7295	35.0212	34.7281*	34.7295
		10	33.7053	33.7037	34.5022	33.7002*	33.7037
		20	36.4000	36.3964	38.0177	36.3926*	36.3964

หมายเหตุ

* หมายถึง วิธีการประมาณค่าสูญหายที่มีค่า MAPE ต่ำที่สุด

ตารางที่ 4.6 ผลการวิเคราะห์การทดสอบความแตกต่างของค่าเฉลี่ยของค่า MAPE ของวิธีการประมาณค่าสูญหายทั้ง 4 วิธี เปรียบเทียบกับกรณีที่ไม่มีข้อมูลสูญหายสำหรับข้อมูลอนุกรมเวลากรณีมีอิทธิพลจากปัจจัยแนวโน้มต่ำแต่ปัจจัยฤดูกาลสูง ($\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 0.3$, $\beta_3 =$ -0.6, $\beta_4 = 0.6$ และ $\beta_5 =$ -0.8)

ส่วนเบี่ยงเบนมาตรฐาน	ขนาดตัวอย่าง	ร้อยละการสูญหาย	วิธีการประมาณค่าสูญหายที่ไม่แตกต่าง จากกรณีที่ไม่มีข้อมูลสูญหาย
5	50	5	RI NNI WNR EM
		10	RI WNR EM
		20	WNR*
	100	5	RI WNR EM
		10	RI WNR EM
		20	RI* EM*
	200	5	RI WNR EM
		10	RI WNR EM
		20	RI EM
10	50	5	RI NNI WNR EM
		10	RI NNI WNR EM
		20	WNR*
	100	5	RI NNI WNR EM
		10	RI WNR EM
		20	WNR
	200	5	RI WNR EM
		10	RI WNR EM
		20	RI WNR EM

* หมายถึง จากการทดสอบความแตกต่างของค่าเฉลี่ยค่า MAPE พบว่า ไม่มีวิธีการใดที่ไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย แต่วิธีการนี้เป็นวิธีที่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหายน้อยที่สุด

ตารางที่ 4.6 (ต่อ) ผลการวิเคราะห์การทดสอบความแตกต่างของค่าเฉลี่ยของค่า MAPE ของวิธีการประมาณค่าสูญหายทั้ง 4 วิธี เปรียบเทียบกับกรณีที่ไม่มีข้อมูลสูญหายสำหรับข้อมูลอนุกรมเวลากรณีมีอิทธิพลจากปัจจัยแนวโน้มต่ำแต่ปัจจัยฤดูกาลสูง ($\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 0.3$, $\beta_3 =$ -0.6, $\beta_4 = 0.6$ และ $\beta_5 = -0.8$)

ส่วนเบี่ยงเบนมาตรฐาน	ขนาดตัวอย่าง	ร้อยละการสูญหาย	วิธีการประมาณค่าสูญหายที่ไม่แตกต่าง จากกรณีที่ไม่มีข้อมูลสูญหาย
15	50	5	RI NNI WNR EM
		10	RI NNI WNR EM
		20	RI WNR EM
	100	5	RI NNI WNR EM
		10	RI NNI WNR EM
		20	RI WNR EM
	200	5	RI NNI WNR EM
		10	RI WNR EM
		20	RI WNR EM
20	50	5	RI NNI WNR EM
		10	RI NNI WNR EM
		20	RI NNI WNR EM
	100	5	RI NNI WNR EM
		10	RI NNI WNR EM
		20	RI NNI WNR EM
	200	5	RI NNI WNR EM
		10	RI WNR EM
		20	RI WNR EM

ตารางที่ 4.6 (ต่อ) ผลการวิเคราะห์การทดสอบความแตกต่างของค่าเฉลี่ยของค่า MAPE ของวิธีการประมาณค่าสูญหายทั้ง 4 วิธี เปรียบเทียบกับกรณีที่ไม่มีข้อมูลสูญหายสำหรับข้อมูลอนุกรมเวลากรณีมีอิทธิพลจากปัจจัยแนวโน้มต่ำแต่ปัจจัยฤดูกาลสูง ($\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 0.3$, $\beta_3 = -0.6$, $\beta_4 = 0.6$ และ $\beta_5 = -0.8$)

ส่วนเบี่ยงเบนมาตรฐาน	ขนาดตัวอย่าง	ร้อยละการสูญหาย	วิธีการประมาณค่าสูญหายที่ไม่แตกต่าง จากกรณีที่ไม่มีข้อมูลสูญหาย
25	50	5	RI NNI WNR EM
		10	RI NNI WNR EM
		20	RI NNI WNR EM
	100	5	RI NNI WNR EM
		10	RI NNI WNR EM
		20	RI NNI WNR EM
	200	5	RI NNI WNR EM
		10	RI NNI WNR EM
		20	RI WNR EM

หมายเหตุ

RI หมายถึง จากการทดสอบความแตกต่างของค่าเฉลี่ย MAPE วิธี Regression Imputation ให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย

NNI หมายถึง จากการทดสอบความแตกต่างของค่าเฉลี่ย MAPE วิธี Nearest Neighbor Imputationให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย

WNR หมายถึง จากการทดสอบความแตกต่างของค่าเฉลี่ย MAPE วิธี Weighted Nearest Neighbor and Regression Imputation ให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย

EM หมายถึง จากการทดสอบความแตกต่างของค่าเฉลี่ย MAPE วิธี EM algorithm (Expectation Maximization) ให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย

เมื่อพิจารณาเปรียบเทียบค่า MAPE สำหรับข้อมูลอนุกรมเวลากรณีมีปัจจัยแนวโน้มต่ำแต่ปัจจัยฤดูกาลสูงคือกรณีที่กำหนดให้สัมประสิทธิ์การถดถอย $\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 0.3$, $\beta_3 =$ -0.6, $\beta_4 = 0.6$ และ $\beta_5 =$ -0.8 จากวิธีการประมาณค่าสูญหายทั้ง 4 วิธีและกรณีที่ไม่มีข้อมูลสูญหายเปรียบเทียบกัน พบว่าค่า MAPE จะเพิ่มขึ้นเมื่อข้อมูลมีส่วนเบี่ยงเบนมาตรฐานเพิ่มขึ้น แต่มีค่าลดลงเมื่อขนาดตัวอย่างใหญ่ และเมื่อร้อยละการสูญหายของข้อมูลเพิ่มสูงขึ้นค่า MAPE จากวิธีการประมาณค่าสูญหายจะแตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย (ค่าจริง) เพิ่มมากขึ้น

เมื่อส่วนเบี่ยงมาตรฐานอยู่ในระดับต่ำ ($\sigma =$ 5) และตัวอย่างมีขนาดเล็ก (n = 50) วิธี WNR มีค่า MAPE ต่ำที่สุด เมื่อขนาดตัวอย่างเพิ่มขึ้นเป็นขนาดปานกลางและขนาดใหญ่ (n = 100 และ 200) วิธีที่มีค่า MAPE ต่ำสุดมี 2 วิธี คือ RIและ EM ยกเว้นเมื่อตัวอย่างมีขนาดปานกลาง (100) ที่ร้อยละการสูญหายอยู่ในระดับต่ำ (5) วิธีการที่มีค่า MAPE ต่ำที่สุด คือ วิธี WNR

จากการทดสอบความแตกต่างของค่าเฉลี่ยค่า MAPE เมื่อตัวอย่างมีขนาดเล็ก (n = 50) ที่ร้อยละของการสูญหายอยู่ในระดับต่ำ (5) วิธีการประมาณค่าสูญหายทั้ง 4 วิธีให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย เมื่อร้อยละของการสูญหายเพิ่มขึ้นเป็นระดับปานกลาง (10) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย และเมื่อร้อยละของการสูญหายเพิ่มขึ้นเป็นระดับสูง (20) ไม่มีวิธีการใดที่ให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย แต่วิธีที่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหายน้อยที่สุด คือ วิธี WNR เมื่อตัวอย่างมีขนาดปานกลาง (n = 100) และร้อยละของการสูญหายอยู่ในระดับต่ำและปานกลาง (5 และ 10) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย และเมื่อร้อยละของการสูญหายเพิ่มขึ้นเป็นระดับสูง (20) ไม่มีวิธีการใดที่ให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย แต่วิธีที่แตกต่างจากกรณีที่ไม่มีข้อข้อมูลสูญหายน้อยที่สุด คือ วิธี RI และ EM เมื่อตัวอย่างมีขนาดใหญ่ (n = 200) และร้อยละของการสูญหายอยู่ในระดับต่ำและปานกลาง (5 และ 10) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย และเมื่อร้อยละของการสูญหายเพิ่มขึ้นเป็นระดับสูง (20)พบว่าวิธี RI และ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย จากกรณีที่ไม่มีข้อมูลสูญหายที่ระดับนัยสำคัญ 0.05

เมื่อส่วนเบี่ยงมาตรฐานอยู่ในระดับที่ค่อนข้างต่ำ ($\sigma =$ 10) เกือบทุกกรณีวิธีที่มีค่า MAPE ต่ำสุดคือวิธี WNR ยกเว้นเมื่อตัวอย่างขนาดใหญ่ (n = 200) และร้อยละการสูญหายอยู่ในระดับสูง (20) วิธีการประมาณค่าสูญหายที่มีค่า MAPE ต่ำที่สุด คือ วิธี RI และ EM

จากการทดสอบความแตกต่างของค่าเฉลี่ยค่า MAPE เมื่อตัวอย่างมีขนาดเล็ก (n = 50) ที่ร้อยละของการสูญหายอยู่ในระดับต่ำและระดับปานกลาง(5 และ10) วิธีการประมาณค่าสูญหายทั้ง 4 วิธีให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย และเมื่อร้อยละของการสูญหายเพิ่มขึ้นเป็นระดับสูง (20) ไม่มีวิธีการใดที่ให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหายหาย แต่วิธีที่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหายน้อยที่สุด คือ วิธี WNR กรณีที่ตัวอย่างมีขนาดปานกลาง (n = 100) ที่ร้อยละของการสูญหายอยู่ในระดับต่ำ (50) วิธีการประมาณค่าสูญหายทั้ง 4 วิธีให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย เมื่อร้อยละของการสูญหายเพิ่มขึ้นเป็นระดับปานกลาง (10) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย และเมื่อร้อยละของการสูญหายเพิ่มขึ้นเป็นระดับสูง (20) มีวิธี WNR เท่านั้นที่ให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย เมื่อตัวอย่างมีขนาดใหญ่ (n = 200) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ที่ระดับนัยสำคัญ 0.05

เมื่อส่วนเบี่ยงมาตรฐานอยู่ในระดับปานกลางและระดับค่อนข้างสูง (σ =15 และ 20) วิธีที่มีค่า MAPE ต่ำสุดคือวิธี WNR ในทุกระดับของขนาดตัวอย่างและร้อยละการสูญหายของข้อมูล

จากการทดสอบความแตกต่างของค่าเฉลี่ยค่า MAPE กรณีที่ส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับปานกลาง (σ =15) เมื่อตัวอย่างมีขนาดเล็กและปานกลาง (n = 50 และ 100) ที่ร้อยละของการสูญหายอยู่ในระดับต่ำและระดับปานกลาง (5 และ10) วิธีการประมาณค่าสูญหายทั้ง 4 วิธีให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย และเมื่อร้อยละของการสูญหายเพิ่มขึ้นเป็นระดับสูง (20) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย เมื่อตัวอย่างมีขนาดใหญ่ (n = 200) ที่ร้อยละของการสูญหายอยู่ในระดับต่ำ (5) วิธีการประมาณค่าสูญหายทั้ง 4 วิธีให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย และเมื่อร้อยละของการสูญหายเพิ่มขึ้นเป็นระดับปานกลางและระดับสูง (10 และ 20) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหายที่ระดับนัยสำคัญ 0.05

สำหรับกรณีที่ส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับค่อนข้างสูง(σ = 20) เมื่อตัวอย่างมีขนาดเล็กและปานกลาง (n = 50 และ 100) วิธีการประมาณค่าสูญหายทั้ง 4 วิธีให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย เมื่อตัวอย่างมีขนาดใหญ่ (n = 200) ที่ร้อยละของการสูญหายอยู่ในระดับต่ำ (5) วิธีการประมาณค่าสูญหายทั้ง 4 วิธีให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย เมื่อร้อยละการสูญหายของข้อมูลเพิ่มขึ้นมาอยู่ในระดับ

ปานกลางและสูง (10 และ 20) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ที่ระดับนัยสำคัญ 0.05

กรณีที่ส่วนเบี่ยงมาตรฐานอยู่ในระดับสูง(σ = 25) เมื่อตัวอย่างมีขนาดเล็กและปานกลาง (n = 50 และ 100) วิธีการที่มีค่า MAPE ต่ำที่สุด คือ วิธี NNI ยกเว้นเมื่อร้อยละการสูญหายของข้อมูลอยู่ในระดับสูง (20) วิธี WNR มีค่า MAPE ต่ำสุด เมื่อตัวอย่างมีขนาดใหญ่ (n = 200) วิธีการประมาณค่าสูญหายที่มีค่า MAPE ต่ำที่สุด คือ วิธี WNR ทุกระดับร้อยละการสูญหาย

จากการทดสอบความแตกต่างของค่าเฉลี่ยค่า MAPE เกือบทุกกรณี วิธีการประมาณค่าสูญหายทั้ง 4 วิธีให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ยกเว้นเมื่อตัวอย่างมีขนาดใหญ่ (n = 200) ที่ร้อยละของการสูญหายอยู่ในระดับสูง (20) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหายที่ระดับนัยสำคัญ 0.05

ข้อมูลอนุกรมเวลากรณีมีอิทธิพลจากปัจจัยแนวโน้มต่ำแต่ปัจจัยฤดูกาลสูง ($\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 0.3$, $\beta_3 = -0.6$, $\beta_4 = 0.6$ และ $\beta_5 = -0.8$) ข้อมูลมีการแกว่งของขึ้นลงตามฤดูกาลสูง ทำให้ฤดูกาลเดียวกันข้อมูลมีความคล้ายคลึงกันวิธี WNR จึงประมาณค่าสูญหายได้ดี และเมื่อส่วนเบี่ยงเบนมาตรฐานสูงการกระจายของข้อมูลเพิ่มมากขึ้น กรณีที่วิธี NNI ซึ่งแทนค่าข้อมูลสูญหายด้วยค่าของหน่วยตัวอย่างที่คล้ายคลึงกัน ประมาณค่าสูญหายได้ดีนั้น เนื่องจากวิธี RI เกิดความคลาดเคลื่อนมากขึ้น ส่งผลให้วิธี WNR เกิดความคลาดเคลื่อนเพิ่มขึ้นเช่นเดียวกัน

ตารางที่ 4.7 แสดงการเปรียบเทียบวิธีการประมาณค่าสูญหาย โดยพิจารณาจาก ค่า MAPE สำหรับข้อมูลอนุกรมเวลากรณีมีอิทธิพลจากปัจจัยแนวโน้มสูงแต่ปัจจัยฤดูกาลต่ำ ($\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 0.8$, $\beta_3 = 0.1$, $\beta_4 =$ -0.1และ $\beta_5 =$-0.2)

σ	n	pm	MAPE				
			Complete	RI	NNI	WNR	EM
5	50	5	4.7734	4.7803*	5.0268	4.7833	4.7803*
		10	4.8213	4.8574*	6.4142	4.8895	4.8574*
		20	4.8402	4.9103*	7.8183	5.0407	4.9103*
	100	5	4.0755	4.0792*	4.6582	4.0826	4.0792*
		10	4.1184	4.1275*	9.1106	4.1879	4.1275*
		20	4.1780	4.2027*	11.4668	4.3811	4.2027*
	200	5	3.1918	3.1927*	4.5948	3.1967	3.1927*
		10	3.1795	3.1818*	10.8755	3.2337	3.1818*
		20	3.2027	3.2098*	14.6138	3.3435	3.2098*
10	50	5	9.8489	9.8722	9.9541	9.8660*	9.8722
		10	9.8679	9.9181	10.7545	9.9126*	9.9181
		20	10.1388	10.3001*	12.6478	10.3937	10.3001*
	100	5	8.4502	8.4583*	9.1100	8.4584	8.4583*
		10	8.2513	8.2694*	9.5106	8.2709	8.2694*
		20	8.5874	8.6350*	12.3813	8.6773	8.6350*
	200	5	6.5013	6.5035*	8.4404	6.5097	6.5035*
		10	6.4853	6.4887*	10.6469	6.5041	6.4887*
		20	6.4703	6.4831*	16.9168	6.6170	6.4831*
15	50	5	15.8888	15.9109	16.1531	15.9107*	15.9109
		10	15.8523	15.9204	16.1987	15.9075*	15.9204
		20	15.8265	16.1051	17.7362	16.0411*	16.1051
	100	5	13.0447	13.0525*	13.8569	13.0586	13.0525*
		10	13.1375	13.1541*	14.2739	13.1564	13.1541*
		20	13.2033	13.2650*	16.3105	13.2732	13.2650*
	200	5	10.2665	10.2694*	11.5462	10.2825	10.2694*
		10	10.4239	10.4229*	15.1876	10.4471	10.4229*
		20	10.5029	10.5112*	18.2338	10.5610	10.5112*

ตารางที่ 4.7 (ต่อ) แสดงการเปรียบเทียบวิธีการประมาณค่าสูญหาย โดยพิจารณาจาก ค่า MAPE สำหรับข้อมูลอนุกรมเวลากรณีมีอิทธิพลจากปัจจัยแนวโน้มสูงแต่ปัจจัยฤดูกาลต่ำ ($\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 0.8$, $\beta_3 = 0.1$, $\beta_4 =$ -0.1และ $\beta_5 =$-0.2)

σ	n	pm	MAPE				
			Complete	RI	NNI	WNR	EM
20	50	5	23.2585	23.2768	23.3868	23.2767*	23.2768
		10	25.4473	25.5433	25.8428	25.5228*	25.5433
		20	23.7298	24.0126	24.7971	23.9072*	24.0126
	100	5	20.6188	20.6252	20.8737	20.6242*	20.6252
		10	21.0774	21.0932	22.6485	21.0917*	21.0932
		20	19.9177	19.9683	22.7593	19.9625*	19.9683
	200	5	14.3546	14.3588*	14.7073	14.3627	14.3588*
		10	14.8671	14.8624*	18.6312	14.8923	14.8624*
		20	15.4187	15.4221*	22.0527	15.5000	15.4221*
25	50	5	34.5454	34.5013	34.6280	34.5009*	34.5013
		10	33.1201	33.1647	33.4393	33.1472*	33.1647
		20	35.5324	35.5855	36.1075	35.4832*	35.5855
	100	5	27.9247	27.9220	28.1339	27.9217*	27.9220
		10	29.6371	29.6378	30.7172	29.6328*	29.6378
		20	30.3721	30.3726	33.2185	30.3709*	30.3726
	200	5	20.8844	20.8730*	21.4546	20.8873	20.8730*
		10	21.1891	21.1895*	23.5693	21.2275	21.1895*
		20	20.5261	20.5346*	29.0837	20.6553	20.5346*

หมายเหตุ

* หมายถึง วิธีการประมาณค่าสูญหายที่มีค่า MAPE ต่ำที่สุด

ตารางที่ 4.8 ผลการวิเคราะห์การทดสอบความแตกต่างของค่าเฉลี่ยของค่า MAPE ของวิธีการประมาณค่าสูญหายทั้ง 4 วิธี เปรียบเทียบกับกรณีที่ไม่มีข้อมูลสูญหายสำหรับข้อมูลอนุกรมเวลากรณีมีอิทธิพลจากปัจจัยแนวโน้มสูงแต่ปัจจัยฤดูกาลต่ำ ($\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 0.8$, $\beta_3 = 0.1$, $\beta_4 =$ -0.1และ $\beta_5 =$-0.2)

ส่วนเบี่ยงเบนมาตรฐาน	**ขนาดตัวอย่าง**	**ร้อยละการสูญหาย**	**วิธีการประมาณค่าสูญหายที่ไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย**
5	50	5	RI WNR EM
		10	RI WNR EM
		20	RI* EM*
	100	5	RI WNR EM
		10	RI EM
		20	RI EM
	200	5	RI WNR EM
		10	RI EM
		20	RI EM
10	50	5	RI WNR EM
		10	RI WNR EM
		20	RI* EM*
	100	5	RI WNR EM
		10	RI WNR EM
		20	RI EM
	200	5	RI WNR EM
		10	RI WNR EM
		20	RI EM

* หมายถึง จากการทดสอบความแตกต่างของค่าเฉลี่ยค่า MAPE พบว่า ไม่มีวิธีการใดที่ไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย แต่วิธีการนี้เป็นวิธีที่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหายน้อยที่สุด

ตารางที่ 4.8 (ต่อ) ผลการวิเคราะห์การทดสอบความแตกต่างของค่าเฉลี่ยของค่า MAPE ของวิธีการประมาณค่าสูญหายทั้ง 4 วิธี เปรียบเทียบกับกรณีที่ไม่มีข้อมูลสูญหายสำหรับข้อมูลอนุกรมเวลากรณีมีอิทธิพลจากปัจจัยแนวโน้มสูงแต่ปัจจัยฤดูกาลต่ำ ($\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 0.8$, $\beta_3 = 0.1$, $\beta_4 =$ -0.1และ $\beta_5 =$ -0.2)

ส่วนเบี่ยงเบนมาตรฐาน	ขนาดตัวอย่าง	ร้อยละการสูญหาย	วิธีการประมาณค่าสูญหายที่ไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย
15	50	5	RI WNR EM
		10	RI WNR EM
		20	WNR
	100	5	RI WNR EM
		10	RI WNR EM
		20	RI WNR EM
	200	5	RI WNR EM
		10	RI WNR EM
		20	RI WNR EM
20	50	5	RI NNI WNR EM
		10	RI NNI WNR EM
		20	RI WNR EM
	100	5	RI NNI WNR EM
		10	RI WNR EM
		20	RI WNR EM
	200	5	RI WNR EM
		10	RI WNR EM
		20	RI WNR EM

ตารางที่ 4.8 (ต่อ) ผลการวิเคราะห์การทดสอบความแตกต่างของค่าเฉลี่ยของค่า MAPE ของวิธีการประมาณค่าสูญหายทั้ง 4 วิธี เปรียบเทียบกับกรณีที่ไม่มีข้อมูลสูญหายสำหรับข้อมูลอนุกรมเวลากรณีมีอิทธิพลจากปัจจัยแนวโน้มสูงแต่ปัจจัยฤดูกาลต่ำ ($\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 0.8$, $\beta_3 = 0.1$, $\beta_4 = -0.1$ และ $\beta_5 = -0.2$)

ส่วนเบี่ยงเบนมาตรฐาน	**ขนาดตัวอย่าง**	**ร้อยละการสูญหาย**	**วิธีการประมาณค่าสูญหายที่ไม่แตกต่าง จากกรณีที่ไม่มีข้อมูลสูญหาย**
25	50	5	RI NNI WNR EM
		10	RI NNI WNR EM
		20	RI NNI WNR EM
	100	5	RI NNI WNR EM
		10	RI WNR EM
		20	RI WNR EM
	200	5	RI NNI WNR EM
		10	RI WNR EM
		20	RI WNR EM

หมายเหตุ

RI หมายถึง จากการทดสอบความแตกต่างของค่าเฉลี่ย MAPE วิธี Regression Imputation ให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย

NNI หมายถึง จากการทดสอบความแตกต่างของค่าเฉลี่ย MAPE วิธี Nearest Neighbor Imputationให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย

WNR หมายถึง จากการทดสอบความแตกต่างของค่าเฉลี่ย MAPE วิธี Weighted Nearest Neighbor and Regression Imputation ให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย

EM หมายถึง จากการทดสอบความแตกต่างของค่าเฉลี่ย MAPE วิธี EM algorithm (Expectation Maximization) ให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย

เมื่อพิจารณาเปรียบเทียบค่า MAPE สำหรับข้อมูลอนุกรมเวลากรณีมีอิทธิพลจากปัจจัยแนวโน้มสูงแต่ปัจจัยฤดูกาลต่ำ คือกรณีที่กำหนดให้สัมประสิทธิ์การถดถอย $\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 0.8$, $\beta_3 = 0.1$, $\beta_4 =$ -0.1และ $\beta_5 =$-0.2 จากวิธีการประมาณค่าสูญหายทั้ง 4 วิธีและกรณีที่ไม่มีข้อมูลสูญหายเปรียบเทียบกัน พบว่าค่า MAPE จะเพิ่มขึ้นเมื่อข้อมูลมีส่วนเบี่ยงเบนมาตรฐานเพิ่มขึ้น แต่มีค่าลดลงเมื่อขนาดตัวอย่างใหญ่ขึ้น และเมื่อร้อยละการสูญหายของข้อมูลเพิ่มสูงขึ้นค่า MAPE จากวิธีการประมาณค่าสูญหายจะมีผลต่างจากกรณีที่ไม่มีข้อมูลสูญหายเพิ่มมากขึ้น

เมื่อส่วนเบี่ยงมาตรฐานอยู่ในระดับต่ำ ($\sigma =$ 5) วิธีที่มีค่า MAPE ต่ำที่สุดมี 2 วิธี คือ RI และ EM

จากการทดสอบความแตกต่างของค่าเฉลี่ยค่า MAPE เมื่อตัวอย่างมีขนาดเล็ก (n = 50) ที่ร้อยละของการสูญหายอยู่ในระดับต่ำและปานกลาง (5 และ10) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย และเมื่อร้อยละของการสูญหายเพิ่มขึ้นเป็นระดับสูง (20) ไม่มีวิธีการประมาณค่าสูญหายที่ให้ผลไม่แตกต่างจากกรณีไม่มีข้อมูลสูญหาย แต่วิธีที่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหายน้อยที่สุด คือ วิธี RI และ EM เมื่อตัวอย่างมีขนาดปานกลางและขนาดใหญ่ (n = 100 และ 200) ที่ร้อยละของการสูญหายอยู่ในระดับต่ำพบว่า 3 วิธีคือ วิธี RI, WNR และ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย และเมื่อร้อยละของการสูญหายเพิ่มขึ้นเป็นระดับปานกลางและระดับสูง (10 และ 20) พบว่า วิธี RIและ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหายที่ระดับนัยสำคัญ 0.05

เมื่อส่วนเบี่ยงมาตรฐานอยู่ในระดับที่ค่อนข้างต่ำ ($\sigma =$ 10) เกือบทุกกรณีวิธีที่มีค่า MAPE ต่ำสุดคือวิธี RI และ EM ยกเว้นเมื่อตัวอย่างมีขนาดเล็ก (n = 50) ที่ร้อยละการสูญหายอยู่ในระดับต่ำและปานกลาง (5 และ 10) วิธีการประมาณค่าสูญหายที่มีค่า MAPE ต่ำที่สุด คือ วิธี WNR

จากการทดสอบความแตกต่างของค่าเฉลี่ยค่า MAPE เมื่อตัวอย่างมีขนาดเล็ก (n = 50) ที่ร้อยละของการสูญหายอยู่ในระดับต่ำและปานกลาง (5 และ 10) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย และเมื่อร้อยละของการสูญหายเพิ่มขึ้นเป็นระดับสูง (20) ไม่มีวิธีการประมาณค่าสูญหายที่ให้ผลไม่แตกต่างจากกรณีไม่มีข้อมูลสูญหาย แต่วิธีที่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหายน้อยที่สุด คือ วิธี RI และ EM เมื่อตัวอย่างมีขนาดปานกลางและขนาดใหญ่ (n = 100 และ 200) ที่ร้อยละของการสูญหายอยู่ในระดับต่ำและปานกลาง (5 และ 10) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ให้ผลไม่แตกต่างกันและไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย และเมื่อร้อยละของการสูญหายเพิ่มขึ้นเป็นระดับสูง

(20) พบว่า วิธี RIและ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหายที่ระดับนัยสำคัญ 0.05

เมื่อส่วนเบี่ยงมาตรฐานอยู่ในระดับปานกลาง (σ = 15) เมื่อตัวอย่างมีขนาดเล็ก (n = 50) วิธีการประมาณค่าสูญหายที่มีค่า MAPE ต่ำที่สุด คือ วิธี WNR เมื่อตัวอย่างมีขนาดปานกลางและขนาดใหญ่ (n = 100 และ 200) วิธีการประมาณค่าสูญหายที่มีค่า MAPE ต่ำสุดคือวิธี RI และ EM

จากการทดสอบความแตกต่างของค่าเฉลี่ยค่า MAPE เมื่อตัวอย่างมีขนาดเล็ก (n = 50) ที่ร้อยละของการสูญหายอยู่ในระดับต่ำและปานกลาง (5 และ10) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย และเมื่อร้อยละของการสูญหายเพิ่มขึ้นเป็นระดับสูง (20) มีวิธีการประมาณค่าสูญหายเพียงวิธีเดียว คือ วิธี WNR ที่ให้ผลไม่แตกต่างจากกรณีไม่มีข้อมูลสูญหาย เมื่อตัวอย่างมีขนาดปานกลางและขนาดใหญ่ (n = 100 และ 200) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหายที่ระดับนัยสำคัญ 0.05 ทุกระดับร้อยละการสูญหายของข้อมูล

เมื่อส่วนเบี่ยงมาตรฐานอยู่ในระดับค่อนข้างสูงและระดับสูง (σ = 20 และ 25) เมื่อตัวอย่างมีขนาดเล็กและปานกลาง (n = 50 และ 100) วิธีการประมาณค่าสูญหายที่มีค่า MAPE ต่ำที่สุด คือ วิธี WNR เมื่อตัวอย่างมีขนาดใหญ่ (200) วิธีการประมาณค่าสูญหายที่มีค่า MAPE ต่ำสุดคือวิธี RI และ EM

สำหรับกรณีที่ส่วนเบี่ยงมาตรฐานอยู่ในระดับค่อนข้างสูง(σ = 20) เมื่อตัวอย่างมีขนาดเล็ก (n = 50) ที่ร้อยละของการสูญหายอยู่ในระดับต่ำและปานกลาง (5 และ 10) วิธีการประมาณค่าสูญหายทั้ง 4 วิธีให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย และเมื่อร้อยละของการสูญหายเพิ่มขึ้นเป็นระดับสูง (20) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย เมื่อตัวอย่างมีขนาดปานกลาง(n = 100) ที่ร้อยละของการสูญหายอยู่ในระดับต่ำ (5) วิธีการประมาณค่าสูญหายทั้ง 4 วิธีให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ที่ร้อยละของการสูญหายอยู่ในระดับปานกลางและสูง (10และ 20) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย เมื่อตัวอย่างมีขนาดใหญ่ (n = 200) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย คือไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหายที่ระดับนัยสำคัญ 0.05 ทุกระดับร้อยละการสูญหายของข้อมูล

สำหรับกรณีที่ส่วนเบี่ยงมาตรฐานอยู่ในระดับสูง(σ = 25) เมื่อตัวอย่างมีขนาดเล็ก (n = 50) วิธีการประมาณค่าสูญหายทั้ง 4 วิธีให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย เมื่อตัวอย่างมีขนาดปานกลางและขนาดใหญ่ (n = 100 และ 200) ที่ร้อยละของการสูญ

หายอยู่ในระดับต่ำ (5) วิธีการประมาณค่าสูญหายทั้ง 4 วิธีให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ที่ร้อยละของการสูญหายอยู่ในระดับปานกลางและสูง (10และ 20) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ที่ระดับนัยสำคัญ 0.05

ข้อมูลอนุกรมเวลากรณีมีอิทธิพลจากปัจจัยแนวโน้มสูงแต่ปัจจัยฤดูกาลต่ำ ($\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 0.8$, $\beta_3 = 0.1$, $\beta_4 =$ -0.1และ $\beta_5 =$-0.2) เนื่องจากมีความสัมพันธ์ระหว่างข้อมูลกับเวลา เมื่อมีอิทธิพลจากปัจจัยแนวโน้มสูง ส่งผลให้ วิธีRI และEM เป็นวิธีที่มีค่า MAPE ต่ำกว่าวิธีอื่นๆที่นำมาเปรียบเทียบ แต่เมื่อส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับสูง การกระจายของข้อมูลสูงขึ้นวิธี WNR จึงเป็นวิธีที่ดีกว่ายกเว้นกรณีที่ตัวอย่างมีขนาดใหญ่ซึ่งข้อมูลมีระยะยาวส่งผลให้วิธี RI และ EM ประมาณค่าสูญหายได้ดีกว่าวิธี WNR เพราะเมื่อข้อมูลมีระยะสั้นจะถูกรบกวนจากปัจจัยฤดูกาลมากกว่าข้อมูลระยะยาว

ตารางที่ 4.9 แสดงการเปรียบเทียบวิธีการประมาณค่าสูญหาย โดยพิจารณาจาก ค่า MAPE สำหรับข้อมูลอนุกรมเวลากรณีมีอิทธิพลจากปัจจัยแนวโน้มและปัจจัยฤดูกาลระดับปานกลาง ($\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 0.5$, $\beta_3 = 0.5$, $\beta_4 =$ -0.1และ $\beta_5 =$ -0.3)

σ	n	pm	MAPE				
			Complete	RI	NNI	WNR	EM
5	50	5	5.3204	5.3341	5.4525	5.3330*	5.3341
		10	5.2755	5.3245*	6.2905	5.3500	5.3245*
		20	5.2564	5.3494*	7.3096	5.4106	5.3494*
	100	5	4.7709	4.7785*	5.3586	4.7801	4.7785*
		10	4.8291	4.8501*	6.7622	4.8592	4.8501*
		20	4.6978	4.7319*	9.5519	4.8111	4.7319*
	200	5	3.9162	3.9181*	5.1596	3.9204	3.9181*
		10	3.8574	3.8613*	7.7572	3.8719	3.8613*
		20	3.8925	3.9043*	13.6655	3.9933	3.9043*
10	50	5	10.8025	10.8242	10.9473	10.8156*	10.8242
		10	10.8499	10.9792	11.2088	10.9481*	10.9792
		20	10.6500	10.8789	11.3635	10.8495*	10.8789
	100	5	9.8799	9.8900	10.2952	9.8847*	9.8900
		10	9.6857	9.7114	10.2612	9.7057*	9.7114
		20	9.8554	9.9322*	13.2092	9.9322*	9.9322*
	200	5	7.9033	7.9058*	8.8689	7.9061	7.9058*
		10	7.8628	7.8690*	9.6635	7.8738	7.8690*
		20	8.0249	8.0448*	16.9548	8.0825	8.0448*
15	50	5	19.3390	19.3634	19.4148	19.3578*	19.3634
		10	17.3099	17.3955	17.6682	17.3851*	17.3955
		20	17.8034	18.0863	18.3859	18.0074*	18.0863
	100	5	16.2161	16.2226	16.3294	16.2224*	16.2226
		10	15.7218	15.7542	16.1673	15.7435*	15.7542
		20	15.4424	15.5326	17.2770	15.5183*	15.5326
	200	5	12.5134	12.5172*	13.1384	12.5174	12.5172*
		10	12.0704	12.0791*	14.0440	12.0795	12.0791*
		20	12.5375	12.5681*	18.5755	12.5800	12.5681*

ตารางที่ 4.9 (ต่อ) แสดงการเปรียบเทียบวิธีการประมาณค่าสูญหาย โดยพิจารณาจาก ค่า MAPE สำหรับข้อมูลอนุกรมเวลากรณีมีอิทธิพลจากปัจจัยแนวโน้มและปัจจัยฤดูกาลระดับปานกลาง ($\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 0.5$, $\beta_3 = 0.5$, $\beta_4 = -0.1$ และ $\beta_5 = -0.3$)

σ	n	pm	MAPE				
			Complete	RI	NNI	WNR	EM
20	50	5	30.2684	30.3338	30.3155	30.3125*	30.3338
		10	29.4462	29.5426	29.4847	29.4813*	29.5426
		20	27.7652	28.0947	28.4244	27.9623*	28.0947
	100	5	23.5135	23.5180	23.5346	23.5106*	23.5180
		10	22.5617	22.5993	22.9909	22.5829*	22.5993
		20	23.9316	24.0695	25.1419	24.0172*	24.0695
	200	5	19.0018	19.0086*	19.2296	19.0980	19.0086*
		10	18.3415	18.3519*	20.3476	18.3701	18.3519*
		20	18.1773	18.1995*	22.5373	18.2050	18.1995*
25	50	5	45.4110	45.3632	45.3441*	45.3465	45.3632
		10	41.0517	41.2093	41.1661*	41.1668	41.2093
		20	40.3084	40.4629	40.5354	40.3427*	40.4629
	100	5	37.6443	37.6409	37.7034	37.6402*	37.6409
		10	35.4019	35.4030	35.9553	35.4019*	35.4030
		20	34.8382	34.9255	35.8635	34.8578*	34.9255
	200	5	26.9383	26.9325*	27.2336	26.9379	26.9325*
		10	27.4334	27.4178*	28.7708	27.4446	27.4178*
		20	28.7274	28.6991*	32.6939	28.7428	28.6991*

หมายเหตุ

* หมายถึง วิธีการประมาณค่าสูญหายที่มีค่า MAPE ต่ำที่สุด

ตารางที่ 4.10 ผลการวิเคราะห์การทดสอบความแตกต่างของค่าเฉลี่ยของค่า MAPE ของวิธีการประมาณค่าสูญหายทั้ง 4 วิธี เปรียบเทียบกับกรณีที่ไม่มีข้อมูลสูญหายสำหรับข้อมูลอนุกรมเวลากรณีมีอิทธิพลจากปัจจัยแนวโน้มและปัจจัยฤดูกาลระดับปานกลาง ($\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 0.5$, $\beta_3 = 0.5$, $\beta_4 = -0.1$ และ $\beta_5 = -0.3$)

ส่วนเบี่ยงเบนมาตรฐาน	ขนาดตัวอย่าง	ร้อยละการสูญหาย	วิธีการประมาณค่าสูญหายที่ไม่แตกต่าง จากกรณีที่ไม่มีข้อมูลสูญหาย
5	50	5	RI WNR EM
		10	RI EM
		20	RI* EM*
	100	5	RI WNR EM
		10	RI WNR EM
		20	RI EM
	200	5	RI WNR EM
		10	RI WNR EM
		20	RI EM
10	50	5	RI WNR EM
		10	WNR
		20	WNR*
	100	5	RI WNR EM
		10	RI WNR EM
		20	RI WNR EM
	200	5	RI WNR EM
		10	RI WNR EM
		20	RI EM

* หมายถึง จากการทดสอบความแตกต่างของค่าเฉลี่ยค่า MAPE พบว่า ไม่มีวิธีการใดที่ไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย แต่วิธีการนี้เป็นวิธีที่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหายน้อยที่สุด

ตารางที่ 4.10 (ต่อ) ผลการวิเคราะห์การทดสอบความแตกต่างของค่าเฉลี่ยของค่า MAPE ของวิธีการประมาณค่าสูญหายทั้ง 4 วิธี เปรียบเทียบกับกรณีที่ไม่มีข้อมูลสูญหายสำหรับข้อมูลอนุกรมเวลา กรณีมีอิทธิพลจากปัจจัยแนวโน้มและปัจจัยฤดูกาลระดับปานกลาง ($\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 0.5$, $\beta_3 = 0.5$, $\beta_4 = -0.1$ และ $\beta_5 = -0.3$)

ส่วนเบี่ยงเบนมาตรฐาน	ขนาดตัวอย่าง	ร้อยละการสูญหาย	วิธีการประมาณค่าสูญหายที่ไม่แตกต่าง จากกรณีที่ไม่มีข้อมูลสูญหาย
15	50	5	RI NNI WNR EM
		10	RI NNI WNR EM
		20	RI WNR EM
	100	5	RI NNI WNR EM
		10	RI WNR EM
		20	RI WNR EM
	200	5	RI WNR EM
		10	RI WNR EM
		20	RI WNR EM
20	50	5	RI NNI WNR EM
		10	RI NNI WNR EM
		20	RI NNI WNR EM
	100	5	RI NNI WNR EM
		10	RI NNI WNR EM
		20	RI WNR EM
	200	5	RI NNI WNR EM
		10	RI WNR EM
		20	RI WNR EM

ตารางที่ 4.10 (ต่อ) ผลการวิเคราะห์การทดสอบความแตกต่างของค่าเฉลี่ยของค่า MAPE ของวิธีการประมาณค่าสูญหายทั้ง 4 วิธี เปรียบเทียบกับกรณีที่ไม่มีข้อมูลสูญหายสำหรับข้อมูลอนุกรมเวลา กรณีมีอิทธิพลจากปัจจัยแนวโน้มและปัจจัยฤดูกาลระดับปานกลาง ($\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 0.5$, $\beta_3 = 0.5$, $\beta_4 =$ -0.1และ $\beta_5 =$ -0.3)

ส่วนเบี่ยงเบนมาตรฐาน	ขนาดตัวอย่าง	ร้อยละการสูญหาย	วิธีการประมาณค่าสูญหายที่ไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย
25	50	5	RI NNI WNR EM
		10	RI NNI WNR EM
		20	RI NNI WNR EM
	100	5	RI NNI WNR EM
		10	RI NNI WNR EM
		20	RI WNR EM
	200	5	RI NNI WNR EM
		10	RI WNR EM
		20	RI WNR EM

หมายเหตุ

RI หมายถึง จากการทดสอบความแตกต่างของค่าเฉลี่ย MAPE วิธี Regression Imputation ให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย

NNI หมายถึง จากการทดสอบความแตกต่างของค่าเฉลี่ย MAPE วิธี Nearest Neighbor Imputationให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย

WNR หมายถึง จากการทดสอบความแตกต่างของค่าเฉลี่ย MAPE วิธี Weighted Nearest Neighbor and Regression Imputation ให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย

EM หมายถึง จากการทดสอบความแตกต่างของค่าเฉลี่ย MAPE วิธี EM algorithm (Expectation Maximization) ให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย

เมื่อพิจารณาเปรียบเทียบค่า MAPE สำหรับข้อมูลอนุกรมเวลากรณีมีอิทธิพลจากปัจจัยแนวโน้มและปัจจัยฤดูกาลระดับปานกลาง คือกรณีที่กำหนดให้สัมประสิทธิ์การถดถอย $\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 0.5$, $\beta_3 = 0.5$, $\beta_4 = -0.1$ และ $\beta_5 = -0.3$ จากวิธีการประมาณค่าสูญหายทั้ง 4 วิธี และกรณีที่ไม่มีข้อมูลสูญหายเปรียบเทียบกัน พบว่าค่า MAPE จะเพิ่มขึ้นเมื่อข้อมูลมีส่วนเบี่ยงเบนมาตรฐานเพิ่มขึ้น แต่มีค่าลดลงเมื่อขนาดตัวอย่างใหญ่ และเมื่อร้อยละการสูญหายของข้อมูลเพิ่มสูงขึ้นค่า MAPE จากวิธีการประมาณค่าสูญหายจะแตกต่างจากกรณีที่ไม่มีข้อมูลสูญหายเพิ่มมากขึ้น

เมื่อส่วนเบี่ยงมาตรฐานอยู่ในระดับต่ำ (σ = 5) เกือบทุกกรณีวิธีที่มีค่า MAPE ต่ำสุดคือวิธี RIและ EM ยกเว้นเมื่อตัวอย่างมีขนาดเล็ก (n = 50) และร้อยละการสูญหายอยู่ในระดับต่ำ (5) วิธีการประมาณค่าสูญหายที่มีค่า MAPE ต่ำที่สุด คือ วิธี WNR

จากการทดสอบความแตกต่างของค่าเฉลี่ยค่า MAPE เมื่อตัวอย่างมีขนาดเล็ก (n = 50) ที่ร้อยละของการสูญหายอยู่ในระดับต่ำ(5) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ที่ร้อยละของการสูญหายอยู่ในระดับปานกลาง (10) พบว่าวิธี RI และ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย และเมื่อร้อยละของการสูญหายเพิ่มขึ้นเป็นระดับสูง (20) ไม่มีวิธีการประมาณค่าสูญหายที่ให้ผลไม่แตกต่างจากกรณีไม่มีข้อมูลสูญหาย แต่วิธีที่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหายน้อยที่สุด คือ วิธี RI และ EM เมื่อตัวอย่างมีขนาดปานกลางและขนาดใหญ่ (n = 100 และ 200) ที่ร้อยละของการสูญหายอยู่ในระดับต่ำและปานกลาง (5 และ10) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ที่ร้อยละของการสูญหายเพิ่มขึ้นเป็นระดับสูง (20) พบว่าวิธี RI และ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ที่ระดับนัยสำคัญ 0.05

เมื่อส่วนเบี่ยงมาตรฐานอยู่ในระดับที่ค่อนข้างต่ำ (σ = 10) เมื่อตัวอย่างมีขนาดเล็กและปานกลาง (n = 50 และ 100) วิธีการประมาณค่าสูญหายที่มีค่า MAPE ต่ำที่สุด คือ วิธี WNR และที่ร้อยละการสูญหายอยู่ในระดับสูง (20) เมื่อตัวอย่างมีขนาดปานกลาง (100) วิธี RIและ EM มีค่า MAPE ต่ำสุดเท่ากันกับวิธี WNR เมื่อตัวอย่างมีขนาดใหญ่ (n = 200) วิธีการประมาณค่าสูญหายที่มีค่า MAPE ต่ำที่สุด คือ วิธี RI และ EM

จากการทดสอบความแตกต่างของค่าเฉลี่ยค่า MAPE เมื่อตัวอย่างมีขนาดเล็ก (n = 50) ที่ร้อยละของการสูญหายอยู่ในระดับต่ำ (5) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ที่ร้อยละของการสูญหายอยู่ในระดับปานกลาง

(10) พบว่าวิธี WNR ให้ผลไม่แตกต่างจากกรณีไม่มีข้อมูลสูญหายเพียงวิธีเดียว และเมื่อร้อยละของการสูญหายเพิ่มขึ้นเป็นระดับสูง (20) ไม่มีวิธีการประมาณค่าสูญหายที่ให้ผลไม่แตกต่างจากกรณีไม่มีข้อมูลสูญหาย แต่วิธีที่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหายน้อยที่สุด คือ วิธี WNR เมื่อตัวอย่างมีขนาดปานกลางและขนาดใหญ่ (n = 100 และ 200) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ยกเว้นเมื่อตัวอย่างมีขนาดใหญ่ (200) ที่ร้อยละการสูญหายของข้อมูลอยู่ในระดับสูง (20) วิธีการประมาณค่าสูญหาย 2 วิธี คือ วิธี RI และ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ที่ระดับนัยสำคัญ 0.05

เมื่อส่วนเบี่ยงมาตรฐานอยู่ในระดับปานกลาง (σ =15) กรณีที่ตัวอย่างมีขนาดเล็กและปานกลาง (n = 50 และ 100) วิธีการประมาณค่าสูญหายที่มีค่า MAPE ต่ำที่สุด คือ วิธี WNR เมื่อตัวอย่างมีขนาดใหญ่ (n = 200) วิธีการประมาณค่าสูญหายที่มีค่า MAPE ต่ำที่สุด คือ วิธี RI และ EM

จากการทดสอบความแตกต่างของค่าเฉลี่ยค่า MAPE เมื่อตัวอย่างมีขนาดเล็ก (n = 50) ที่ร้อยละของการสูญหายอยู่ในระดับต่ำและปานกลาง (5 และ10) พบว่าวิธีการประมาณค่าสูญหายทั้ง 4 วิธีให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย และเมื่อร้อยละของการสูญหายเพิ่มขึ้นเป็นระดับสูง (20) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย เมื่อตัวอย่างมีขนาดปานกลางและขนาดใหญ่ (n = 100 และ 200) พบว่าเกือบทุกกรณีวิธี RI, WNR และ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ยกเว้นเมื่อตัวอย่างมีขนาดปานกลาง (100) และร้อยละการสูญหายอยู่ในระดับต่ำ (5) วิธีการประมาณค่าสูญหายทั้ง 4 วิธี ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ที่ระดับนัยสำคัญ 0.05

เมื่อส่วนเบี่ยงมาตรฐานอยู่ในระดับค่อนข้างสูง (σ =20) เมื่อตัวอย่างมีขนาดเล็กและปานกลาง (n = 50 และ100) วิธีการประมาณค่าสูญหายที่มีค่า MAPE ต่ำที่สุด คือ วิธี WNR และเมื่อตัวอย่างมีขนาดใหญ่ (n = 200) วิธีการประมาณค่าสูญหายที่มีค่า MAPE ต่ำที่สุด คือ วิธี RI และ EM

จากการทดสอบความแตกต่างของค่าเฉลี่ยค่า MAPE สำหรับกรณีที่ส่วนเบี่ยงมาตรฐานอยู่ในระดับค่อนข้างสูง (σ = 20) เมื่อตัวอย่างมีขนาดเล็ก (n = 50) วิธีการประมาณค่าสูญหายทั้ง 4 วิธีให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย เมื่อตัวอย่างมีขนาดปานกลาง (n = 100) ที่ร้อยละการสูญหายอยู่ในระดับต่ำและปานกลาง (5 และ 10) วิธีการประมาณ

ค่าสูญหายทั้ง 4 วิธี ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ที่ร้อยละการสูญหายอยู่ในระดับสูง (20) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย และเมื่อตัวอย่างมีขนาดใหญ่ (200) ที่ร้อยละการสูญหายอยู่ในระดับต่ำวิธีการประมาณค่าสูญหายทั้ง 4 วิธี ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ที่ร้อยละการสูญหายของข้อมูลอยู่ในระดับปานกลางและสูง(10 และ 20) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหายที่ระดับนัยสำคัญ 0.05

เมื่อส่วนเบี่ยงมาตรฐานอยู่ในระดับสูง($\sigma =$ 25) เมื่อตัวอย่างมีขนาดเล็ก (n = 50) ที่ร้อยละการสูญหายอยู่ในระดับต่ำ (5) วิธีการประมาณค่าสูญหายที่มีค่า MAPE ต่ำที่สุด คือวิธี NNI เมื่อร้อยละการสูญหายอยู่ในระดับปานกลาง และ ระดับสูง (10 และ 20) วิธี WNR เป็นวิธีที่มีค่า MAPE ต่ำที่สุด เมื่อตัวอย่างมีขนาดปานกลาง (n = 100) วิธีการประมาณค่าสูญหายที่มีค่า MAPE ต่ำที่สุด คือ วิธี WNR และเมื่อตัวอย่างมีขนาดใหญ่ (n = 200) วิธีการประมาณค่าสูญหายที่มีค่า MAPE ต่ำที่สุด คือ วิธี RI และ EM

สำหรับกรณีที่ส่วนเบี่ยงมาตรฐานอยู่ในระดับสูง ($\sigma =$ 25) เมื่อตัวอย่างมีขนาดเล็กและปานกลาง (n = 50 และ 100) วิธีการประมาณค่าสูญหายทั้ง 4 วิธีให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ยกเว้นเมื่อตัวอย่างมีขนาดปานกลาง (100) ที่ร้อยละการสูญหายของข้อมูลอยู่ในระดับสูง (20) 3 วิธีคือ วิธี RI, WNR และ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย และเมื่อตัวอย่างมีขนาดใหญ่ (200) ที่ร้อยละการสูญหายอยู่ในระดับต่ำ (5) วิธีการประมาณค่าสูญหายทั้ง 4 วิธี ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย และที่ร้อยละการสูญหายของข้อมูลอยู่ในระดับปานกลางและสูง (10 และ 20) พบว่า 3 วิธีคือ วิธี RI, WNR และ EM ให้ผลไม่แตกต่างกัน และไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหายที่ระดับนัยสำคัญ 0.05

ข้อมูลอนุกรมเวลากรณีมีอิทธิพลจากปัจจัยแนวโน้มและปัจจัยฤดูกาลระดับปานกลาง ($\beta_0 = 0.5$, $\beta_1 = 1, \beta_2 = 0.5, \beta_3 = 0.5, \beta_4 = -0.1$และ $\beta_5 = -0.3$) ส่งผลให้อิทธิพลจากปัจจัยฤดูกาลจะเห็นได้ชัดเจนในข้อมูลระยะสั้น (ตัวอย่างขนาดเล็ก) เห็นได้จากวิธี WNR จะมีค่า MAPE ต่ำสำหรับข้อมูลระยะสั้น และวิธี NNI มีค่า MAPE ต่ำเมื่อส่วนเบี่ยงเบนมาตรฐานเพิ่มขึ้นมาอยู่ในระดับสูง สำหรับอิทธิพลจากปัจจัยแนวโน้มจะเห็นได้ชัดเจนในข้อมูลระยะยาว (ตัวอย่างขนาดใหญ่) โดยวิธี RI และ EM เป็นวิธีที่มีค่า MAPE ต่ำสุดทุกกรณี

บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบวิธีการประมาณค่าสูญหายของตัวแปรตามในสมการถดถอยเชิงเส้นพหุเพื่อการพยากรณ์ เมื่อการสูญหายของข้อมูลตัวแปรตามเป็นไปอย่างสุ่ม โดยทำการประมาณค่าสูญหายของตัวแปรตาม 4 วิธี ซึ่งได้แก่ วิธีการประมาณค่าสูญหายแบบ Regression Imputation (RI) วิธีการประมาณค่าสูญหายแบบ Nearest Neighbor Imputation (NNI) วิธีการประมาณค่าสูญหายแบบ Weighted Nearest Neighbor and Regression Imputation (WNR) และวิธีการประมาณค่าสูญหายแบบด้วย EM algorithm (Expectation Maximization) ซึ่งศึกษาในกรณีต่าง ๆ ดังนี้

กรณีที่1 ความถดถอยเชิงเส้นพหุ เมื่อเป็นข้อมูลภาคตัดขวางภายใต้ตัวแบบ

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \qquad ; i = 1,2,\ldots,r,r+1,\ldots,n$$

เมื่อ r เป็นจำนวนค่าสังเกตที่ทราบค่า

$n-r$ เป็นจำนวนค่าสังเกตที่สูญหาย

เมื่อกำหนดค่าสหสัมพันธ์ (Correlation) ระหว่างตัวแปรดังนี้

1) ค่าสหสัมพันธ์ (Correlation) ระหว่าง ค่าสังเกตของตัวแปรตาม(y_i)กับค่าสังเกตของตัวแปรอิสระตัวที่ 1(x_{1i})และค่าสังเกตของตัวแปรตาม(y_i)กับค่าสังเกตของตัวแปรอิสระตัวที่ 2 (x_{2i}) เท่ากัน คือ 0.7

2) ค่าสหสัมพันธ์ (Correlation) ระหว่าง ค่าสังเกตของตัวแปรตาม(y_i)กับค่าสังเกตของตัวแปรอิสระตัวที่ 1(x_{1i})และค่าสังเกตของตัวแปรตาม(y_i)กับค่าสังเกตของตัวแปรอิสระตัวที่ 2 (x_{2i}) คือ 0.9 และ 0.4 ตามลำดับ

และกำหนดให้ตัวแปรอิสระทั้ง 2 ไม่มีความสัมพันธ์กัน

กรณีที่2 ความถดถอยเชิงเส้นพหุ เมื่อเป็นข้อมูลอนุกรมเวลา ที่ประกอบด้วยปัจจัยแนวโน้มและปัจจัยฤดูกาล

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 t + \beta_3 Q_{1i} + \beta_4 Q_{2i} + \beta_5 Q_{3i} + \varepsilon_i \qquad ; i = 1,2,\ldots,r,r+1,\ldots,n$$

เมื่อ r เป็นจำนวนค่าสังเกตที่ทราบค่า

$n-r$ เป็นจำนวนค่าสังเกตที่สูญหาย

t เป็นค่าแนวโน้มของอนุกรมเวลา

Q_{li} เป็นตัวแปรบ่งชี้ของข้อมูลรายไตรมาส (มีทั้งหมด 4 ฤดูกาล) $l=1,2$ และ 3

เมื่อกำหนดให้ค่าสัมประสิทธิ์การถดถอยแตกต่างกัน 3 ลักษณะเพื่อให้ได้ข้อมูลอนุกรมเวลาที่มีลักษณะแตกต่างกัน 3 รูปแบบ คือ

รูปแบบที่ 1 $\beta_0=0.5$, $\beta_1=1$, $\beta_2=0.3$, $\beta_3=$-0.6, $\beta_4=0.6$ และ $\beta_5=$-0.8 (อิทธิพลจากปัจจัยฤดูกาลสูง)

รูปแบบที่ 2 $\beta_0=0.5$, $\beta_1=1$, $\beta_2=0.8$, $\beta_3=0.1$, $\beta_4=$ -0.1และ $\beta_5=$-0.2 (อิทธิพลจากปัจจัยแนวโน้มสูง)

รูปแบบที่ 3 $\beta_0=0.5$, $\beta_1=1$, $\beta_2=0.5$, $\beta_3=0.5$, $\beta_4=$ -0.1และ $\beta_5=$-0.3(อิทธิพลจากทั้งปัจจัยแนวโน้มและปัจจัยฤดูกาลระดับปานกลาง)

การวิจัยครั้งนี้ใช้วิธีการประมาณค่าสูญหายทั้ง 4 วิธีกับข้อมูลทั้ง 2 กรณี ภายใต้ขอบเขตการวิจัยที่เหมือนกันดังนี้

1. กำหนดค่าความคลาดเคลื่อนของข้อมูลเป็นตัวแปรสุ่มมีการแจกแจงแบบปกติ $(\varepsilon_i \sim N(0,\sigma^2))$ เมื่อ $\sigma=5$,10 ,15 , 20 และ25 และสำหรับข้อมูลอนุกรมเวลา $Cov(\varepsilon_i,\varepsilon_j)=0$ เมื่อ $i \neq j$

2. ขนาดตัวอย่าง(n)ที่ใช้ในการศึกษาเท่ากับ 50 , 100 และ 200

3. กำหนดการสูญหายของข้อมูลเกิดขึ้นกับตัวแปรตาม และเป็นการสูญหายแบบสุ่ม (Missing at Random (MAR)) ข้อมูลตัวแปรตามที่สูญหายคิดเป็นร้อยละ 5 , 10 และ 20

สำหรับเกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าสัมประสิทธิ์ความถดถอย ผู้วิจัยพิจารณาจากความผิดพลาดโดยเฉลี่ยของค่าพยากรณ์ของตัวแปรตามกับค่าจริง ในรูปแบบ Mean absolute percentage error (MAPE) ซึ่งวิธีการที่ให้ค่า MAPE ต่ำกว่าแสดงว่าเป็นวิธีการประมาณที่ดีกว่าซึ่งผู้วิจัยใช้ข้อมูลที่ได้จากการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โลด้วยโปรแกรม MATLAB

5.1 สรุปผลการวิจัย

5.1.1 ผลการเปรียบเทียบค่าความผิดพลาดโดยเฉลี่ยของค่าพยากรณ์ของตัวแปรตามกับค่าจริง (MAPE)

กรณีที่ 1 ความถดถอยเชิงเส้นพหุ เมื่อเป็นข้อมูลภาคตัดขวาง

เมื่อค่าสหสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระทั้ง 2 ตัวสูง(0.7,0.7)

เมื่อตัวอย่างมีขนาดเล็ก (50) วิธี RI และ EM เป็นวิธีที่มีค่า MAPE ต่ำสุด ทุกระดับของส่วนเบี่ยงเบนมาตรฐาน เมื่อตัวอย่างมีขนาดปานกลาง (100) ที่ส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับต่ำถึงค่อนข้างสูง (5-20) วิธี RI และ EM เป็นวิธีที่มีค่า MAPE ต่ำสุดเท่ากัน เมื่อตัวอย่างมีขนาดใหญ่ (200) ที่ส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับต่ำถึงค่อนข้างสูง (5-20) วิธี RI ,WNR และ EM เป็นวิธีที่มีค่า MAPE ต่ำสุดเท่ากัน เมื่อส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับสูง (25) ทั้งกรณีที่ตัวอย่างมีขนาดปานกลางและขนาดใหญ่ (100 – 200) วิธี WNR เป็นวิธีที่มีค่า MAPE ต่ำสุดเพียงวิธีเดียว เมื่อส่วนเบี่ยงเบนมาตรฐานเพิ่มขึ้นการกระจายของข้อมูลเพิ่มขึ้น ทำให้การแทนค่าสูญหายด้วยวิธี WNR คือ แทนค่าสูญหายด้วยหน่วยตัวอย่างที่คล้ายคลึงกัน จากวิธี NNI ร่วมกับการใช้สมการพยากรณ์จากวิธี RI ให้ผลดีกว่าวิธีอื่นๆ

จากการทดสอบความแตกต่างของค่าเฉลี่ยค่า MAPE กรณีที่ตัวอย่างมีขนาดเล็ก (50) ทุกระดับของส่วนเบี่ยงเบนมาตรฐาน เมื่อร้อยละการสูญหายของข้อมูลอยู่ในระดับปานกลางและระดับสูง (10 – 20) พบว่า 3 วิธีที่ให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย คือ วิธี RI WNR และ EM ที่ระดับนัยสำคัญ 0.05 เมื่อตัวอย่างมีขนาดปานกลาง (100) และส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับค่อนข้างสูงและระดับสูง (20 – 25) วิธีการประมาณค่าสูญหายทั้ง 4 วิธีให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ทุกระดับร้อยละการสูญหายของข้อมูล เมื่อตัวอย่างมีขนาดใหญ่(200) เกือบทุกกรณีวิธีการประมาณค่าสูญหายทั้ง 4 วิธีให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย

เมื่อค่าสหสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระตัวหนึ่งสูงมากและอีกตัวหนึ่งปานกลาง (0.9, 0.4)

กรณีที่ตัวอย่างมีขนาดเล็ก (50) ทุกระดับของส่วนเบี่ยงเบนมาตรฐาน วิธี RI และ EM เป็นวิธีที่มีค่า MAPE ต่ำสุด เมื่อตัวอย่างมีขนาดปานกลางและขนาดใหญ่ (100-200) ที่ส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับต่ำ (5) วิธี RI, WNR และ EM เป็นวิธีที่มีค่า MAPE ต่ำสุดเท่ากัน เมื่อส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับปานกลางถึงระดับสูง (15-25) ขนาดตัวอย่างไม่ส่งผลให้วิธีการประมาณค่าสูญหายที่มีค่า MAPE ต่ำสุดเปลี่ยนแปลง โดยวิธี RI และวิธีEM มีค่า MAPE ต่ำกว่าวิธีการประมาณค่าอื่นๆ ในกรณีนี้ค่าสหสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระตัวหนึ่งสูงอีก

ตัวหนึ่งปานกลาง ส่งผลให้วิธี NNI มีประสิทธิภาพในการประมาณค่าสูญหายลดลง การใช้สมการถดถดถอยในการประมาณค่าสูญหายเพียงวิธีเดียวจึงดีกว่าการใช้วิธี WNR

จากการทดสอบความแตกต่างของค่าเฉลี่ยค่า MAPE เมื่อตัวอย่างมีขนาดเล็ก (50) ทุกระดับของส่วนเบี่ยงเบนมาตรฐาน ที่ร้อยละการสูญหายของข้อมูลอยู่ในระดับปานกลางและระดับสูง (10 – 20) พบว่า 3 วิธีที่ให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย คือ วิธี RI WNR และ EM ที่ระดับนัยสำคัญ 0.05 เมื่อส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับค่อนข้างสูงและระดับสูง (20 – 25) ทั้งกรณีที่ตัวอย่างมีขนาดปานกลางและขนาดใหญ่ (100 – 200) วิธีการประมาณค่าสูญหายทั้ง 4 วิธีให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ทุกระดับร้อยละการสูญหายของข้อมูล

กรณีข้อมูลภาคตัดขวาง ปัจจัยที่มีผลกระทบทำให้วิธีการประมาณค่าสูญหายมีค่า MAPE ต่ำสุดแตกต่างกันที่เห็นได้ชัดเจนที่สุด คือ ส่วนเบี่ยงเบนมาตรฐาน เมื่อส่วนเบี่ยงเบนมาตฐานอยู่ในระดับสูงการกระจายของข้อมูลเพิ่มขึ้น ทำให้การแทนค่าสูญหายด้วยวิธี WNR คือ แทนค่าสูญหายด้วยหน่วยตัวอย่างที่คล้ายคลึงกัน จากวิธี NNI ร่วมกับการใช้สมการพยากรณ์จากวิธี RI ให้ผลดีกว่าวิธีอื่นๆ เพราะสมการถดถอยที่ส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับสูงอาจเกิดความคลาดเคลื่อน จึงต้องเพิ่มความแม่นยำด้วยการพิจารณาความคล้ายคลึงของข้อมูล โดยวิธี WNR ประมาณค่าสูญหายได้ดี เมื่อตัวแปรอิสระทั้ง 2 มีค่าสหสัมพันธ์กับตัวแปรตามอยู่ในระดับสูง เนื่องจากในกรณีที่ค่าสหสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระตัวหนึ่งสูงอีกตัวหนึ่งปานกลาง การแทนค่าข้อมูลสูญหายจากการพิจารณาความคล้ายคลึงของตัวแปรอิสระเพื่อนแทนค่าในตัวแปรตามนั้นอาจเกิดความไม่สอดคล้องกัน ส่งผลให้วิธี NNI มีประสิทธิภาพในการประมาณค่าสูญหายลดลง การใช้สมการถดถอยในการประมาณค่าสูญหายจึงดีกว่าการใช้วิธี WNR แม้ว่าส่วนเบี่ยงเบนมาตรฐานจะสูง

กรณีที่ 2 ความถดถอยเชิงเส้นพหุ เมื่อเป็นข้อมูลอนุกรมเวลา ที่ประกอบด้วยปัจจัยแนวโน้มและปัจจัยฤดูกาล

รูปแบบที่ 1 $\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 0.3$, $\beta_3 = -0.6$, $\beta_4 = 0.6$ และ $\beta_5 = -0.8$ (อิทธิพลจากปัจจัยฤดูกาลสูง)

กรณีที่ส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับต่ำ (5) ที่ตัวอย่างมีขนาดเล็ก (50) วิธี WNR เป็นวิธีที่มีค่า MAPE ต่ำสุด เมื่อตัวอย่างมีขนาดปานกลางและขนาดใหญ่ (100 และ 200) วิธี RI และ EM มีค่า MAPE ต่ำสุด เมื่อส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับค่อนข้างต่ำถึงสูง (10-25) เกือบทุกกรณีวิธี WNR เป็นวิธีที่มีค่า MAPE ต่ำสุด ยกเว้นเมื่อส่วนเบี่ยงเบนมาตรฐานสูง (25) ที่ตัวอย่างมี

ขนาดเล็กและปานกลาง (50 – 100) วิธี NNI เป็นวิธีที่มีค่า MAPE ต่ำสุด เนื่องจากข้อมูลมีการแกว่งขึ้นลงตามฤดูกาลสูง ทำให้ฤดูกาลเดียวกันข้อมูลมีความคล้ายคลึงกันวิธี WNR จึงประมาณค่าสูญหายได้ดี และเมื่อส่วนเบี่ยงเบนมาตรฐานสูงการกระจายของข้อมูลเพิ่มมากขึ้น วิธี NNI ซึ่งแทนค่าข้อมูลสูญหายด้วยค่าของหน่วยตัวอย่างที่คล้ายคลึงกัน จึงประมาณค่าสูญหายได้ดีขึ้น

จากการทดสอบความแตกต่างของค่าเฉลี่ยค่า MAPE ที่ระดับนัยสำคัญ 0.05 กรณีที่ส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับต่ำ (5) เมื่อร้อยละการสูญหายอยู่ในระดับสูง (20) ที่ตัวอย่างมีขนาดเล็ก (50) วิธี WNR เท่านั้นที่ไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย เมื่อตัวอย่างมีขนาดปานกลางและขนาดใหญ่(100 – 200) วิธี RI และ EM ให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย กรณีที่ส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับปานกลาง (15) เมื่อร้อยละการสูญหายอยู่ในระดับต่ำและปานกลาง (5 – 10) ที่ทุกระดับของขนาดตัวอย่าง เกือบทุกกรณี วิธีการประมาณค่าสูญหายทั้ง 4 วิธีให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย กรณีที่ส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับค่อนข้างสูงและระดับสูง (20 – 25) เกือบทุกกรณีวิธีการประมาณค่าสูญหายทั้ง 4 วิธีให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย

รูปแบบที่ 2 $\beta_0 = 0.5$, $\beta_1 = 1, \beta_2 = 0.8, \beta_3 = 0.1, \beta_4 =$ -0.1และ $\beta_5 =$ -0.2 (อิทธิพลจากปัจจัยแนวโน้มสูง)

กรณีที่ตัวอย่างมีขนาดเล็ก (50) ที่ส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับต่ำ (5) วิธี RI และ EM เป็นวิธีที่มีค่า MAPE ต่ำสุด เมื่อส่วนเบี่ยงเบนมาตรฐานเพิ่มสูงขึ้น (10 – 25) วิธี WNR เป็นวิธีที่มีค่า MAPE ต่ำสุด กรณีที่ตัวอย่างมีขนาดปานกลาง (100) ที่ส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับต่ำถึงปานกลาง (5-15) วิธี RI และ EM เป็นวิธีที่มีค่า MAPE ต่ำสุด เมื่อส่วนเบี่ยงเบนมาตรฐานเพิ่มขึ้นมาอยู่ในระดับสูง (20-25) วิธี WNR มีค่า MAPE ต่ำสุด กรณีที่ตัวอย่างมีขนาดใหญ่ (200) วิธี RI และ EM เป็นวิธีที่มีค่า MAPE ต่ำสุด ทุกระดับของส่วนเบี่ยงเบนมาตรฐาน เมื่อมีอิทธิพลจากปัจจัยแนวโน้มสูง วิธีRI และEM เป็นวิธีที่มีค่า MAPE ต่ำกว่าวิธีอื่นๆที่นำมาเปรียบเทียบ แต่เมื่อส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับสูง การกระจายของข้อมูลสูงขึ้นวิธี WNR จึงเป็นวิธีที่ดีกว่ายกเว้นกรณีที่ตัวอย่างมีขนาดใหญ่ซึ่งข้อมูลมีระยะยาวส่งผลให้วิธี RI และ EM ประมาณค่าสูญหายได้ดีกว่าวิธี WNR

จากการทดสอบความแตกต่างของค่าเฉลี่ยค่า MAPE กรณีที่ส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับต่ำ (5) เมื่อตัวอย่างมีขนาดปานกลางและขนาดใหญ่ (100 – 200) ที่ร้อยละการสูญหายอยู่ในระดับปานกลางและระดับสูง (10 – 20) วิธี RI และ EM ให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ที่ระดับนัยสำคัญ 0.05 กรณีที่ส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับปานกลาง (15) เกือบทุก

กรณี 3 วิธีที่ให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย คือ RI WNR และ EM ยกเว้นเมื่อตัวอย่างขนาดเล็ก (50) ที่ร้อยละการสูญหายอยู่ในระดับสูง (20) วิธี WNR เป็นวิธีที่ไม่แตกต่าจากกรณีที่ไม่มีข้อมูลสูญหายเพียงวิธีเดียว กรณีที่ส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับสูง (20 – 25) เมื่อตัวอย่างมีขนาดเล็ก (50) เกือบทุกกรณีวิธีการประมาณค่าสูญหายทั้ง 4 วิธีให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย เมื่อตัวอย่างมีปานกลางและขนาดใหญ่ (100 – 200) เกือบทุกกรณีพบว่า 3 วิธี คือ วิธี RI WNR และ EM ให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย

รูปแบบที่ 3 $\beta_0 = 0.5$, $\beta_1 = 1, \beta_2 = 0.5, \beta_3 = 0.5, \beta_4 = -0.1$ และ $\beta_5 = -0.3$ (อิทธิพลจากทั้งปัจจัยแนวโน้มและปัจจัยฤดูกาลระดับปานกลาง)

กรณีที่ตัวอย่างมีขนาดเล็กและปานกลาง (50 – 100) วิธี RI และ EM มีค่า MAPE ต่ำสุด เมื่อส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับต่ำ (5) เมื่อส่วนเบี่ยงเบนมาตรฐานเพิ่มสูงขึ้น (10 – 25) วิธี WNR จะเป็นวิธีที่มีค่า MAPE ต่ำกว่าวิธีอื่นที่นำมาเปรียบเทียบ และที่ตัวอย่างมีขนาดเล็ก (50) เมื่อส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับสูง (25) วิธี NNI มีค่า MAPE ต่ำสุด กรณีที่ตัวอย่างมีขนาดใหญ่ (200) วิธี RI และ EM เป็นวิธีที่มีค่า MAPE ต่ำที่สุดทุกระดับส่วนเบี่ยงเบนมาตรฐาน

จากการทดสอบความแตกต่างของค่าเฉลี่ยค่า MAPE เมื่อส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับต่ำ (5) ที่ร้อยละการสูญหายอยู่ในระดับสูง (20) วิธี RI และ EM ไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย เมื่อส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับปานกลาง (15) เกือบทุกกรณี 3 วิธี คือ วิธี RI WNR และ EM ให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย เมื่อส่วนเบี่ยงเบนมาตรฐานสูง (20 – 25) ที่ตัวอย่างมีขนาดเล็กและปานกลาง (50 -100) เกือบทุกกรณีวิธีการประมาณค่าสูญหายทั้ง 4 วิธีให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย ที่ตัวอย่างขนาดใหญ่ (200) 3 วิธี คือ วิธี RI WNR และ EM ให้ผลไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย

กรณีข้อมูลอนุกรมเวลาเมื่อข้อมูลได้รับอิทธิพลจากปัจจัยทั้ง 2 ในระดับที่แตกต่างกัน สำหรับข้อมูลที่ได้รับอิทธิพลจากปัจจัยฤกาลสูง ข้อมูลมีการแกว่งขึ้นลงตามฤดูกาลสูง วิธี WNR จึงประมาณค่าสูญหายได้ดี เพราะต้องอาศัยการพิจารณาความคล้ายคลึงของข้อมูลที่มีฤดูกาลเดียวกันเพื่อเพิ่มความแม่นยำจากวิธี NNI แต่ข้อมูลยังคงมีอิทธิพลจากปัจจัยแนวโน้มในระดับต่ำจึงต้องใช้วิธี RI ร่วมกันในการประมาณค่า และเมื่อส่วนเบี่ยงเบนมาตรฐานสูงการกระจายของข้อมูลเพิ่มมากขึ้นสมการถดถอยอาจเกิดความคลาดเคลื่อนมากขึ้นสำหรับข้อมูลระยะสั้น วิธี NNI ซึ่งแทนค่าข้อมูลสูญหายด้วยค่าของหน่วยตัวอย่างที่คล้ายคลึงกัน จึงประมาณค่าสูญหายได้ดีกว่าวิธี WNR ที่พิจารณาสมการถดถอยร่วมกับการพิจารณาความคล้ายคลึง สำหรับปัจจัยแนวโน้มของข้อมูลมี

ลักษณะเชิงเส้น เนื่องจากมีความสัมพันธ์ระหว่างข้อมูลกับเวลา เมื่อมีอิทธิพลจากปัจจัยแนวโน้มสูง ส่งผลให้ วิธีRI และEM เป็นวิธีที่มีค่า MAPE ต่ำกว่าวิธีอื่นๆที่นำมาเปรียบเทียบ แต่เมื่อส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับสูง การกระจายของข้อมูลสูงขึ้นวิธี WNR จึงเป็นวิธีที่ดีกว่า ยกเว้นกรณีที่ตัวอย่างมีขนาดใหญ่ซึ่งข้อมูลมีระยะยาวส่งผลให้วิธี RI และ EM ประมาณค่าสูญหายได้ดีกว่าวิธี WNR เพราะเมื่อข้อมูลมีระยะสั้นจะถูกรบกวนจากปัจจัยฤดูกาลมากกว่าข้อมูลระยะยาว กรณีที่ข้อมูลมีทั้งอิทธิพลจากทั้งปัจจัยแนวโน้มและปัจจัยฤดูกาลในระดับปานกลาง อิทธิพลจากปัจจัยฤดูกาลจะเห็นได้ชัดเจนในข้อมูลระยะสั้น (ตัวอย่างขนาดเล็ก) เห็นได้จากสำหรับข้อมูลระยะสั้น วิธี WNR จะมีค่า MAPE ต่ำ และวิธี NNI มีค่า MAPE ต่ำเมื่อส่วนเบี่ยงเบนมาตรฐานเพิ่มขึ้นมาอยู่ในระดับสูง สำหรับอิทธิพลจากปัจจัยแนวโน้มจะเห็นได้ชัดเจนในข้อมูลระยะยาว (ตัวอย่างขนาดใหญ่) โดยวิธี RI และ EM เป็นวิธีที่มีค่า MAPE ต่ำสุดทุกกรณี

5.1.2 สรุปความแตกต่างของวิธีการประมาณค่าสูญหายที่นำมาพิจารณาเปรียบเทียบ

วิธี WNR เป็นวิธีประมาณค่าสูญหายที่ดีกว่าวิธีการประมาณค่าอื่นๆที่นำมาเปรียบเทียบสำหรับข้อมูลภาคตัดขวางที่ค่าสหสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระทั้งสองอยู่ในระดับสูง เมื่อส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับสูง และข้อมูลอนุกรมเวลาที่ได้รับอิทธิพลจากปัจจัยฤดูกาลสูง

วิธี NNI เป็นวิธีประมาณค่าสูญหายที่ดีกว่าวิธีอื่นที่นำมาเปรียบเทียบสำหรับข้อมูลอนุกรมเวลาที่ได้รับอิทธิพลจากปัจจัยฤดูกาลสูง เมื่อส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับสูง

วิธี RI เป็นวิธีประมาณค่าสูญหายที่ดีกว่าวิธีอื่นที่นำมาเปรียบเทียบสำหรับข้อมูลภาคตัดขวางที่ค่าสหสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระตัวหนึ่งสูงมากและอีกตัวหนึ่งปานกลาง สำหรับข้อมูลอนุกรมเวลาจะเหมาะสมกับข้อมูลที่ได้รับอิทธิพลจากปัจจัยแนวโน้มสูง โดยพบว่าทุกกรณีที่วิธี RI เป็นวิธีที่มีค่า MAPE ต่ำสุดนั้นวิธี EM จะมีค่า MAPE ต่ำสุดเท่ากันกับวิธี RI จากวิธี EM ที่เป็นวิธีที่ประมาณค่าสูญหายที่ประมาณค่าจากกระบวนการวนซ้ำเพื่อค้นหาค่าประมาณภาวะน่าจะเป็นสูงสุด (Maximum Likelihood) ของค่าพารามิเตอร์ พบว่าพารามิเตอร์ที่ประมาณค่าได้ในแต่ละรอบมีค่าแตกต่างกันน้อยมาก โดยจะแตกต่างกันที่ทศนิยมหลักที่สิบขึ้นไป ทำให้ค่าพารามิเตอร์ที่นำมาใช้ในการประมาณค่าสูญหายนั้นมีความแตกต่างจากค่าพารา – มิเตอร์จากวิธี RI ตั้งแต่ทศนิยมหลักที่สิบขึ้นไปเช่นกัน เพราะค่าประมาณของพารามิเตอร์จากวิธี RI คือค่าประมาณพารามิเตอร์เริ่มต้นในวิธี EM จึงส่งผลให้ผลลัพธ์ที่ปรากฏทั้งของวิธี RI และ วิธี EM ไม่แตกต่างกันทุกกรณี และจากการศึกษาผลจากงานวิจัยที่ผ่านมาของ วารุณี ตรีบำรุงศักดิ์ (2537) เกี่ยวกับการประมาณค่าสูญหายที่มีการนำวิธี RI และวิธี EM มาเปรียบเทียบกันนั้นพบว่า

วิธี RI และวิธี EM จะมีความแตกต่างกันเมื่อร้อยละการสูญหายของข้อมูลอยู่ในระดับสูงถึงร้อยละ 60 – 70 โดยขนาดตัวอย่างที่ทำการศึกษามีค่าระหว่าง 10 – 70 ในทางปฏิบัติไม่ควรนำข้อมูลที่มีขนาดเล็กและค่าสูญหายมากถึงร้อยละ 60 – 70 มาใช้ในการวิเคราะห์ข้อมูลทางสถิติ

5.1.3 ปัจจัยที่มีผลต่อค่าความผิดพลาดโดยเฉลี่ยของค่าพยากรณ์ของตัวแปรตามกับค่าจริง (MAPE)

กรณีที่ 1 ความถดถอยเชิงเส้นพหุ เมื่อเป็นข้อมูลภาคตัดขวาง

ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อน ε_i (σ)

เมื่อส่วนเบี่ยงเบนมาตรฐานเพิ่มขึ้นข้อมูลมีประสิทธิภาพในการพยากรณ์ลดลงทุกกรณี (ค่า MAPE เพิ่มขึ้น) เนื่องจากสมการการถดถอยเชิงเส้นสร้างขึ้นจากข้อมูลตัวแปรอิสระและตัวแปรตามที่เก็บรวบรวมไว้ ซึ่งตัวแปรทั้งสองประเภทนี้มีความสัมพันธ์กันในลักษณะใดลักษณะหนึ่ง เมื่อส่วนเบี่ยงเบนมาตรฐานมีค่าสูงขึ้นสมการที่ได้มีความสามารถในการอธิบายความผันแปรลดลง จึงส่งผลโดยตรงต่อความคลาดเคลื่อนในการพยากรณ์

ขนาดตัวอย่าง (n)

เมื่อขนาดตัวอย่างเพิ่มขึ้นส่งผลให้ค่า MAPE มีแนวโน้มลดลงเพราะขนาดตัวอย่างที่เพิ่มขึ้นจะส่งผลให้ความคลาดเคลื่อนในการพยากรณ์ลดลง แต่เป็นการลดลงเพียงเล็กน้อยโดยต่างกันน้อยกว่า 0.1 เนื่องจากข้อมูลภาคตัดขวางไม่มีผลกระทบจากปัจจัยแนวโน้มและฤดูกาล ทำให้แม้ตัวอย่างมีขนาดเล็กความคลาดเคลื่อนในการพยากรณ์มีค่าไม่สูงนัก เมื่อขนาดตัวอย่างเพิ่มขึ้นความคลาดเคลื่อนจึงลดลงเพียงเล็กน้อย

ร้อยละการสูญหายของข้อมูลตัวแปรตาม

ที่ตัวอย่างขนาดเดียวกัน เมื่อร้อยละการสูญหายของข้อมูลเพิ่มขึ้นส่งผลให้ค่า MAPE จากวิธีการประมาณค่าสูญหายทั้ง 4 วิธี มีผลต่างจากกรณีที่ไม่มีข้อมูลสูญหายเพิ่มขึ้น

ค่าสหสัมพันธ์ระหว่าง y_i กับ x_{ji} (ρ_j)

กรณีที่ค่าสหสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระทั้ง 2 ตัวสูง(0.7,0.7) มีค่า MAPE ต่ำกว่า กรณีที่ค่าสหสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระตัวหนึ่งสูงมากและอีกตัวหนึ่งปานกลาง (0.9, 0.4) เนื่องจากค่าสหสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระตัวหนึ่งที่อยู่ในระดับปานกลางส่งผลให้ความสามารถในการอธิบายความผันแปรได้น้อยกว่ากรณีที่ค่าสหสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระทั้ง 2 ตัวสูง

กรณีที่ 2 ความถดถอยเชิงเส้นพหุ เมื่อเป็นข้อมูลอนุกรมเวลา ที่ประกอบด้วยปัจจัยแนวโน้มและปัจจัยฤดูกาล

ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อน ε_i (σ)

เมื่อส่วนเบี่ยงเบนมาตรฐานเพิ่มขึ้นข้อมูลมีประสิทธิภาพในการพยากรณ์ลดลงทุกกรณี (ค่า MAPE เพิ่มขึ้น) เนื่องจากเมื่อส่วนเบี่ยงเบนมาตรฐานมีค่าสูงขึ้นความสามารถในการอธิบายความผันแปรลดลง จึงส่งผลโดยตรงต่อความคลาดเคลื่อนในการพยากรณ์ โดยผลของค่า MAPE จากวิธีการประมาณค่าสูญหายทั้ง 4 วิธีที่ส่วนเบี่ยงเบนมาตรฐานระดับต่างๆจะสอดคล้องกับรูปแบบของข้อมูล

ขนาดตัวอย่าง (n)

เมื่อขนาดตัวอย่างเพิ่มขึ้นส่งผลให้ค่า MAPE มีแนวโน้มลดลงเพราะขนาดตัวอย่างที่เพิ่มขึ้นจะส่งผลให้ความคลาดเคลื่อนในการพยากรณ์ลดลง ซึ่งเห็นได้ชัดเจนกว่ากรณีข้อมูลภาคตัดขวาง เพราะเมื่อข้อมูลมีระยะสั้นจะถูกรบกวนจากปัจจัยฤดูกาลมากกว่าข้อมูลระยะยาว

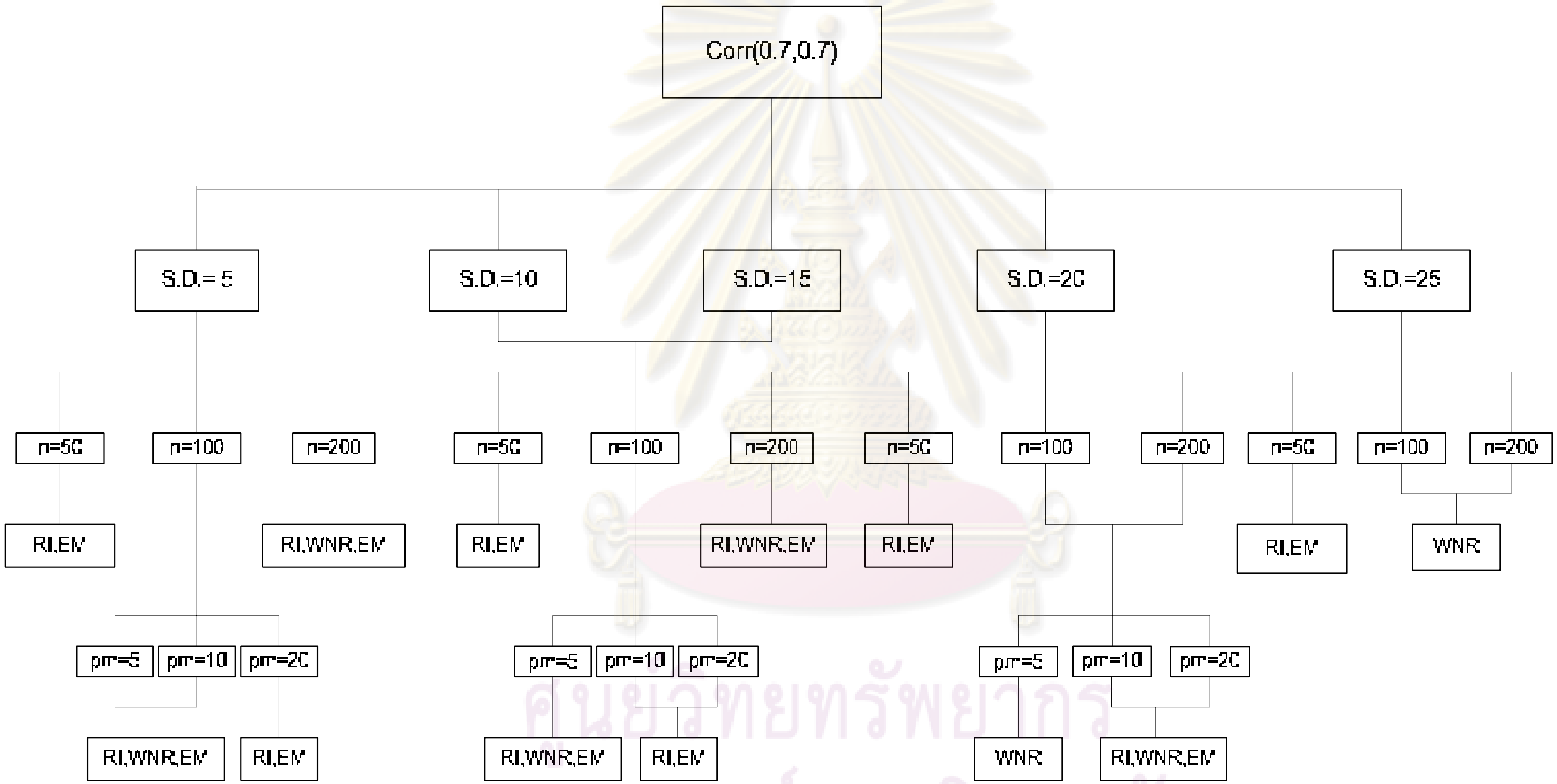
ร้อยละการสูญหายของข้อมูลตัวแปรตาม

ที่ตัวอย่างขนาดเดียวกัน เมื่อร้อยละการสูญหายของข้อมูลเพิ่มขึ้นส่งผลให้ค่า MAPE จากวิธีการประมาณค่าสูญหายทั้ง 4 วิธี มีผลต่างจากกรณีที่ไม่มีข้อมูลสูญหายเพิ่มขึ้น เช่นเดียวกัน

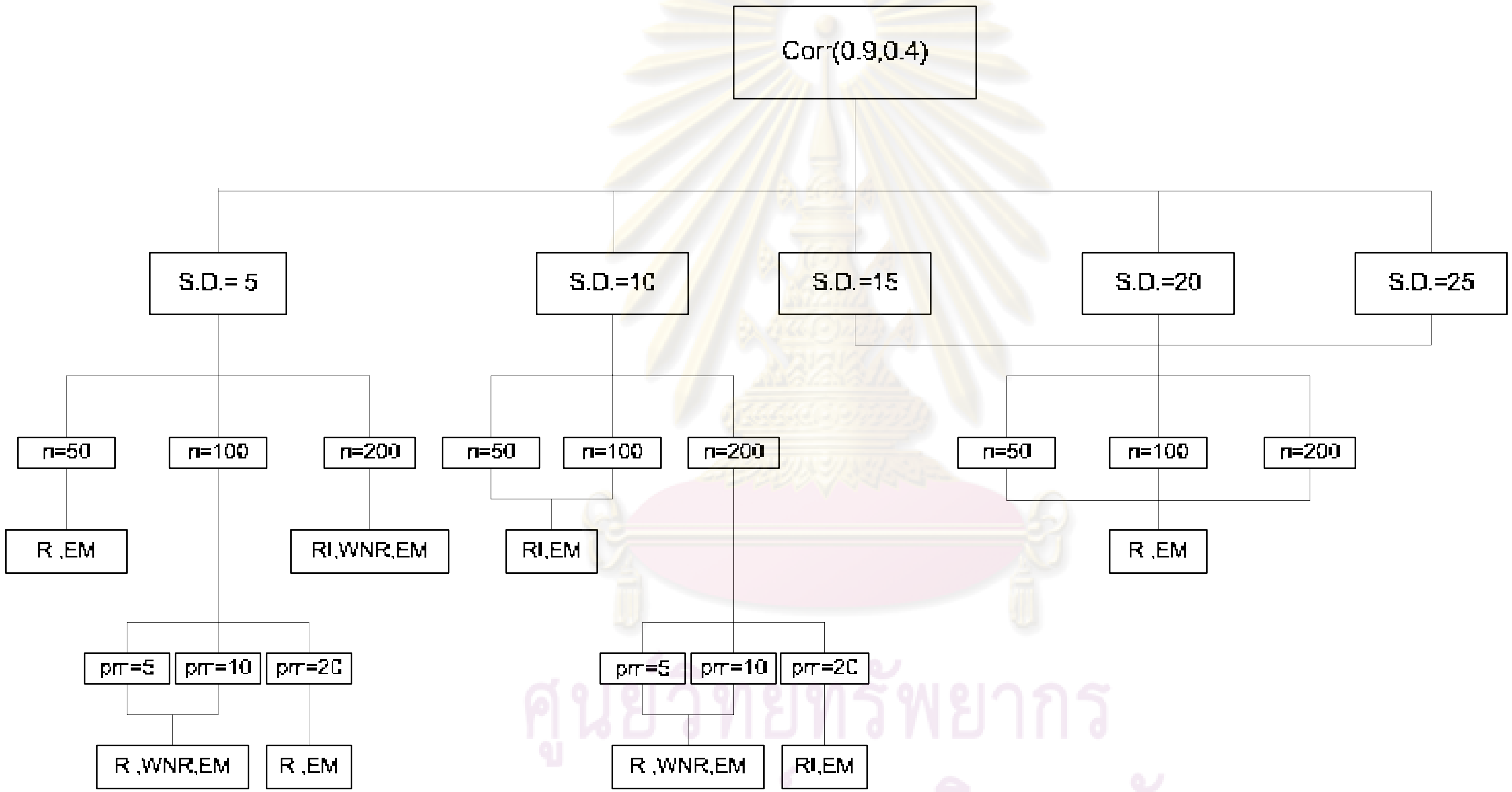
สัมประสิทธิ์การถดถอย

สัมประสิทธิ์การถดถอยที่กำหนดขึ้นมีจุดประสงค์เพื่อให้ข้อมูลที่สร้างขึ้นได้รับอิทธิพลจากปัจจัยแนวโน้มและปัจจัยฤดูการแตกต่างกัน ส่งผลให้ค่า MAPE เมื่อข้อมูลที่มีอิทธิพลจากปัจจัยแนวโน้มสูง มีค่าต่ำกว่า กรณีที่ข้อมูลมีอิทธิพลจากทั้งปัจจัยแนวโน้มและปัจจัยฤดูกาลระดับปานกลาง และกรณีที่ข้อมูลมีอิทธิพลจากปัจจัยฤดูกาลสูง ตามลำดับ เนื่องจากข้อมูลที่มีอิทธิพลจากปัจจัยแนวโน้มสูง สมการมีรูปแบบเชิงเส้นชัดเจนกว่ารูปแบบอื่นๆ ความคลาดเคลื่อนในการพยากรณ์จึงต่ำกว่ากรณีที่มีปัจจัยฤดูกาลสูงซึ่งมีการแกว่งของข้อมูลตามฤดูกาล

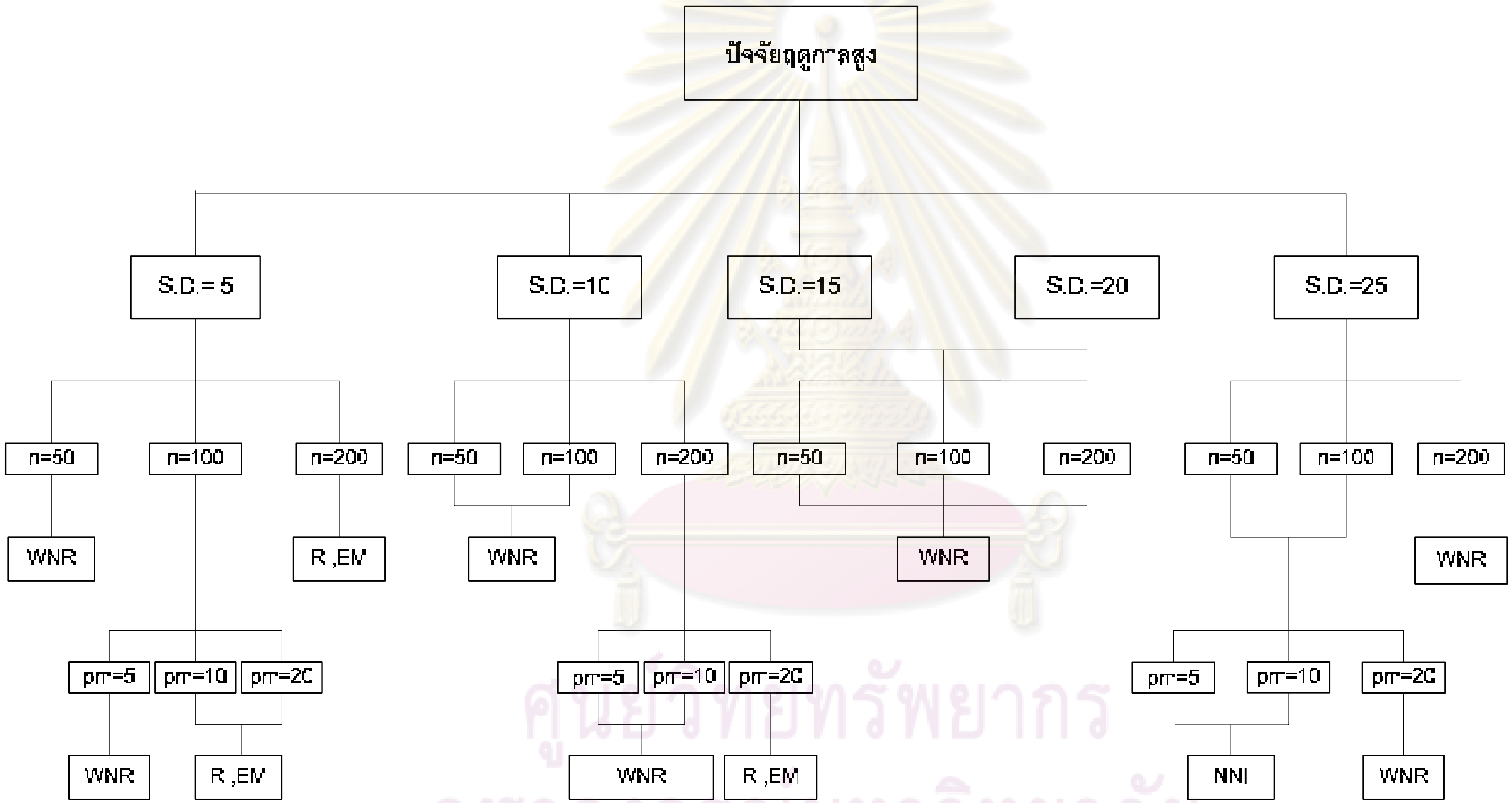
รูปที่ 5.1 แผนผังแสดงวิธีที่มีค่า MAPE ต่ำสุดเมื่อเป็นข้อมูลภาคตัดขวางเมื่อค่าสหสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระทั้ง 2 ตัวสูง (0.7,0.7)



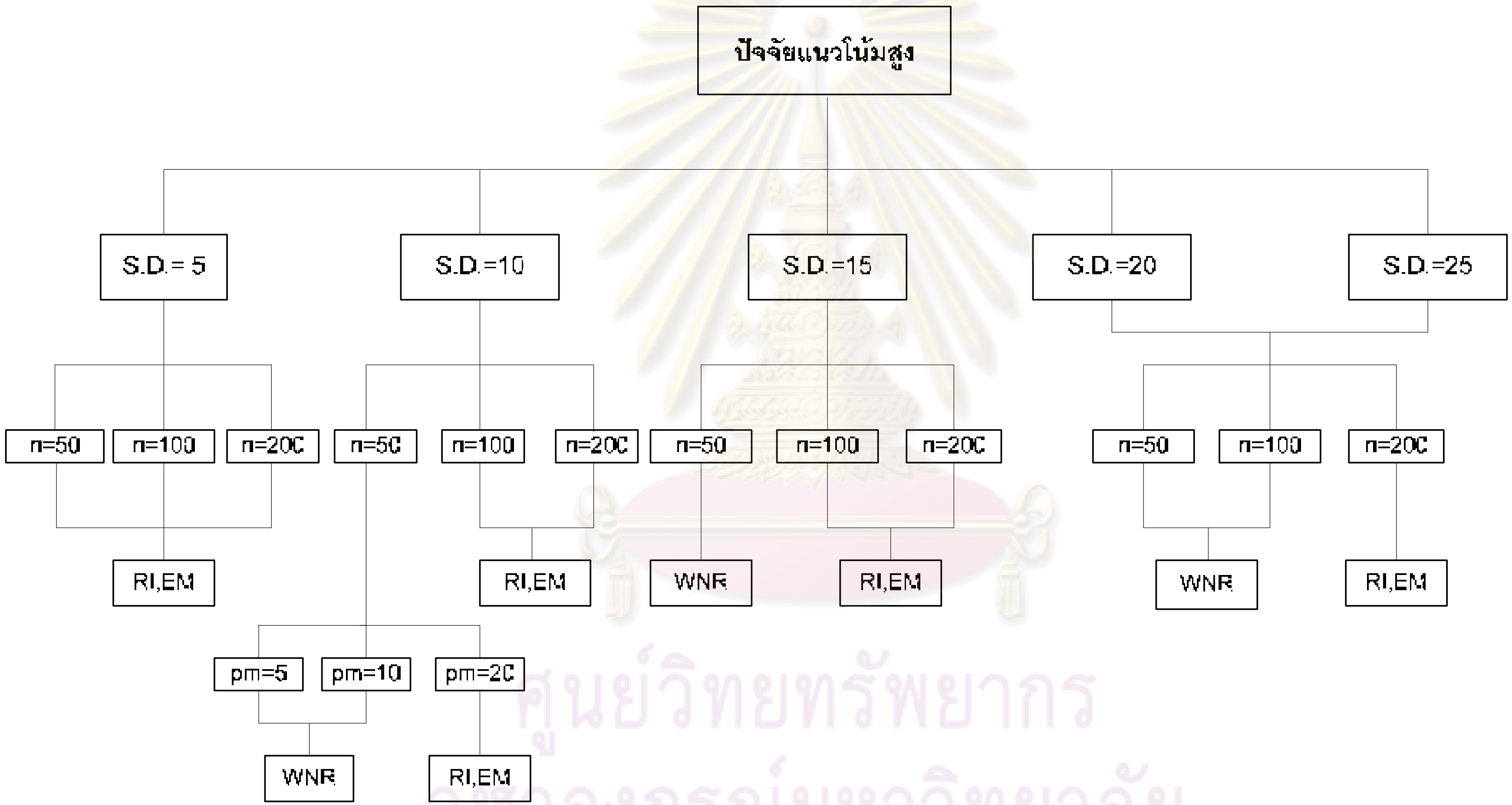
รูปที่ 5.2 แผนผังแสดงวิธีที่มีค่า MAPE ต่ำสุดเมื่อเป็นข้อมูลภาคตัดขวางเมื่อค่าสหสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระตัวหนึ่งสูงมากและอีกตัวหนึ่งปานกลาง (0.9, 0.4)



รูปที่ 5.3 แผนผังแสดงวิธีที่มีค่า MAPE ต่ำสุดเมื่อเป็นข้อมูลอนุกรมเวลารูปแบบที่ 1 $\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 0.3$, $\beta_3 = -0.6$, $\beta_4 = 0.6$ และ $\beta_5 = -0.8$ (อิทธิพลจากปัจจัยฤดูกาลสูง)

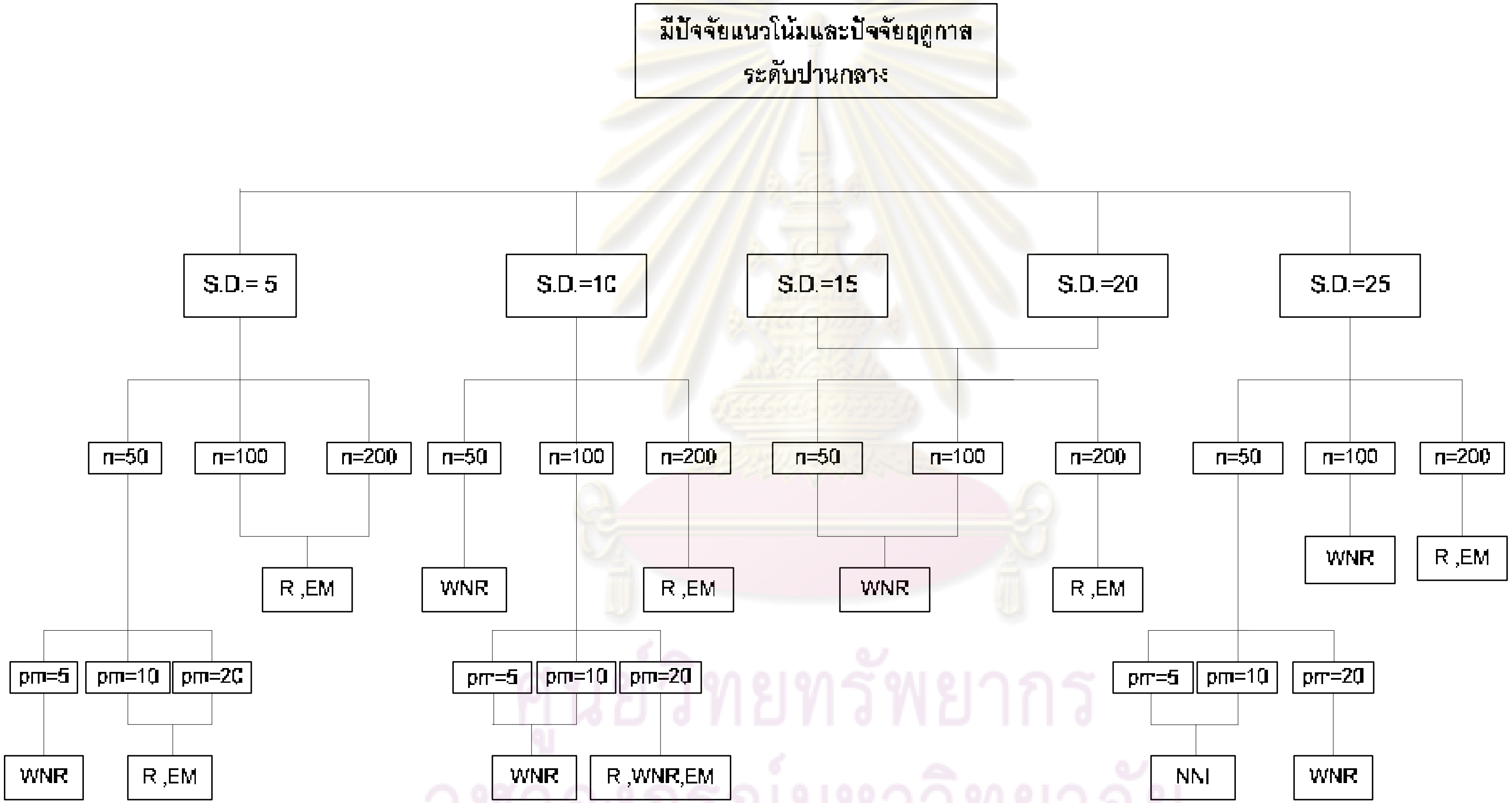


รูปที่ 5.4 แผนผังแสดงวิธีที่มีค่า MAPE ต่ำสุดเมื่อเป็นข้อมูลอนุกรมเวลารูปแบบที่ 2 $\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 0.8$, $\beta_3 = 0.1$, $\beta_4 =$ -0.1และ $\beta_5 =$ -0.2 (อิทธิพลจากปัจจัยแนวโน้มสูง)



77

รูปที่ 5.5 แผนผังแสดงวิธีที่มีค่า MAPE ต่ำสุดเมื่อเป็นข้อมูลอนุกรมเวลารูปแบบที่ 3 $\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 0.5$, $\beta_3 = 0.5$, $\beta_4 =$ -0.1และ $\beta_5 =$ -0.3 (อิทธิพลจากปัจจัยแนวโน้มและปัจจัยฤดูกาลระดับปานกลาง)



5.2 ข้อเสนอแนะ

5.2.1 ด้านการนำไปใช้

จากผลการวิจัยจะเห็นได้ว่าทุกปัจจัยส่งผลให้ค่า MAPE จากวิธีการประมาณค่าสูญหายทั้ง 4 วิธีมีค่าต่ำที่สุดแตกต่างกันซึ่งผู้ใช้สามารถนำไปตัดสินใจเลือกวิธีการประมาณค่าสูญหายได้ตามกรณีต่างๆที่พบ แต่จากการทดสอบความแตกต่างของค่าเฉลี่ยค่า MAPE กรณีที่วิธีการประมาณค่าสูญหายทั้ง 4 วิธีไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย และให้ผลไม่แตกต่างกัน ผู้ใช้จึงสามารถเลือกวิธีการที่ง่ายและเงื่อนไขน้อยที่สุดในการนำไปใช้คือวิธี NNI และในกรณีที่วิธีการประมาณค่าสูญหาย 3 วิธีที่ไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหาย และให้ผลไม่แตกต่างกัน คือ วิธี RI WNR แล EM ในกรณีนี้วิธี RI เป็นวิธีที่ขั้นตอนน้อยที่สุด มีเพียงบางกรณีเท่านั้นที่วิธี RI แล EM (2 วิธีนี้ให้ผลไม่แตกต่างกันทุกกรณี) หรือวิธี WNR เป็นเพียงวิธีเดียวที่ไม่แตกต่างจากกรณีที่ไม่มีข้อมูลสูญหายจึงควรเลือกใช้ตามความเหมาะสมของข้อมูล

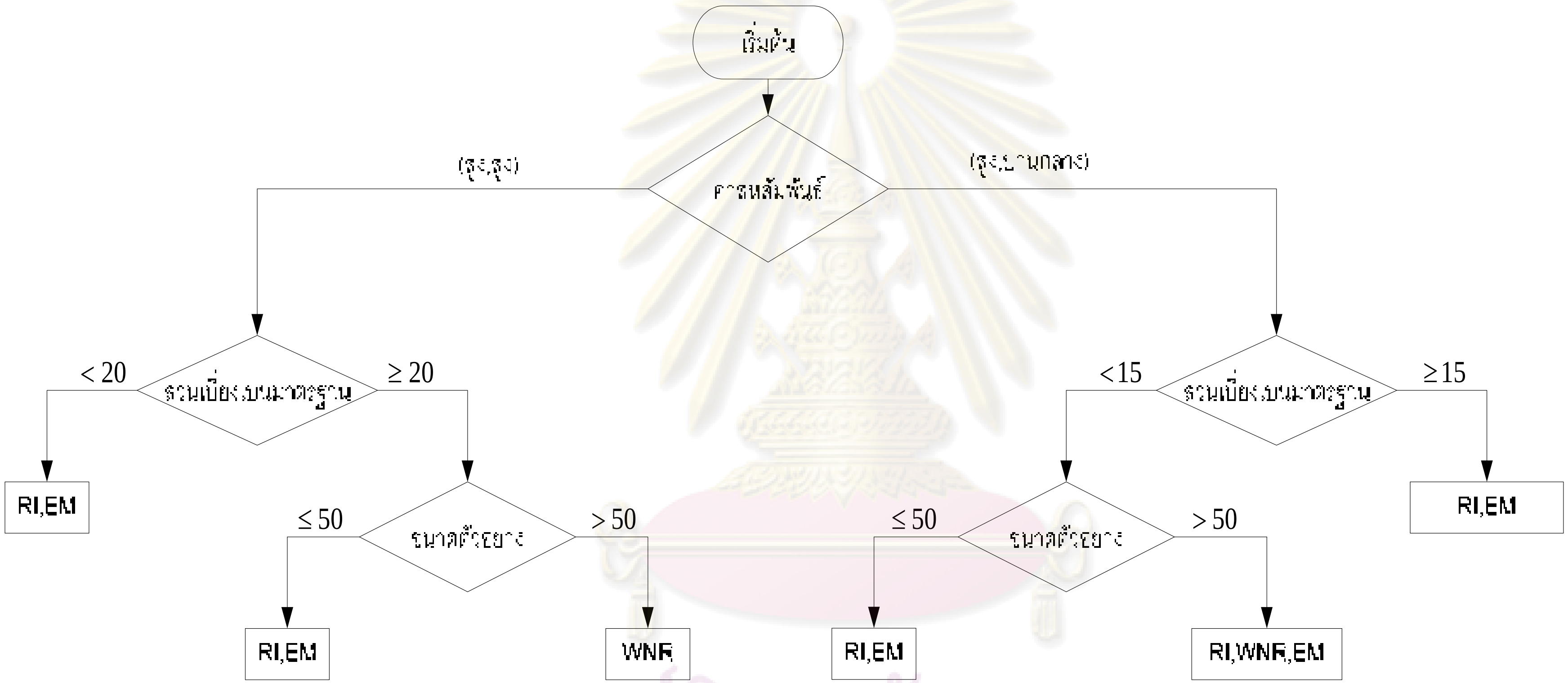
5.2.2 ด้านการศึกษาวิจัย

5.2.2.1 สำหรับการวิจัยครั้งนี้ผู้วิจัยได้ศึกษาเฉพาะกรณีที่ตัวแปรอิสระ x_i มีการแจกแจงแบบปกติ สำหรับการวิจัยในครั้งต่อไปอาจทำการศึกษากรณีที่ตัวแปรอิสระมีการแจกแจงแบบอื่นเพื่อศึกษาถึงผลกระทบของตัวแปรอิสระที่มีต่อวิธีการประมาณค่าสูญหาย เนื่องจากวิธีการประมาณค่าสูญหายอาศัยข้อมูลตัวแปรอิสระในการประมาณค่า

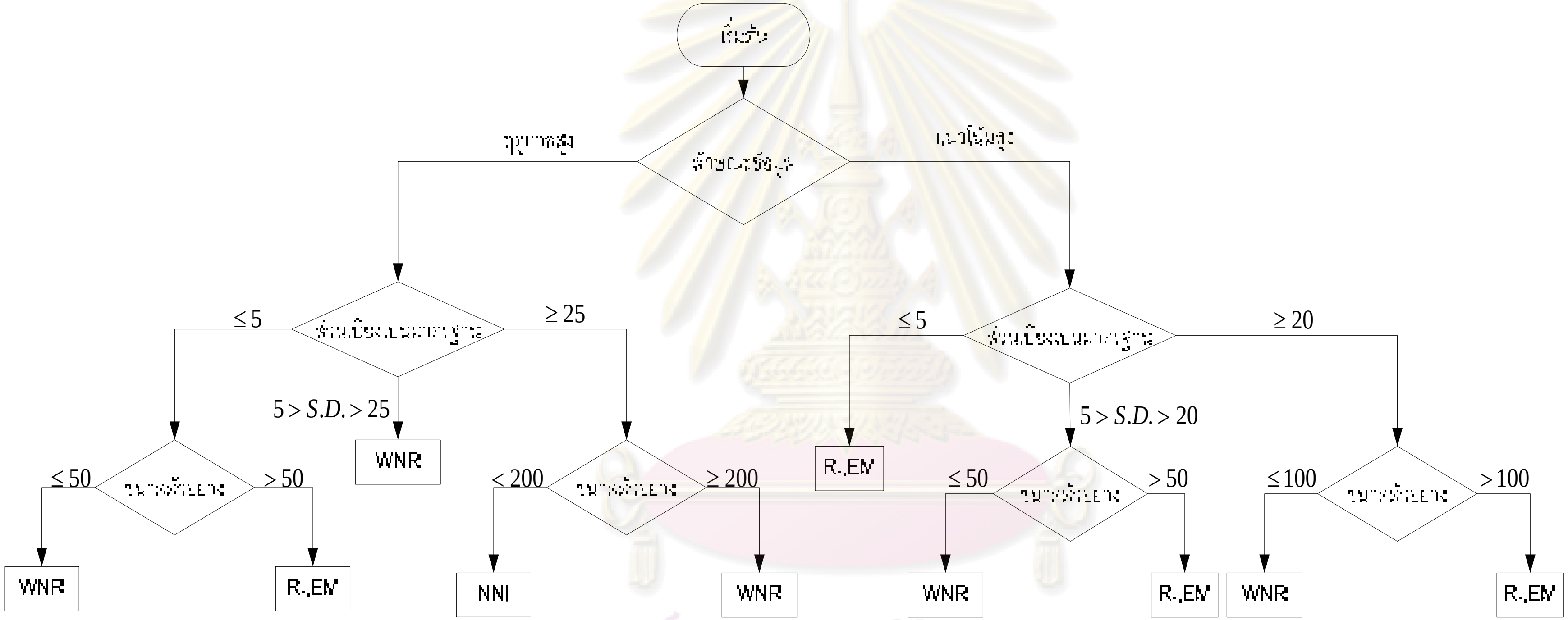
5.2.2.2 ในการวิจัยครั้งนี้ได้ทำการศึกษาวิธีการประมาณค่าสูญหายเปรียบเทียบกันระหว่างข้อมูลภาคตัดขวางและข้อมูลอนุกรมเวลาซึ่งสำหรับข้อมูลทั้ง 2 ลักษณะอาจยังมีวิธีอื่นๆที่เหมาะสมแตกต่างกันในงานวิจัยครั้งต่อไปอาจจะทำการศึกษาวิธีการประมาณค่าสูญหายด้วยวิธีอื่นๆเพิ่มเติมตามลักษณะข้อมูลที่แตกต่างกัน

5.2.2.3 งานวิจัยครั้งนี้ศึกษาเฉพาะกรณีการวิเคราะห์การถดถอยเชิงเส้น ในการวิจัยครั้งต่อไปอาจทำการศึกษาสมการถดถอยรูปแบบอื่นๆเช่นในกรณีที่ไม่ใช่เชิงเส้น

รูปที่ 5.6 แผนผังแสดงการเลือกใช้วิธีการประมาณค่าสูญหายในการวิเคราะห์การถดถอยเชิงเส้นพหุเมื่อเป็นข้อมูลภาคตัดขวางในทางปฏิบัติ



รูปที่ 5.7 แผนผังแสดงการเลือกใช้วิธีการประมาณค่าสูญหายในการวิเคราะห์การถดถอยเชิงเส้นพหุเมื่อเป็นข้อมูลอนุกรมเวลาในทางปฏิบัติ



81

รายการอ้างอิง

ภาษาไทย

กุศยา ปลั่งพงษ์พันธุ์. วิธีพยากรณ์ทางสถิติ. เอกสารประกอบการสอน.นครปฐม: ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร พระราชวังสนามจันทร์, 2547.

ชุติมา ชัยมุสิก. การวิเคราะห์การถดถอยเชิงซ้อนเมื่อข้อมูลของตัวแปรอิสระสูญหาย. วิทยานิพนธ์ ปริญญามหาบัณฑิต ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย, 2533.

ธีระพร วีระถาวร. ตัวแบบเชิงเส้นทฤษฎีและการประยุกต์. กรุงเทพมหานคร: สำนักพิมพ์วิทยพัฒน์, 2541.

วารุณี ตรีบำรุงศักดิ์. การพยากรณ์ด้วยวิธีการวิเคราะห์การถดถอยเชิงเส้นพหุเมื่อตัวแปรตามมีค่าสูญหาย. วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย, 2537.

สุพล ดุรงค์วัฒนา. การวิเคราะห์ขั้นสูง. กรุงเทพมหานคร: ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย, 2545.

ภาษาอังกฤษ

James, H. and Gray, K. What to do about Missing Values in Time Series Cross-Section Data.[Online]. Available from: http://gking.harvard.edu/amelia., [21/11/2007].

Jiahua, Ch. and Jun, Sh. Jackknife Variance Estimation for Nearest – Neighbor Imputation. Journal of the American Statistical Association . 96,453 (2001):pp.260 – 269.

Roderick J.A. L. and Donald B. R. Statistical Analysis with Missing Data. Wiley Series in Probability and Mathematical Statistic. Applied Probability and Statistics. New York: John Wiley and Sons,1987.

Chaimongkol, W., and Suwattee,P. The Effect of Weighted Nearest Neighbor – Regression Imputation on Simple Linear Regression Analysis: Proceedings of the symposium. School of Applied Statistics,National Institute of Development Administration (NIDA), 2005.

Chaimongkol, W. Three Composite Imputation Methods for Item Nonresponse Estimation in Sample Surveys. Doctoral dissertation, Program in Statistics. School of Applied Statistics , National Institute of Development Administration (NIDA), 2005.

ภาคผนวก

ภาคผนวก ก

กรณีที่ 1ข้อมูลภาคตัดขวาง

โปรแกรมหลัก (การกำหนดค่า การเรียกใช้ function)

```
n =50;
N = 1000;
L = 1000;

LAPEComp=0;
LAPERI=0;
LAPENNI=0;
LAPEWRI=0;
LAPEEM=0;

stdevX1 = 10;
stdevX2 = 10;
stdevy = 5;

varX1 = (stdevX1).^2;
varX2 = (stdevX2).^2;
vary = (stdevy).^2;

CorrX1y = 0.9;
CorrX2y = 0.4;
CovX1y = stdevX1*stdevy*CorrX1y;
CovX2y = stdevX2*stdevy*CorrX2y;
SigmaXy = [varX1 0 CovX1y;0 varX2 CovX2y;CovX1y CovX2y vary]
CholSigmaXy = chol(SigmaXy)
CSigmaXy = CholSigmaXy'

while N > 0
  APEComp=0;
  APERI=0;
  APENNI=0;
  APEWRI=0;
  APEEM=0;
 [X,y,yobs,Xobs,ymis,Xmis,Nobs,NOMis]=Gencase1_1(n,CSigmaXy)
```

กรณีข้อมูลสมบูรณ์

```
[yhatcomp,APEComp] = Compcase1(y,X,n)
APECompL(N,1) = APEComp
```

วิธี RI

```
[yRI,yNewRI,APERI] = RIcase1(y,X,Xobs,yobs,Xmis,n)
APERIL(N,1) = APERI
```

วิธีNNI

```
[ED,minED,EDpost,yNNI,DiffNNI,DiffRNNI,APENNI]=NNIcase1(y,X,ymis,yobs,Xobs,Xmis,No
bs,NOMis,n)
APENNIL(N,1) = APENNI
```

วิธีWRI

```
[yWRI,APEWRI] = WRIcase1(y,X,yRI,yNNI,Xmis,Xobs,yobs,n,NOMis,vary)
APEWRIL(N,1) = APEWRI
```

วิธีEM

```
[Xnew,yhat,y_after,b_before,b_after,yEM,APEEM] = EMcase1New(X,y,Xobs,yobs,Xmis,n)
APEEML(N,1) = APEEM
```

เกณฑ์ MAPE เก็บสะสมแต่ละรอบ

```
LAPEComp = LAPEComp + APEComp;
LAPERI = LAPERI + APERI;
LAPENNI = LAPENNI + APENNI;
LAPEWRI = LAPEWRI + APEWRI;
LAPEEM = LAPEEM + APEEM;
N=N-1
end
end
Std_APEComp = std(APECompL(:,1))
Std_APERI = std(APERIL(:,1))
Std_APENNI = std(APENNIL(:,1))
Std_APEWRI = std(APEWRIL(:,1))
Std_APEEM = std(APEEML(:,1))
```

```
MAPEComp = LAPEComp/L
MAPERI = LAPERI/L
MAPENNI = LAPENNI/L
MAPEWRI = LAPEWRI/L
MAPEEM = LAPEEM/L
```

ฟังก์ชั่นการจำลองค่า

```
function[X,y,yobs,Xobs,ymis,Xmis,Nobs,NOMis]=Gencase1_1(n,CSigmaXy)

z1 = randn(n,1);
z2 = randn(n,1);
z3 = randn(n,1);

X1 = CSigmaXy(1,1)*z1+60;
X2 = CSigmaXy(2,1)*z1+CSigmaXy(2,2)*z2+60;
y = CSigmaXy(3,1)*z1+CSigmaXy(3,2)*z2+CSigmaXy(3,3)*z3+60;
X = [ones(n,1) X1 X2];

pm = 20;
po = 100-pm;
Nobs = ((n*po)/100);
Xobs = X(1:Nobs,:);
yobs = y(1:Nobs,:);
Nmis = Nobs + 1;
NOMis = n - Nobs;
Xmis = X(Nmis:n,:);
ymis = y(Nmis:n,:);
```

ฟังก์ชั่นวิธี Regression Imputation

```
function [yRI,yNewRI,APERI]=RIcase1(y,X,Xobs,yobs,Xmis,n)

bRI = (inv(Xobs'*Xobs))*(Xobs'*yobs);
yRI = (Xmis*bRI);
```

เกณฑ์ MAPE

```
yNewRI = [yobs;yRI];
bhatRI = (inv(X'*X))*(X'*yNewRI);
yhatRI = (X*bhatRI);
DiffRI = (y(:,1))-(yhatRI(:,1));
DiffRRI = (DiffRI(:,1))./(y(:,1));
APERI = ((sum(abs(DiffRRI(:,1))))/n)*100;
```

ฟังก์ชั่นวิธี Nearest Neighbor Imputation

```
function[ED,minED,EDpost,yNNI,DiffNNI,DiffRNNI,APENNI]=NNIcase1(y,X,ymis,yobs,Xobs,
Xmis,Nobs,NOMis,n)

ED = zeros(Nobs,NOMis);
for j=1:NOMis
ED(:,j) = sqrt((Xmis(j,2)-Xobs(:,2)).^2+(Xmis(j,3)-Xobs(:,3)).^2)
 End

[minED EDpost] = min(ED);
ymis(:,1) = yobs(EDpost(1,:),1);
yNNI = ymis;
```

เกณฑ์ MAPE

```
yNewNNI = [yobs;yNNI];
bhatNNI = (inv(X'*X))*(X'*yNewNNI);
yhatNNI = (X*bhatNNI);
DiffNNI = (y(:,1))-(yhatNNI(:,1));
DiffRNNI = (DiffNNI(:,1))./(y(:,1));
APENNI = ((sum(abs(DiffRNNI(:,1))))/n)*100;
```

ฟังก์ชั่นวิธี Weighted Nearest Neighbor and Regression Imputation

```
function [yWRI,APEWRI] = WRIcase1(y,X,yRI,yNNI,Xmis,Xobs,yobs,n,NOMis,vary)

for i = 1:NOMis
```

```
varyRI = Xmis(i,:)*(vary*(inv(Xobs'*Xobs)))*(Xmis(i,:))'
varyNNI = vary
W = varyRI/(varyNNI+varyRI)
yWRI = W*yNNI(:,1) + (1-W)*yRI(:,1);
end
```

เกณฑ์ MAPE

```
yNewWRI = [yobs;yWRI];
bhatWRI = (inv(X'*X))*(X'*yNewWRI);
yhatWRI = (X*bhatWRI);
DiffWRI = (y(:,1))-(yhatWRI(:,1));
DiffRWRI = (DiffWRI(:,1))./(y(:,1));
APEWRI = ((sum(abs(DiffRWRI(:,1))))/n)*100;
```

ฟังก์ชั่นวิธี EM algorithm (Expectation Maximization)

```
function [Xnew,yhat,y_after,b_before,b_after,yEM,APEEM] =
EMcase1New(X,y,Xobs,yobs,Xmis,n)

diff=1;

b0 = (inv(Xobs'*Xobs))*(Xobs'*yobs);
b_after=b0;
y_after = (Xmis*b0);

while diff >0.00001

  b_before = b_after
  yhat = [yobs;y_after ];
  Xnew = [Xobs;Xmis];
  b_after = (inv(Xnew'*Xnew))*(Xnew'*yhat )
  y_after = (Xmis*b_after);
  diff = abs(b_after(:,1) - b_before(:,1))

end

yEM = y_after;
```

เกณฑ์ MAPE

```
yNewEM = [yobs;yEM];
bhatEM = (inv(X'*X))*(X'*yNewEM);
yhatEM = (X*bhatEM);
DiffEM = (y(:,1))-(yhatEM(:,1));
DiffREM = (DiffEM(:,1))./(y(:,1));
APEEM = ((sum(abs(DiffREM(:,1))))/n)*100;
```

กรณีที่ 2 ข้อมูลอนุกรมเวลา

โปรแกรมหลัก(การกำหนดค่า การเรียกใช้ function)

```
clc;
stdeverr =25;
varerr = (stdeverr).^2;
n = 50;
p = 1;
N = 1000;
L=1000;

LAPEComp=0;
LAPERI=0;
LAPENNI=0;
LAPEWRI=0;
LAPEEM=0;

X1 = normrnd(60,10,n,p)
t = [1:n];
tr = t';
A = eye(3);
B = zeros(1,3);
C = [A;B];
D=
[C;C;C;C;C;C;C;C;C;C;C;C;C;C;C;C;C;C;C;C;C;C;C;C;C;C;C;C;C;C;C;C;C;C;C;C;C;C;C;C;C;
C;C;C;C;C;C;C;C;C];
Q = D(1:n,:);
X = [ones(n,1) X1 tr Q];

 while N > 0

   APEComp=0;
   APERI=0;
   APENNI=0;
   APEWRI=0;
   APEEM=0;
```

```
[y,X,X1,yobs,Xobs,ymis,Xmis,Nobs,NOMis] = GenCase2(X,stdeverr,n,t,X1);
```

กรณีข้อมูลสมบูรณ์

```
[yhatcomp,APEComp] = Compcase2(y,X,n);
APECompL(N,1) = APEComp
```

วิธี RI

```
[yRI,yNewRI,yhatRI,DiffRI,DiffRRI,APERI] = RI(y,X,Xobs,yobs,Xmis,n);
APERIL(N,1) = APERI
```

วิธี NNI

```
[ED,minED,EDpost,yNNI,APENNI] = NNI(y,X,ymis,yobs,Xobs,Xmis,Nobs,NOMis,n);
APENNIL(N,1) = APENNI
```

วิธีWRI

```
[yWRI,APEWRI] = WRI(y,X,yRI,yNNI,Xmis,Xobs,yobs,varerr,n,NOMis);
APEWRIL(N,1) = APEWRI
```

วิธีEM

```
[Xnew,yhat,y_after,b_before,b_after,yEM,APEEM] = EMnew1(X,y,Xobs,yobs,Xmis,n);
APEEML(N,1) = APEEM
```

เกณฑ์ MAPE เก็บสะสมแต่ละรอบ

```
LAPEComp = LAPEComp + APEComp;
LAPERI = LAPERI + APERI;
LAPENNI = LAPENNI + APENNI;
LAPEWRI = LAPEWRI + APEWRI;
LAPEEM = LAPEEM + APEEM;

N=N-1
end

end
```

```
Std_APEComp = std(APECompL(:,1))
Std_APERI = std(APERIL(:,1))
Std_APENNI = std(APENNIL(:,1))
Std_APEWRI = std(APEWRIL(:,1))
Std_APEEM = std(APEEML(:,1))

MAPEComp = LAPEComp/L
MAPERI = LAPERI/L
MAPENNI = LAPENNI/L
MAPEWRI = LAPEWRI/L
MAPEEM = LAPEEM/L
```

ฟังก์ชั่นการจำลองค่า

```
function [y,X,X1,yobs,Xobs,ymis,Xmis,Nobs,NOMis] = GenCase2(X,stdeverr,n,t,X1)

%b = [0.5 1 0.3 -0.6 0.6 -0.8]';
%b = [0.5 1 0.8 0.1 -0.1 -0.2]';
b = [0.5 1 0.5 0.5 -0.1 -0.3]';
err = normrnd(0,stdeverr,[1 n]);
y = (X*b)+ err';

pm =20;
po = 100-pm;
Nobs = ((n*po)/100);
Xobs = X(1:Nobs,:);
yobs = y(1:Nobs,:);
Nmis = Nobs + 1;
NOMis = n - Nobs;
Xmis = X(Nmis:n,:);
ymis = y(Nmis:n,:);
```

ฟังก์ชั่นวิธี Regression Imputation

```
function [yRI,yNewRI,yhatRI,DiffRI,DiffRRI,APERI] = RI(y,X,Xobs,yobs,Xmis,n)
```

```
bRI = (inv(Xobs'*Xobs))*(Xobs'*yobs)
yRI = (Xmis*bRI);
```

เกณฑ์ MAPE

```
yNewRI = [yobs;yRI];
XNewRI = [Xobs;Xmis];
bhatRI = (inv(XNewRI'*XNewRI))*(XNewRI'*yNewRI);
yhatRI = (X*bhatRI);
DiffRI = (y(:,1))-(yhatRI(:,1));
DiffRRI = (DiffRI(:,1))./(y(:,1));
APERI = ((sum(abs(DiffRRI(:,1))))/n)*100;
```

ฟังก์ชั่นวิธี Nearest Neighbor Imputation

```
function [ED,minED,EDpost,yNNI,APENNI] = NNI(y,X,ymis,yobs,Xobs,Xmis,Nobs,NOMis,n)

ED = zeros(Nobs,NOMis);

for j=1:NOMis
ED(:,j) = sqrt((Xmis(j,2)-Xobs(:,2)).^2);
End

[minED EDpost] = min(ED);
ynni = ymis;
ynni(:,1) = yobs(EDpost(1,:),1);
yNNI = ynni;
```

เกณฑ์ MAPE

```
yNewNNI = [yobs;yNNI];
bhatNNI = (inv(X'*X))*(X'*yNewNNI);
yhatNNI = (X*bhatNNI);
DiffNNI = (y(:,1))-(yhatNNI(:,1));
DiffRNNI = (DiffNNI(:,1))./(y(:,1));
absDiffRNNI = abs(DiffRNNI(:,1));
APENNI = mean(absDiffRNNI)*100;
```

ฟังก์ชั่นวิธี Weighted Nearest Neighbor and Regression Imputation

```
function [yWRI,APEWRI] = WRI(y,X,yRI,yNNI,Xmis,Xobs,yobs,varerr,n,NOMis)

for i = 1:NOMis
varyRI = Xmis(i,:)*(varerr*(inv(Xobs'*Xobs)))*(Xmis(i,:))'
varyNNI = varerr
W = varyRI/(varyNNI+varyRI)
yWRI = W*yNNI(:,1) + (1-W)*yRI(:,1);
end
```

เกณฑ์ MAPE

```
yNewWRI = [yobs;yWRI];
bhatWRI = (inv(X'*X))*(X'*yNewWRI);
yhatWRI = (X*bhatWRI);
DiffWRI = (y(:,1))-(yhatWRI(:,1));
DiffRWRI = (DiffWRI(:,1))./(y(:,1));
APEWRI = ((sum(abs(DiffRWRI(:,1))))/n)*100;
```

ฟังก์ชั่นวิธี EM algorithm (Expectation Maximization)

```
function [Xnew,yhat,y_after,b_before,b_after,yEM,APEEM] = EMnew1(X,y,Xobs,yobs,Xmis,n)


diff=1;
b0 = (inv(Xobs'*Xobs))*(Xobs'*yobs);
b_after=b0;
y_after = (Xmis*b0);
while diff >0.00001
 b_before = b_after
 yhat = [yobs;y_after ];
 Xnew = [Xobs;Xmis];
 b_after = (inv(Xnew'*Xnew))*(Xnew'*yhat )
 y_after = (Xmis*b_after);
 diff = abs(b_after(:,1) - b_before(:,1))
end

yEM = y_after;
```

เกณฑ์ MAPE

```
yNewEM = [yobs;yEM];
bhatEM = (inv(X'*X))*(X'*yNewEM);
yhatEM = (X*bhatEM);
DiffEM = (y(:,1))-(yhatEM(:,1));
DiffREM = (DiffEM(:,1))./(y(:,1));
APEEM = ((sum(abs(DiffREM(:,1))))/n)*100;
```

ภาคผนวก ข

ตารางที่ 1 แสดงส่วนเบี่ยงเบนมาตรฐานค่า MAPE สำหรับข้อมูลภาคตัดขวางเมื่อค่าสหสัมพันธ์ระหว่าง ค่าสังเกตของตัวแปรตามกับค่าสังเกตของตัวแปรอิสระตัวที่ 1และ ตัวที่ 2 เท่ากันคือ 0.7

σ	n	pm	S.D.			
			RI	NNI	WNR	EM
5	50	5	.102747281	.105254174	.102618692	.102747281
		10	.106593127	.117923849	.106451059	.106593127
		20	.097327499	.150168167	.097384565	.097327499
	100	5	.071526194	.072183068	.071538217	.071526194
		10	.069524304	.072936513	.069541294	.069524304
		20	.074038326	.085091876	.074080922	.074038326
	200	5	.049918806	.050254035	.049922285	.049918806
		10	.049897466	.050355759	.049900090	.049897466
		20	.051150947	.053066675	.051145126	.051150947
10	50	5	.213847339	.234787352	.213763977	.213847339
		10	.212718005	.245786980	.212771996	.212718005
		20	.222987641	.296156513	.222856083	.222987641
	100	5	.160287282	.161003667	.160254855	.160287282
		10	.154705058	.160279138	.154726781	.154705058
		20	.161121597	.185295907	.161023018	.161121597
	200	5	.104677371	.105826290	.104679283	.104677371
		10	.108777682	.108977084	.108774099	.108777682
		20	.106668194	.111956181	.106714312	.106668194
15	50	5	.497040964	.646421934	.502037894	.497040964
		10	.662265153	.433304197	.639906144	.662265153
		20	.385516479	.609621801	.392459119	.385516479
	100	5	.257328773	.261384995	.257296746	.257328773
		10	.318070139	.331751924	.318270867	.318070139
		20	.323998751	.466727267	.327156565	.323998751
	200	5	.200060528	.163320335	.199800584	.200060528
		10	.212232073	.239012691	.212411213	.212232073
		20	.357024693	.224280963	.335917843	.357024693

ตารางที่1(ต่อ) แสดงส่วนเบี่ยงเบนมาตรฐานค่า MAPE สำหรับข้อมูลภาคตัดขวางเมื่อค่าสหสัมพันธ์ ระหว่าง ค่าสังเกตของตัวแปรตามกับค่าสังเกตของตัวแปรอิสระตัวที่ 1และ ตัวที่ 2 เท่ากันคือ 0.7

σ	n	pm	S.D.			
			RI	NNI	WNR	EM
20	50	5	2.853505294	2.870852672	2.578500066	2.853505294
		10	2.388075170	2.759723153	2.395763483	2.388075170
		20	2.005880383	3.674496697	2.061120982	2.005880383
	100	5	2.699815687	2.485043959	2.689597035	2.699815687
		10	2.431944804	2.349798384	2.430338077	2.431944804
		20	1.699256382	2.246273742	1.715414592	1.699256382
	200	5	2.428598314	2.353608286	2.428499092	2.428598314
		10	2.849075041	2.787051837	2.845617519	2.849075041
		20	2.493878049	3.837782794	2.520919966	2.493878049
25	50	5	4.120969715	5.356176134	4.199896300	4.120969715
		10	3.745432416	4.565665725	3.765083043	3.745432416
		20	3.225555947	4.273262709	3.161712001	3.225555947
	100	5	4.534170497	4.324000379	4.512347189	4.534170497
		10	4.467201920	4.873593037	4.444299716	4.467201920
		20	5.011061779	4.955474497	4.993356418	5.011061779
	200	5	3.967072850	3.087900801	3.968594660	3.967072850
		10	4.476730092	4.143945657	4.455100726	4.476730092
		20	3.636226600	3.072103982	3.639993737	3.636226600

ตารางที่ 2 แสดงส่วนเบี่ยงเบนมาตรฐานค่า MAPE สำหรับข้อมูลภาคตัดขวางเมื่อค่าสหสัมพันธ์ระหว่าง ค่าสังเกตของตัวแปรตามกับค่าสังเกตของตัวแปรอิสระตัวที่ 1และ ตัวที่ 2 คือ 0.9 และ 0.4 ตามลำดับ

σ	n	pm	S.D.			
			RI	NNI	WNR	EM
5	50	5	.122498921	.125561823	.122523004	.122498921
		10	.128005949	.134218419	.127820661	.128005949
		20	.128791144	.153039647	.128838251	.128791144
	100	5	.092117040	.093023398	.092131996	.092117040
		10	.092244372	.093194971	.092264199	.092244372
		20	.090415014	.096772731	.090446246	.090415014
	200	5	.060623563	.060965934	.060625811	.060623563
		10	.062965918	.062900009	.062966168	.062965918
		20	.062585494	.063783879	.062580468	.062585494
10	50	5	.263492553	.277974910	.263347666	.263492553
		10	.265866778	.275312917	.265665540	.265866778
		20	.260402096	.349431365	.260923938	.260402096
	100	5	.188228849	.189784793	.188253999	.188228849
		10	.181902002	.185324261	.181884655	.181902002
		20	.193299179	.208535388	.193568407	.193299179
	200	5	.127745790	.128237947	.127750271	.127745790
		10	.129371306	.129665217	.129362234	.129371306
		20	.128332054	.131084325	.128336934	.128332054
15	50	5	.570868702	.591966511	.573229656	.570868702
		10	.545426186	.764735498	.558995194	.545426186
		20	.483600088	.600922688	.483886710	.483600088
	100	5	.328183849	.331481879	.328143953	.328183849
		10	.343382412	.362663201	.343657386	.343382412
		20	.351157126	.524247456	.355402669	.351157126
	200	5	.274917392	.327362797	.275705900	.274917392
		10	.294650073	.309156321	.294836856	.294650073
		20	.235550780	.247983380	.235465708	.235550780

ตารางที่ 2 (ต่อ) แสดงส่วนเบี่ยงเบนมาตรฐานค่า MAPE สำหรับข้อมูลภาคตัดขวางเมื่อค่าสหสัมพันธ์ ระหว่าง ค่าสังเกตของตัวแปรตามกับค่าสังเกตของตัวแปรอิสระตัวที่ 1และ ตัวที่ 2 คือ 0.9 และ 0.4 ตามลำดับ

σ	n	pm	S.D.			
			RI	NNI	WNR	EM
20	50	5	4.100657830	4.224949319	4.107685076	4.100657830
		10	3.887782718	4.181407883	3.858459340	3.887782718
		20	2.598345118	3.901573817	2.685033281	2.598345118
	100	5	2.306401686	2.159565626	2.303063183	2.306401686
		10	2.727323808	2.987997160	2.712736019	2.727323808
		20	2.906807098	3.417646945	2.914427789	2.906807098
	200	5	1.461908098	1.608632727	1.462840481	1.461908098
		10	2.381029917	2.499772606	2.382866951	2.381029917
		20	2.644187475	2.832638500	2.639313809	2.644187475
25	50	5	5.930337068	6.546253947	5.960646366	5.930337068
		10	5.051202164	4.821781757	4.984178392	5.051202164
		20	5.861764395	8.449287382	5.871900106	5.861764395
	100	5	5.017474669	4.348721871	4.900118293	5.017474669
		10	5.172665657	5.557950622	5.174389044	5.172665657
		20	4.835303427	4.536488432	4.837464376	4.835303427
	200	5	5.371255196	5.133708890	5.367766977	5.371255196
		10	4.763593665	4.843042987	4.763199214	4.763593665
		20	5.327721638	5.068061493	5.314495527	5.327721638

ตารางที่ 3 แสดงส่วนเบี่ยงเบนมาตรฐานค่า MAPE สำหรับข้อมูลอนุกรมเวลากรณีมีอิทธิพลจากปัจจัยแนวโน้มต่ำแต่ปัจจัยฤดูกาลสูง ($\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 0.3$, $\beta_3 = -0.6$, $\beta_4 = 0.6$ และ $\beta_5 = -0.8$)

σ	n	pm	S.D.			
			RI	NNI	WNR	EM
5	50	5	.704286665	.707825926	.704541503	.704286665
		10	.689338120	.704167093	.687744654	.689338120
		20	.695840696	.783941663	.695076482	.695840696
	100	5	.425276454	.428975147	.425389565	.425276454
		10	.418688484	.429667697	.418418743	.418688484
		20	.443538518	.509989805	.439969995	.443538518
	200	5	.257587436	.256970253	.257775085	.257587436
		10	.258322245	.276969316	.257601254	.258322245
		20	.267403356	.370089351	.268322872	.267403356
10	50	5	1.506254251	1.510533190	1.505510529	1.506254251
		10	1.263466858	1.348777712	1.275909357	1.263466858
		20	1.534598908	1.539547638	1.518706464	1.534598908
	100	5	.940838866	.936093131	.940264048	.940838866
		10	.976053137	.973931236	.975062677	.976053137
		20	.999771667	1.027239187	.999317804	.999771667
	200	5	.604532968	.609020970	.604912070	.604532968
		10	.577483405	.606427331	.577531005	.577483405
		20	.587064934	.763439435	.586856101	.587064934
15	50	5	5.049213912	5.114781379	5.058783286	5.049213912
		10	4.998208575	5.189774245	5.024965661	4.998208575
		20	4.859674552	4.798815705	4.898154294	4.859674552
	100	5	3.028215908	3.020783742	3.027864157	3.028215908
		10	2.895908579	2.977878095	2.901989842	2.895908579
		20	3.494438348	3.768183877	3.515872003	3.494438348
	200	5	2.069714138	2.145077615	2.072752757	2.069714138
		10	2.030450800	2.167141225	2.040632015	2.030450800
		20	1.912087733	2.192334618	1.925212094	1.912087733

ตารางที่ 3 (ต่อ) แสดงส่วนเบี่ยงเบนมาตรฐานค่า MAPE สำหรับข้อมูลอนุกรมเวลากรณีมีอิทธิพลจากปัจจัยแนวโน้มต่ำแต่ปัจจัยฤดูกาลสูง ($\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 0.3$, $\beta_3 = -0.6$, $\beta_4 = 0.6$ และ $\beta_5 = -0.8$)

σ	n	pm	S.D.			
			RI	NNI	WNR	EM
20	50	5	6.692100596	6.742899450	6.697815345	6.692100596
		10	7.149294008	7.218910246	7.157128116	7.149294008
		20	6.939901110	7.151539095	6.982845601	6.939901110
	100	5	5.308970168	5.381438917	5.314147109	5.308970168
		10	5.798220371	5.951372688	5.812395270	5.798220371
		20	5.637586301	5.912753670	5.662401740	5.637586301
	200	5	3.898143752	4.256991793	3.911873250	3.898143752
		10	4.258278682	4.534798687	4.273464296	4.258278682
		20	4.160816261	4.646595769	4.186230619	4.160816261
25	50	5	9.692123946	9.759203605	9.699231716	9.692123946
		10	9.623314658	9.755500947	9.636468649	9.623314658
		20	9.682053466	9.968287573	9.684941963	9.682053466
	100	5	7.224523147	7.189699279	7.220714730	7.224523147
		10	6.462415165	6.552890295	6.468862910	6.462415165
		20	6.725638424	6.944122418	6.739909079	6.725638424
	200	5	4.909222095	4.932306089	4.922604193	4.909222095
		10	5.303069815	5.036790732	5.335017252	5.303069815
		20	5.601339745	5.454593302	5.649150467	5.601339745

ตารางที่ 4 แสดงส่วนเบี่ยงเบนมาตรฐานค่า MAPE สำหรับข้อมูลอนุกรมเวลากรณีมีอิทธิพลจากปัจจัยแนวโน้มสูงแต่ปัจจัยฤดูกาลต่ำ ($\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 0.8$, $\beta_3 = 0.1$, $\beta_4 =$ -0.1และ $\beta_5 =$-0.2)

σ	n	pm	S.D.			
			RI	NNI	WNR	EM
5	50	5	.564878331	.552836968	.564057956	.564878331
		10	.584805041	.608317049	.584121114	.584805041
		20	.579413941	.702500817	.576964133	.579413941
	100	5	.329343814	.321083214	.329098349	.329343814
		10	.336506960	.408098595	.335506964	.336506960
		20	.342361054	.487154196	.348247050	.342361054
	200	5	.181905216	.183776694	.182179991	.181905216
		10	.183909170	.232267369	.185955516	.183909170
		20	.190767189	.272902167	.194820272	.190767189
10	50	5	.901926610	.897090837	.900356393	.901926610
		10	.852668221	.869827690	.845897252	.852668221
		20	.913370221	.938408369	.913515849	.913370221
	100	5	.757612040	.758510025	.757422751	.757612040
		10	.754242727	.780631832	.754830773	.754242727
		20	.740581611	.907156443	.739252820	.740581611
	200	5	.414749462	.423223151	.414879146	.414749462
		10	.508287899	.577195206	.510614427	.508287899
		20	.409564558	.545867090	.412272119	.409564558
15	50	5	2.757101751	2.820698851	2.767887322	2.757101751
		10	2.726806667	2.789111041	2.734750491	2.726806667
		20	2.916655562	3.171881059	2.911523358	2.916655562
	100	5	2.725232393	3.250532547	2.764503281	2.725232393
		10	1.427916786	1.527634609	1.433822583	1.427916786
		20	1.575715562	1.872292563	1.587180538	1.575715562
	200	5	.853991757	.969755406	.962612526	.853991757
		10	1.060431539	1.672450242	1.084832244	1.060431539
		20	.981429487	1.007311525	.983136106	.981429487

ตารางที่ 4 (ต่อ) แสดงส่วนเบี่ยงเบนมาตรฐานค่า MAPE สำหรับข้อมูลอนุกรมเวลากรณีมีอิทธิพลจากปัจจัยแนวโน้มสูงแต่ปัจจัยฤดูกาลต่ำ ($\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 0.8$, $\beta_3 = 0.1$, $\beta_4 =$ -0.1และ $\beta_5 =$-0.2)

σ	n	pm	S.D.			
			RI	NNI	WNR	EM
20	50	5	4.312269458	4.448867121	4.329900551	4.312269458
		10	5.268754755	5.550270517	5.311792215	5.268754755
		20	4.636361307	4.759233931	4.615536425	4.636361307
	100	5	3.131097724	3.256662219	3.140786465	3.131097724
		10	3.449298845	3.932637912	3.484301911	3.449298845
		20	3.540364990	3.944948407	3.571099016	3.540364990
	200	5	2.484736587	2.602187670	2.489425221	2.484736587
		10	2.532313200	2.445009223	2.572027910	2.532313200
		20	2.220039442	2.299285763	2.292607206	2.220039442
25	50	5	8.503998607	8.686194991	8.529790861	8.503998607
		10	9.171992490	9.128482923	9.037596656	9.171992490
		20	8.558520068	8.929046050	8.611705650	8.558520068
	100	5	7.004809187	7.261183011	7.027495956	7.004809187
		10	6.566425977	7.395259311	6.637056168	6.566425977
		20	6.784077821	7.990583069	6.902608090	6.784077821
	200	5	4.271503235	4.412245609	4.292769341	4.271503235
		10	4.158666947	4.332942495	4.110873929	4.158666947
		20	4.131527564	4.431977485	4.155493603	4.131527564

ตารางที่ 5 แสดงส่วนเบี่ยงเบนมาตรฐานค่า MAPE สำหรับข้อมูลอนุกรมเวลากรณีมีอิทธิพลจากปัจจัยแนวโน้มและปัจจัยฤดูกาลระดับปานกลาง ($\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 0.5$, $\beta_3 = 0.5$, $\beta_4 =$ -0.1และ $\beta_5 =$ -0.3)

σ	n	pm	S.D.			
			RI	NNI	WNR	EM
5	50	5	.627625354	.630587585	.628100469	.627625354
		10	.616282698	.697510543	.613589329	.616282698
		20	.637765839	.759177295	.636192030	.637765839
	100	5	.386786425	.393490465	.385613208	.386786425
		10	.397375303	.441733041	.395430684	.397375303
		20	.395731267	.523032689	.398522157	.395731267
	200	5	.237327558	.234898393	.237681274	.237327558
		10	.222223983	.247667996	.222231232	.222223983
		20	.225055652	.321254079	.226333695	.225055652
10	50	5	1.403771074	1.395622336	1.401542708	1.403771074
		10	1.370652374	1.384844310	1.366299460	1.370652374
		20	1.419934492	1.437165592	1.402767298	1.419934492
	100	5	.883621639	.888924754	.883686652	.883621639
		10	.849122730	.853191618	.847549520	.849122730
		20	.856121077	1.022632806	.847856059	.856121077
	200	5	.478946590	.481972413	.479043363	.478946590
		10	.483870917	.502930833	.484137829	.483870917
		20	.520186042	.743330363	.524577555	.520186042
15	50	5	4.226021229	4.177254825	4.219967535	4.226021229
		10	3.881219649	4.145018391	3.923271435	3.881219649
		20	3.207673388	3.407561744	3.226093582	3.207673388
	100	5	2.612290525	2.689392104	2.618647561	2.612290525
		10	2.296907538	2.361237265	2.302447461	2.296907538
		20	2.746852569	3.684323629	2.839966472	2.746852569
	200	5	1.107890188	1.137537141	1.108823595	1.107890188
		10	.914256530	1.034992217	.918135586	.914256530
		20	1.232644664	1.557113041	1.244589237	1.232644664

ตารางที่ 5 (ต่อ) แสดงส่วนเบี่ยงเบนมาตรฐานค่า MAPE สำหรับข้อมูลอนุกรมเวลากรณีมีอิทธิพลจากปัจจัยแนวโน้มและปัจจัยฤดูกาลระดับปานกลาง ($\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 0.5$, $\beta_3 = 0.5$, $\beta_4 = -0.1$ และ $\beta_5 = -0.3$)

σ	n	pm	S.D.			
			RI	NNI	WNR	EM
20	50	5	5.617116792	5.553491422	5.604370392	5.617116792
		10	6.309708996	6.140501176	6.244413314	6.309708996
		20	5.355526809	5.525255622	5.371146667	5.355526809
	100	5	4.998567061	5.055580501	5.002529650	4.998567061
		10	4.432687812	4.538650227	4.441551120	4.432687812
		20	4.129933576	4.421450686	4.154666572	4.129933576
	200	5	3.774847855	3.656504526	3.765841137	3.774847855
		10	3.518325615	3.388149875	3.557675699	3.518325615
		20	3.403551809	3.967532127	3.429225319	3.403551809
25	50	5	8.748477676	8.845919225	8.759780819	8.748477676
		10	9.262359995	9.497377130	9.289677921	9.262359995
		20	8.627894040	8.811089417	8.647930616	8.627894040
	100	5	6.742692084	6.875641938	6.752304198	6.742692084
		10	7.321546581	7.853176205	7.364144321	7.321546581
		20	7.550971503	7.871654673	7.593053029	7.550971503
	200	5	4.354184010	4.627227257	4.364726271	4.354184010
		10	4.161196463	4.287628715	4.208636672	4.161196463
		20	4.324677776	5.894700384	4.421697150	4.324677776

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวเพียงออ ยีสา เกิดวันที่ 24 มิถุนายน พ.ศ.2526 ที่จังหวัดนครศรีธรรมราช สำเร็จการศึกษาปริญญาตรีวิทยาศาสตรบัณฑิต สาขาวิชาสถิติ มหาวิทยาลัยศิลปากร ปีการศึกษา 2548 และเข้าศึกษาต่อในหลักสูตรสถิติศาสตรมหาบัณฑิตที่จุฬาลงกรณ์มหาวิทยาลัย เมื่อ พ.ศ.2549