

การตรวจสอบลักษณะของข้อมูลอนุกรมเวลาสำหรับการพยากรณ์
อุปสงค์การท่องเที่ยวในประเทศไทย
Time Series Data Characteristics Checking for Forecasting
Tourism Demand in Thailand

อัครพงศ์ อันทอง*

สถาบันศึกษานโยบายสาธารณะ

มหาวิทยาลัยเชียงใหม่ อำเภอเมือง จังหวัดเชียงใหม่ 50200

ปวีณา คำพุกกะ

คณะบริหารศาสตร์ มหาวิทยาลัยอุบลราชธานี

อำเภอวารินชำราบ จังหวัดอุบลราชธานี 34190

บทคัดย่อ

บทความฉบับนี้มีวัตถุประสงค์เพื่อตรวจสอบลักษณะของข้อมูลอนุกรมเวลาในการพยากรณ์อุปสงค์การท่องเที่ยวในประเทศไทย โดยใช้วิธีการวิเคราะห์การพยากรณ์แบบ ex post ด้วยการประยุกต์ใช้วิธีการบ็อกซ์และเจนกินส์ในการสร้างแบบจำลองพยากรณ์ ผลการศึกษาพบว่า ลักษณะของข้อมูลอนุกรมเวลาที่แตกต่างกันมีผลต่อความแม่นยำในการพยากรณ์อุปสงค์การท่องเที่ยวในประเทศไทย โดยข้อมูลอนุกรมเวลาที่แบ่งตามลักษณะเฉพาะของนักท่องเที่ยวและมีความถี่ที่สูงกว่า ให้ผลการพยากรณ์ที่มีความแม่นยำมากกว่า เนื่องจากอุปสงค์การท่องเที่ยวมีลักษณะที่ไม่เหมือนกันในแต่ละบุคคล อีกทั้งมีอิทธิพลของฤดูกาล และอ่อนไหวต่อเหตุการณ์วิกฤติต่างๆ ดังนั้นควรพิจารณาถึงลักษณะพื้นฐาน องค์ประกอบ และความคงที่ของข้อมูลอนุกรมเวลาก่อนนำมาใช้สร้างแบบจำลองพยากรณ์ เนื่องจากลักษณะพื้นฐานของข้อมูลอนุกรมเวลาที่มีผลต่อการตัดสินใจเลือกวิธีการหรือแบบจำลองพยากรณ์ ในขณะที่ความคงที่ของข้อมูลอนุกรมเวลาเป็นข้อสมมติเบื้องต้นที่สำคัญในการใช้วิธีการพยากรณ์ และเศรษฐกิจ

คำสำคัญ : ลักษณะของข้อมูลอนุกรมเวลา ความแม่นยำในการพยากรณ์ อุปสงค์การท่องเที่ยวในประเทศไทย

* บทความฉบับนี้เป็นส่วนหนึ่งของโครงการท่องเที่ยวไทย: จากนโยบายสู่การปฏิบัติ ของ ศ.ดร.มิ่งสรรพ์ ขาวสอาด ซึ่งได้รับการสนับสนุนจากสำนักงานกองทุนสนับสนุนการวิจัย (สกว.) ภายใต้ทุนส่งเสริมกลุ่มวิจัย (เมธีวิจัยอาวุโส สกว.).

Abstract

The aim of this article was to check time series data characteristics in forecasting tourism demand in Thailand. The study used the ex post forecasting analysis and applied the Box and Jenkins method to create a forecasting model. The empirical results showed that the difference in time series data characteristics affect accuracy in forecasting tourism demand in Thailand. The time series data characteristics of tourists and high frequency data provided more accurate forecasts as tourism demand characteristics have heterogeneous individuals, seasonality, and are sensitive to crises. Therefore, analysts should consider basic characteristics and components of time series data before using it to create a forecasting model as its data affect decisions, method, and/or the forecasting model. Moreover, the stationary nature of time series data is a basic assumption for using a forecasting method and econometrics.

Keywords: time series data characteristics, accuracy of forecasting, tourism demand in Thailand

บทนำ

นับตั้งแต่ปี พ.ศ. 2503 เป็นต้นมา การพยากรณ์อุปสงค์การท่องเที่ยวเป็นหนึ่งในงานวิจัยทางการท่องเที่ยวที่ได้รับความสนใจ และมีการตีพิมพ์หรือเผยแพร่เพิ่มขึ้นอย่างต่อเนื่องมาโดยตลอด 4 ทศวรรษที่ผ่านมา (Li et al., 2005) โดยในช่วงระหว่างปี พ.ศ. 2504-2536 พบว่า งานวิจัยเกี่ยวกับการสร้างแบบจำลองพยากรณ์อุปสงค์การท่องเที่ยวมีจำนวนมากถึง 120 เรื่อง จากจำนวนงานวิจัยทางการท่องเที่ยวทั้งหมด 300 เรื่อง หรือประมาณร้อยละ 40 ของจำนวนงานวิจัยทางการท่องเที่ยวทั้งหมด (Crouch, 1994) ต่อมาในช่วงระหว่างปี พ.ศ. 2543-2550 หรือ 8 ปีที่ผ่านมา พบว่า งานวิจัยทางด้านอุปสงค์การท่องเที่ยวที่ได้รับการตีพิมพ์ในวารสารระดับนานาชาติมีมากถึง 121 เรื่อง (Song and Li, 2008) จากข้อมูลดังกล่าวแสดงให้เห็นถึงความสนใจที่เพิ่มขึ้นอย่างต่อเนื่องในการวิจัยหรือศึกษาแบบจำลองพยากรณ์อุปสงค์การท่องเที่ยวในระดับนานาชาติ

วิธีการที่นิยมใช้ในการสร้างแบบจำลองพยากรณ์อุปสงค์การท่องเที่ยวในปัจจุบันประกอบด้วย แบบจำลองอนุกรมเวลา (time series models), เศรษฐมิติ (econometrics) เช่น Error Correction Model (ECM) เป็นต้น และวิธีการเชิงปริมาณใหม่ๆ ทั้งที่เป็นวิธีการ

ทางสถิติ เช่น SARIMA Intervention เป็นต้น และที่ไม่ใช้วิธีการทางสถิติ เช่น Artificial Neural Network (ANN) เป็นต้น (Song and Li, 2008) แต่อย่างไรก็ตามแบบจำลองอนุกรมเวลายังคงเป็นวิธีการที่นิยมนำมาใช้ในการพยากรณ์อุปสงค์การท่องเที่ยวมากกว่าวิธีการอื่นๆ ส่วนใหญ่นำเสนอผลลัพธ์ของการพยากรณ์ในลักษณะของการพยากรณ์ไปในอนาคต (ex ante forecast) หรือการเปรียบเทียบลักษณะของวิธีการพยากรณ์เพื่อนำเสนอถึงความเหมาะสมของแบบจำลองที่นำมาใช้ (model fit) (Witt and Witt, 1995) นอกจากแบบจำลองอนุกรมเวลาแล้ว วิธีการวิเคราะห์ข้อมูลตัวแปรเดียว (univariate analysis) เช่น วิธีการบ็อกซ์และเจนกินส์ (Box and Jenkins) (ARIMA, SARIMA) เป็นอีกวิธีการหนึ่งที่มีความนิยมนำมาใช้ในการพยากรณ์อุปสงค์การท่องเที่ยวในปัจจุบัน (Witt and Witt, 1995) แต่อย่างไรก็ตามงานศึกษาที่ผ่านมาเน้นการนำเสนอผลการพยากรณ์ไปในอนาคต และการหาแบบจำลองที่เหมาะสม มากกว่าการให้ความสำคัญกับความแม่นยำในการพยากรณ์ ซึ่งถือได้ว่าเป็นวัตถุประสงค์หลักที่สำคัญของการสร้างแบบจำลองพยากรณ์ หรืออาจกล่าวได้ว่างานวิจัยเพื่อศึกษาถึงความแม่นยำ (accuracy) ของการพยากรณ์มีค่อนข้างน้อยเมื่อเทียบกับงานวิจัยประเภทอื่นๆ โดยการศึกษาในลักษณะนี้นิยมใช้รูปแบบการวิเคราะห์ที่เรียกว่า ex-post forecast ในการศึกษาความแม่นยำของการพยากรณ์หรือของแบบจำลองพยากรณ์ที่สร้างขึ้น

ข้อมูลที่ใช้ในการสร้างแบบจำลองพยากรณ์อุปสงค์การท่องเที่ยวเกือบทั้งหมดเป็นข้อมูลอนุกรมเวลาประเภททุติยภูมิที่อ้างอิงจากหน่วยงานที่เกี่ยวข้อง โดยปกติมีการแบ่งประเภทตามลักษณะเฉพาะของนักท่องเที่ยว เช่น ประเทศที่อยู่อาศัย (country of residence) วัตถุประสงค์ในการเดินทาง (purpose of the trip) เป็นต้น หรือแบ่งตามความถี่ของข้อมูลอนุกรมเวลา เช่น ข้อมูลรายปี รายไตรมาส รายเดือน เป็นต้น เนื่องจากอุปสงค์การท่องเที่ยวมีลักษณะที่ไม่เหมือนกันในแต่ละบุคคล (heterogeneous or disaggregated individuals) มากกว่าที่มีลักษณะเหมือนกันในทุกๆ คน (homogeneous or aggregated individuals) จึงทำให้ต้องมีการสร้างแบบจำลองพยากรณ์ที่แตกต่างกัน และอาจมีผลต่อความแม่นยำในการพยากรณ์ ที่ผ่านมามีการศึกษาอย่างกว้างขวางถึงความแม่นยำของการนำข้อมูลที่มีลักษณะแยกย่อย (disaggregated) และลักษณะรวมกัน (aggregated) มาใช้ ส่วนใหญ่เป็นการศึกษาโดยใช้ข้อมูลเศรษฐกิจมหภาค เช่น Grunfeld and Griliches, 1960; Edwards and Orcutt, 1969; Rose, 1977; Tiao and Guttman, 1980; Lütkepohl, 1984; Palm and Zellner, 1992; Clark, 2000; Zellner and Tobias, 2000; Espasa and Albacete, 2007 เป็นต้น แต่มีการศึกษาโดยใช้ข้อมูลอุปสงค์การท่องเที่ยวจำนวนน้อย (Kim and Moosa, 2005) อย่างไรก็ตามงานศึกษาในอดีตได้สะท้อนและให้ข้อสังเกตบางประการ

ในการใช้ข้อมูลอนุกรมเวลาของอุปสงค์การท่องเที่ยวสำหรับการพยากรณ์ว่า การใช้ข้อมูลที่มีลักษณะแยกย่อยอาจมีความแม่นยำในการพยากรณ์มากกว่าการใช้ข้อมูลที่มีลักษณะรวมกัน เนื่องจากข้อมูลลักษณะรวมกันไม่ได้พิจารณาถึงความแตกต่างของพฤติกรรมและองค์ประกอบย่อยอื่นๆ ที่มีอยู่ในอุปสงค์การท่องเที่ยว จึงทำให้เกิดความคลาดเคลื่อนจากการพยากรณ์มากกว่า (Kim and Moosa, 2005) แต่อย่างไรก็ตามในบางกรณีข้อมูลที่มีลักษณะรวมกันอาจให้ผลการพยากรณ์ที่มีความแม่นยำมากกว่า (Vu and Turner, 2005; Kon and Turner, 2005) โดยเฉพาะในกรณีที่อุปสงค์การท่องเที่ยวที่มีลักษณะคล้ายคลึงกัน (homogeneous) หรือแหล่งท่องเที่ยวมีขนาดเล็กและขาดความหลากหลายของสิ่งดึงดูดใจด้านการท่องเที่ยว ดังนั้นนักท่องเที่ยวที่เดินทางมาท่องเที่ยวจึงมีลักษณะเฉพาะที่ใกล้เคียงกัน เช่น ในกรณีของฮ่องกงที่นักท่องเที่ยวส่วนใหญ่เดินทางไปฮ่องกงเพื่อซื้อของเป็นสำคัญ เป็นต้น ดังนั้นจากผลการศึกษาในอดีตจึงยังไม่สามารถสรุปได้อย่างชัดเจนว่า ข้อมูลอนุกรมเวลาของอุปสงค์การท่องเที่ยวที่มีลักษณะแยกย่อยหรือลักษณะรวมกัน แบบไหนที่ให้การพยากรณ์มีความแม่นยำมากกว่ากัน

นอกจากนี้ แม้ว่าในอดีตการสร้างแบบจำลองพยากรณ์อุปสงค์การท่องเที่ยวจะนิยมใช้ข้อมูลรายปี แต่ในปัจจุบันมีการใช้ข้อมูลรายไตรมาส และรายเดือนเพิ่มขึ้นอย่างต่อเนื่อง เนื่องจากการใช้ข้อมูลรายไตรมาส และรายเดือนมีอิทธิพลของฤดูกาลที่เป็นหนึ่งในองค์ประกอบที่สำคัญของอุปสงค์การท่องเที่ยว จากการพิจารณางานวิจัยทางด้านอุปสงค์การท่องเที่ยวในอดีต พบว่า งานวิจัยส่วนใหญ่นิยมใช้ข้อมูลรายปีที่มีขนาดตัวอย่างเล็กประมาณ 44-99 ตัวอย่างในการวิเคราะห์ (Lim, 1997) และมีงานวิจัยเพียงเล็กน้อยที่ใช้ข้อมูลรายไตรมาส ที่มีขนาดตัวอย่างประมาณ 44-99 ตัวอย่าง (Lim and McAleer, 2003) ดังนั้นการใช้ข้อมูลรายปีในการประมาณค่าสัมประสิทธิ์ของแบบจำลองพยากรณ์ด้วยวิธีการถดถอย (regression) อาจได้ค่าสัมประสิทธิ์ที่มีอคติ (bias) มีปัญหาความเที่ยงตรง (precision problem) และมีปัญหาความสอดคล้อง (non consistent) เนื่องจากขนาดของตัวอย่างมีผลต่อค่าสัมประสิทธิ์ของวิธีการถดถอย และค่า degree of freedom ของสมการพยากรณ์ โดยทั่วไปการวิเคราะห์ข้อมูลหลายตัวแปร (multivariate analysis) โดยเฉพาะการใช้วิธีการถดถอยจะต้องมีขนาดตัวอย่างอย่างน้อย 10-20 ตัวอย่างต่อค่าสัมประสิทธิ์ที่ต้องการประมาณค่า 1 ตัว (Hair et al., 1998) ดังนั้นการใช้ข้อมูลรายไตรมาสหรือรายเดือนสามารถแก้ไขปัญหาดังกล่าวได้ เนื่องเป็นการเพิ่มขนาดของตัวอย่างที่ใช้ในการวิเคราะห์ อีกทั้งข้อมูลรายไตรมาสหรือรายเดือนยังมีอิทธิพลของฤดูกาลหรืออิทธิพลของการเปลี่ยนแปลงภายในรอบปีไว้ในข้อมูลอีกด้วย ซึ่งเป็นองค์ประกอบหนึ่งที่สำคัญของข้อมูลอุปสงค์การท่องเที่ยว นอกจากนี้การใช้ข้อมูลรายไตรมาสหรือรายเดือนในการสร้างแบบจำลอง

พยากรณ์อุปสงค์การท่องเที่ยวยังสามารถตรวจสอบผลกระทบจากอิทธิพลของฤดูกาลได้อีกด้วย (Lim and McAleer, 2003; Akarapong and Mingsarn, 2009) ซึ่งสอดคล้องกับพฤติกรรมโดยทั่วไปของนักท่องเที่ยวที่นิยมท่องเที่ยวตามฤดูกาล (Lim and McAleer, 2000; Rossello, Riera and Sanso, 2004; Fernandez-Morales and Mayorga-Toledano, 2008, Akarapong and Mingsarn, 2009)

จากการศึกษาที่ผ่านมามีข้อสังเกตที่น่าสนใจศึกษาเพิ่มเติมในประเด็นที่ว่า ลักษณะของข้อมูลอนุกรมเวลาที่แตกต่างกันอาจมีผลต่อความแม่นยำในการพยากรณ์อุปสงค์การท่องเที่ยว ดังนั้นบทความนี้จึงมีวัตถุประสงค์เพื่อตรวจสอบความแม่นยำของการพยากรณ์อุปสงค์การท่องเที่ยวในประเทศไทยภายใต้ลักษณะของข้อมูลอนุกรมเวลาที่แตกต่างกัน ทั้งที่เป็นความแตกต่างในเรื่องของลักษณะเฉพาะของนักท่องเที่ยว (ประเทศที่อยู่อาศัย) และความถี่ของข้อมูล (ข้อมูลรายปี และรายเดือน) รวมทั้งผลการศึกษาที่ได้จะเป็นข้อมูลที่เป็นประโยชน์ต่อผู้ใช้ข้อมูลอนุกรมเวลาในการสร้างแบบจำลองพยากรณ์อุปสงค์การท่องเที่ยวในกรณีอื่นๆ ที่ว่า ผู้ใช้ข้อมูลควรตระหนักถึงลักษณะเฉพาะต่างๆ ของข้อมูลอนุกรมเวลาที่เราจะใช้ และควรมีการตรวจสอบองค์ประกอบต่างๆ ของข้อมูลอนุกรมเวลาก่อนนำมาใช้ เพื่อให้ผลการพยากรณ์มีความแม่นยำมากขึ้น และใช้เป็นข้อมูลประกอบการตัดสินใจเลือกใช้ในการหรือแบบจำลองพยากรณ์ที่เหมาะสม ตลอดจนกล่าวถึงแนวคิด วิธีการศึกษา และการสร้างแบบจำลองพยากรณ์ จากนั้นเป็นการนำเสนอผลการศึกษาเชิงประจักษ์ และการอภิปรายผลการศึกษารวมทั้งเสนอแนะแนวทางโดยทั่วไปในการตรวจสอบข้อมูลอนุกรมเวลา ตามด้วยสรุปและข้อคิดเห็นบางประการในการนำข้อมูลอนุกรมเวลามาใช้สร้างแบบจำลองพยากรณ์อุปสงค์การท่องเที่ยวของประเทศไทยที่เหมาะสม เพื่อให้ได้ผลการพยากรณ์ที่มีความแม่นยำมากยิ่งขึ้น

2. แนวคิด วิธีการศึกษา และการสร้างแบบจำลองพยากรณ์

หนึ่งในประโยชน์ของข้อมูลอนุกรมเวลา คือ การนำมาใช้สร้างแบบจำลองสำหรับการพยากรณ์ค่าในอนาคต ปัจจุบันมีวิธีการพยากรณ์อยู่หลายวิธี เช่น วิธีการปรับให้เรียบแบบเอ็กซ์โปเนนเชียล (exponential smoothing) การแยกองค์ประกอบ (decomposition) การถดถอยพหุคูณ (multiple regression) นิวรอลเน็ตเวิร์ค (neural network) บ็อกซ์และเจนกินส์ (Box and Jenkins) เป็นต้น (อัศวพงศ์ อันทอง และปวีณา คำพุทะ, 2552) สำหรับบทความนี้เลือกใช้วิธีการบ็อกซ์และเจนกินส์ (Box and Jenkins) ที่เสนอโดย George E.P. Box และ Gwilym M. Jenkins ในปี ค.ศ. 1970 ในการสร้างแบบจำลองพยากรณ์อุปสงค์การท่องเที่ยวในประเทศไทย เนื่องจากเป็นวิธีการวิเคราะห์ข้อมูลตัวแปรเดียวที่ได้

กับความนิยมในการนำมาใช้สร้างแบบจำลองพยากรณ์อุปสงค์การท่องเที่ยว (Li, Song and Witt, 2005; Song and Li, 2008) เนื่องจากมีความแม่นยำสูงในการพยากรณ์ในระยะสั้นและให้ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (mean square error: MSE) ต่ำกว่าวิธีการอื่นๆ ดังนั้นวิธีการนี้จึงเหมาะสำหรับการพยากรณ์ในระยะสั้น แต่ต้องมีขนาดของข้อมูลอนุกรมเวลาที่มากพอสมควร (Enders, 2004; Li, Song and Witt, 2005; Song and Li, 2008) โดยบทความนี้จะใช้วิธีการบ็อกซ์และเจนกินส์ทั้งที่เป็นแบบจำลอง ARIMA (AutoRegressive Integrated Moving Average) และ SARIMA (Seasonal AutoRegressive Integrated and Moving Average) ในการสร้างแบบจำลองพยากรณ์อุปสงค์การท่องเที่ยวในประเทศไทยตามความถี่ของข้อมูลอนุกรมเวลาที่แตกต่างกัน นอกจากนี้ยังได้เพิ่มตัวแปร outlier ที่เป็นอิทธิพลของเหตุการณ์ความไม่แน่นอน (irregular) ที่ไม่สามารถอธิบายได้เข้าไปในแบบจำลอง เนื่องจากเหตุการณ์ความไม่แน่นอนเป็นหนึ่งในองค์ประกอบสำคัญที่มีอิทธิพลต่ออุปสงค์การท่องเที่ยวในประเทศไทยและมีผลต่อความแม่นยำในการพยากรณ์ของแบบจำลองที่สร้างขึ้น (Akarapong et al., 2006; อัศวพงศ์ อันทองและปวีณา คำพุทะ, 2552)

ก. วิธีการบ็อกซ์และเจนกินส์ (Box and Jenkins)

แบบจำลอง ARIMA โดยทั่วไปประกอบด้วยส่วนประกอบที่สำคัญ 3 ส่วน ได้แก่ Autoregressive model (AR(p)) Integrated (I(d)) และ Moving Average model (MA(q)) โดยที่ AR(p) คือ ค่าสังเกตที่เกิดขึ้นก่อนหน้า p ค่า ($Y_{t-1}, \dots, Y_{t-p}$) ส่วน MA(q) คือ ค่าคลาดเคลื่อนจากการพยากรณ์ที่อยู่ก่อนหน้า q ค่า ($\varepsilon_{t-1}, \dots, \varepsilon_{t-q}$) สำหรับ I(d) คือ ผลต่าง (difference) ของอนุกรมเวลา ณ ปัจจุบันกับข้อมูลในอดีต d ช่วงเวลา ดังนั้นรูปแบบทั่วไปของแบบจำลอง ARIMA(p,d,q) สามารถแสดงได้ดังนี้ (Greene, 2003; Enders, 2004)

$$\phi_p(B)\Delta^d Y_t = \delta + \theta_q(B)\varepsilon_t$$

$$\text{หรือ} \quad (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1 - B)^d Y_t = \delta + (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \varepsilon_t \dots \dots \dots (1)$$

$$\text{โดยที่} \quad Y_t = \text{ค่าสังเกตของอนุกรมเวลา ณ เวลา } t$$

$$B = \text{backward shift operation โดยที่ } B^m = \Delta Y_{t-m}$$

$$D = \text{จำนวนครั้งของการหาผลต่างเพื่อให้อนุกรมเวลามีคุณสมบัติคงที่ (stationary)}$$

p = อันดับของออโตรีเกรสซีฟ (autoregressive order)
 q = อันดับของค่าเฉลี่ยเคลื่อนที่ (moving average order)
 δ = ค่าคงที่ (constant term)
 Δ^d = ผลต่างอันดับที่ d
 $\phi_1, \dots, \phi_p$ = พารามิเตอร์ของออโตรีเกรสซีฟ (autoregressive parameter)
 $\theta_1, \dots, \theta_q$ = พารามิเตอร์ของค่าเฉลี่ยเคลื่อนที่ (moving average parameter)
 ε_t = กระบวนการ white noise คือ ค่าความคลาดเคลื่อน ณ เวลา t ภายใต้ข้อสมมติว่า ค่าความคลาดเคลื่อน ณ เวลาที่แตกต่างกันเป็นตัวแปรสุ่มที่เป็นอิสระต่อกัน และมีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็นศูนย์และความแปรปรวนคงที่ [$\varepsilon_t \sim N(0, \sigma_{\varepsilon_t}^2)$]
 จากรูปแบบทั่วไปของสมการ (1) จะเห็นได้ว่า วิธีการบ็อกซ์และเจนกินส์มีเงื่อนไขบางประการเกี่ยวกับค่าพารามิเตอร์ในแบบจำลอง เพื่อให้ข้อมูลอนุกรมเวลามีคุณสมบัติคงที่ (stationary) และคุณสมบัติผกผัน (invertibility) โดยที่คุณสมบัติคงที่ (stationary) เป็นคุณสมบัติของแบบจำลอง AR(p) ที่ทำให้ $E(Y_t)$, $\text{Var}(Y_t)$ และ $\text{Cov}(Y_t, \dots, Y_{t-p})$ มีค่าคงที่ ส่วนคุณสมบัติผกผัน (invertible) เป็นคุณสมบัติของแบบจำลอง MA(q) ที่ทำให้ค่าคลาดเคลื่อนของการพยากรณ์ ε_t ในเทอมของ Y_t, Y_{t-1} มีค่าคงที่
 ในกรณีที่ข้อมูลอนุกรมเวลามีอิทธิพลของฤดูกาล สามารถประยุกต์ใช้วิธีการบ็อกซ์และเจนกินส์ได้เช่นเดียวกัน โดยเรียกแบบจำลองดังกล่าวว่า SARIMA ซึ่งมีส่วนประกอบที่สำคัญ 3 ส่วนเช่นเดียวกับแบบจำลอง ARIMA แต่มีข้อสมมติเบื้องต้น คือ ข้อมูลอนุกรมเวลาที่อยู่ภายในฤดูกาลเดียวกันต้องไม่มีสหสัมพันธ์กัน และข้อมูลอนุกรมเวลาต้องมีค่าสหสัมพันธ์ในตัวเองที่แตกต่างกันไปจากศูนย์เฉพาะช่วงห่างที่ $S, 2S, \dots, PS$ เท่านั้น ข้อสมมตินี้เป็นข้อจำกัดที่สำคัญในการนำแบบจำลอง SARIMA มาใช้งาน เนื่องจากข้อมูลอนุกรมเวลาอาจมีความสัมพันธ์ต่อกันทั้งภายในฤดูกาลเดียวกันและอาจมีความสัมพันธ์ข้ามฤดูกาลได้ (วิชิต หล่อจิระขุนทดกุล และคณะ, 2539; อัครพงศ์ อันทอง และปวีณา คำพุกกะ, 2552) ดังนั้น Box et al. (1994) จึงได้เสนอแบบจำลองข้อมูลอนุกรมเวลาที่มีฤดูกาลเชิงผลคูณ (multiplicative seasonal model) เพื่อให้สามารถวิเคราะห์ข้อมูลอนุกรมเวลาที่มีอิทธิพลของฤดูกาลทั้งตัวแบบเชิงผลบวกและตัวแบบเชิงผลคูณ รูปแบบทั่วไปของแบบจำลอง SARIMA(p,d,q)(P,D,Q)S สามารถแสดงได้ดังนี้ (Box et al., 1994; Enders, 2004; Kim and Moosa, 2005)

$$\begin{aligned}
 \phi_p(B)\Phi_p(B^S)\Delta^d\Delta_S^D Y_t &= \delta + \theta_q(B)\Theta_q(B^S)\varepsilon_t \\
 \text{หรือ } (1 - \phi_1 B^S - \dots - \phi_p B^{PS})(1 - \theta_1 B - \dots - \theta_q B^q)(1 - B^S)^D(1 - B)^d Y_t \\
 &= \delta + (1 - \theta_1 B^S - \dots - \theta_q B^{QS})(1 - \theta_1 B - \dots - \theta_q B^q)\varepsilon_t
 \end{aligned}
 \tag{2}$$

นอกจากนี้ในกรณีที่ข้อมูลอนุกรมเวลามี outlier สามารถเพิ่มตัวแปร outlier เข้าไปในแบบจำลอง ARIMA และ SARIMA ในฐานะของตัวแปรถดถอย ซึ่งรูปแบบของ outlier ที่พบอยู่เสมอ คือ additive outliers และ level shifts (Akarapong et al., 2006)

ข. ความคงที่ของข้อมูลอนุกรมเวลา (stationary) และการทดสอบ unit root

วิธีการบ็อกซ์และเจนกินส์เป็นวิธีการที่ต้องใช้กับข้อมูลอนุกรมเวลาที่มีความคงที่ (stationary) ดังนั้นถ้าหากข้อมูลอนุกรมเวลาที่นำมาใช้ไม่มีความคงที่ (non stationary) ต้องมีการแปลงข้อมูลอนุกรมเวลาดังกล่าวให้มีความคงที่ก่อน ความคงที่ของข้อมูลอนุกรมเวลาหมายถึง ข้อมูลอนุกรมเวลาที่อยู่ในสภาวะสมดุลเชิงสถิติ (statistical equilibrium) หรือการที่คุณสมบัติทางสถิติของข้อมูลอนุกรมเวลาไม่มีการเปลี่ยนแปลงไปตามกาลเวลา ดังนั้นเมื่อสมมติให้ตัวแปร Y_t เป็นข้อมูลอนุกรมเวลาที่มีความคงที่ จะมีคุณสมบัติดังนี้

$$\begin{aligned}
 \text{Mean: } E(Y_t) &= \mu \\
 \text{Variance: } \text{Var}(Y_t) &= E(y_t - \mu)^2 = \sigma^2 \\
 \text{Covariance: } \text{Cov}(Y_t, \dots, Y_{t+k}) &= E[(y_t - \mu)(y_{t+k} - \mu)] = \gamma_k
 \end{aligned}$$

แต่ถ้าตัวแปร Y_t เป็นข้อมูลอนุกรมเวลาที่ไม่มีความคงที่ จะมีคุณสมบัติดังนี้

$$\begin{aligned}
 \text{Mean: } E(Y_t) &= t\mu \\
 \text{Variance: } \text{Var}(Y_t) &= E(y_t - \mu)^2 = t\sigma^2 \\
 \text{Covariance: } \text{Cov}(Y_t, \dots, Y_{t+p}) &= E[(y_t - \mu)(y_{t+p} - \mu)] = t\gamma_k
 \end{aligned}$$

ดังนั้นก่อนนำข้อมูลอนุกรมเวลาไปใช้สร้างแบบจำลองพยากรณ์ด้วยวิธีการบ็อกซ์และเจนกินส์ จึงต้องมีการตรวจสอบความคงที่ของข้อมูลอนุกรมเวลา ด้วยการพิจารณาจากกราฟคอเรโลแกรม (correlogram) หรือการทดสอบ unit root ที่สามารถใช้ทดสอบได้ทั้งข้อมูลอนุกรมเวลาที่ไม่อิทธิพลของฤดูกาล เช่น วิธีการ Augmented Dickey-Fuller (ADF) test (Dickey and Fuller, 1979) Phillips-Perron (PP) test (Phillips and Perron, 1988) Kwiatkowski, Phillips, Schmidt and Shin (KPSS) test (Kwiatkowski et al., 1992) เป็นต้น และข้อมูลอนุกรมเวลาที่มีอิทธิพลของฤดูกาล เช่น HEGY test (Hylleberg et al.,

1990) เป็นต้น ในบทความนี้ใช้วิธี KPSS test ในการทดสอบความคงที่ของข้อมูลอนุกรมเวลาที่ไม่มียุทธพลของฤดูกาล (ข้อมูลรายปี) และใช้วิธี HEGY test ในการทดสอบความคงที่ของข้อมูลอนุกรมเวลาที่มียุทธพลของฤดูกาล (ข้อมูลรายเดือน) ก่อนที่จะอธิบายถึงวิธีการทั้งสอง ขอกล่าวถึงแนวคิดของการทดสอบ unit root พอสังเขป โดยการพิจารณาจากกระบวนการ AR(1) ที่เป็นพื้นฐาน ดังนี้ (Davidson and MacKinnon, 1993; Hamilton, 1994; Hayashi, 2000)

$$Y_t = \rho Y_{t-1} + X_t' \delta + \varepsilon_t \dots\dots\dots (3)$$

โดยที่ X_t คือ ตัวแปรภายนอก ซึ่งอาจเป็นค่าคงที่ หรือ ค่าแนวโน้ม หรือทั้งสองก็ได้ ส่วน ρ และ δ คือ ค่าพารามิเตอร์ที่ได้จากการประมาณค่า ส่วน ε_t คือ กระบวนการ white noise ดังนั้นถ้า $|\rho| \geq 1$ ข้อมูลอนุกรมเวลา Y_t จะไม่มีความคงที่ และค่าความแปรปรวนของ Y_t จะเพิ่มขึ้นตามเวลา แต่ถ้าค่า $|\rho| < 1$ ข้อมูลอนุกรมเวลาจะมีความคงที่ ดังนั้นสมมติฐานที่ใช้ในการทดสอบความคงที่ของข้อมูลอนุกรมเวลาสามารถพิจารณาได้จากการทดสอบค่าสมมุติฐานของ ρ ว่า น้อยกว่า 1 หรือไม่ โดยมีสมมติฐานหลักและสมมติฐานทางเลือก คือ $H_0: \rho = 1$ และ $H_a: \rho \neq 1$ ถ้าปฏิเสธสมมติฐานหลัก แสดงว่า $\rho < 1$ ดังนั้น Y_t มีคุณสมบัติคงที่ หรือมี integration of order zero แต่ถ้าไม่สามารถปฏิเสธสมมติฐานหลัก แสดงว่า Y_t มีคุณสมบัติไม่คงที่

วิธีการ KPSS test ที่เสนอโดย Kwiatkowski et al. (1992) เป็นหนึ่งในวิธีการทดสอบ unit root ที่นิยมในปัจจุบัน เนื่องจากให้ผลการทดสอบที่เที่ยงตรงมากกว่าวิธีการอื่นๆ วิธีการนี้แตกต่างจากวิธีการอื่นๆ โดยค่าสถิติ KPSS ถูกคำนวณจากส่วนที่เหลือ (residual) จากการถดถอยด้วยวิธีการกำลังสองน้อยที่สุด (ordinary least square) ระหว่างตัวแปร Y_t กับตัวแปร X_t ดังนี้

$$Y_t = X_t' \delta + u_t \dots\dots\dots (4)$$

จากสมการ (4) สามารถคำนวณหาค่าสถิติ KPSS ได้จาก

$$KPSS = \frac{\sum_{t=1}^T S_t^2}{\hat{\sigma}_\varepsilon^2}$$

$$\text{โดยที่ } S_t = \sum_{i=1}^t \hat{u}_i ; (\hat{u}_t = Y_t - X_t' \delta)$$

$$\hat{\sigma}_\varepsilon^2 = T^{-1} \sum_{t=1}^T \hat{u}_t^2$$

สำหรับในกรณีที่ค่าคลาดเคลื่อน ($\hat{u}_t$) มีสหสัมพันธ์เชิงอันดับ (serial correlation)

ให้สูตร $\hat{\sigma}_\varepsilon^2$ นี้แทน

$$\hat{\sigma}_\varepsilon^2 = T^{-1} \sum_{t=1}^T \hat{u}_t^2 + 2 \sum_{j=1}^q \omega_j \left(T^{-1} \sum_{t=k+1}^T \hat{u}_t \hat{u}_{t-k} \right)$$

โดยที่ $\omega_j = 1 - \frac{j}{q+1}$ สำหรับ q (Lag length) อาจเป็น $q_4 \approx 4(T/100)^{1/4}$

หรือ $q_2 \approx 12(T/100)^{1/4}$

โดยมีสมมติฐานหลักและสมมติฐานทางเลือกของการทดสอบ คือ $H_0: \sigma_v^2 = 0$ $H_a: \sigma_v^2 \neq 0$ เมื่อ $X_t = X_{t-1} + v_t$ ถ้าปฏิเสธสมมติฐานหลัก แสดงว่า $\sigma_v^2 > 0$ หมายความว่า $y_t \sim I(0)$ แต่ถ้าไม่สามารถปฏิเสธสมมติฐานหลัก แสดงว่า $y_t \sim I(1)$ วิธีการข้างต้นไม่สามารถใช้ได้ในกรณีที่ข้อมูลอนุกรมเวลามียุทธพลของฤดูกาล ดังนั้น Hylleberg et al. (1990) ได้เสนอวิธีการทดสอบ seasonal unit root สำหรับข้อมูลรายไตรมาส และต่อมา Franses (1991) และ Beaulieu and Miron (1993) ได้พัฒนามาใช้ในการทดสอบกับข้อมูลรายเดือน โดยแบบจำลองพื้นฐานที่ใช้ในการทดสอบ seasonal unit root ของข้อมูลรายเดือนมีลักษณะดังนี้

$$\begin{aligned} \Delta_{12} Y_t = & \pi_1 Y_{1,t-1} + \pi_2 Y_{2,t-1} + \pi_3 Y_{3,t-1} + \pi_4 Y_{3,t-2} + \pi_5 Y_{4,t-1} + \pi_6 Y_{4,t-2} \\ & + \pi_7 Y_{5,t-1} + \pi_8 Y_{5,t-2} + \pi_9 Y_{6,t-1} + \pi_{10} Y_{6,t-2} + \pi_{11} Y_{7,t-1} + \pi_{12} Y_{7,t-2} \\ & + \sum_{j=1}^p \phi_j \Delta_{12} Y_{t-j} + \varepsilon_t \dots\dots\dots (5) \end{aligned}$$

$$\begin{aligned} \text{โดยที่ } Y_{1,t} &= (1+L)(1+L^2)(1+L^4+L^8)Y_t \\ Y_{2,t} &= -(1-L)(1+L^2)(1+L^4+L^8)Y_t \\ Y_{3,t} &= -(1-L^2)(1+L^4+L^8)Y_t \\ Y_{4,t} &= -(1-L^4)(1-\sqrt{3}L+L^2)(1+L^4+L^8)Y_t \\ Y_{5,t} &= -(1-L^4)(1+\sqrt{3}L+L^2)(1+L^4+L^8)Y_t \\ Y_{6,t} &= -(1-L^4)(1-L^2+L^4)(1-L+L^2)Y_t \\ Y_{7,t} &= -(1-L^4)(1-L^2+L^4)(1+L+L^2)Y_t \end{aligned}$$

$\Delta_{12}Y_{t-j}$ = ตัวแปรล่า (lag) ของอนุกรมเวลา $\Delta_{12}Y_t$ ณ เวลาที่ $t-j$
 $\pi_1, \dots, \pi_{12}, \phi_j$ = ค่าพารามิเตอร์ที่ได้จากการประมาณค่า
 ε_t = กระบวนการ white noise ที่เป็นตัวแปรสุ่มที่เป็นอิสระต่อกัน และ
 มีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็นศูนย์และความแปรปรวนคงที่
 $[\varepsilon_t \sim N(0, \sigma_{\varepsilon_t}^2)]$

นอกจากนี้สามารถเพิ่มองค์ประกอบเชิงกำหนด (deterministic components) ที่ประกอบด้วยค่าคงที่ (intercept) ตัวแปรหุ่นของความเป็นฤดูกาล 11 ตัวแปร และค่าแนวโน้มเวลา (time trend) เข้าไปในสมการ (5) สำหรับจำนวนตัวแปรล่าที่ใช้ในสมการ (5) พิจารณาจากค่า Akaike Information Criterion (AIC) Hanan-Quinn Criterion (HQC) Schwarz Criterion (SC) และ Final Prediction Error (FPE) (Lütkepohl and Krätzing, 2004) โดยมีสมมติฐานที่ใช้ในการทดสอบ คือ $H_0: \pi_1 = 0$ สำหรับการทดสอบ regular unit root และ $H_0: \pi_i = 0$ ($i = 2, \dots, 12$) สำหรับการทดสอบ seasonal unit root โดยใช้ค่าสถิติ t และ F ในการคำนวณและนำค่าที่ได้ไปเปรียบเทียบกับค่าวิกฤติ (critical values) ที่เสนอโดย Franses and Hobijn (1997)

ในกรณีที่ข้อมูลอนุกรมเวลาที่ใช้ในการสร้างแบบจำลองพยากรณ์ด้วยวิธีการบ็อกซ์และเจนกินส์มีคุณสมบัติไม่คงที่ หรือ มี unit root ทั้งที่เป็น regular หรือ seasonal จะต้องแปลงข้อมูลอนุกรมเวลาดังกล่าวให้มีความคงที่ โดยการหาผลต่างของข้อมูลอนุกรมเวลาจนกระทั่งข้อมูลอนุกรมเวลามีคุณสมบัติคงที่ ซึ่งในที่สุดแล้วจะทราบถึงจำนวนครั้งของการหาผลต่างทั้งที่เป็น regular integrated และ seasonal integrated ที่ใช้ในการแปลงข้อมูลอนุกรมเวลาให้มีความคงที่ นอกจากนี้นักศึกษายังสามารถเปลี่ยนข้อมูลอนุกรมเวลาที่มีความแปรปรวนไม่คงที่ด้วย natural logarithm (อัครพงศ์ อันทอง และปวีณา คำพุทะ, 2552)

ค. ขั้นตอนการพยากรณ์ด้วยวิธีการบ็อกซ์และเจนกินส์

โดยทั่วไปวิธีการบ็อกซ์และเจนกินส์มีขั้นตอนพื้นฐานสำหรับการพยากรณ์ 4 ขั้นตอน ดังนี้ (Gujarati, 1995; ทรงศิริ แด่สมบัติ, 2539; อัครพงศ์ อันทองและปวีณา คำพุทะ, 2552)

1. การกำหนดรูปแบบ (identification): หารูปแบบ AR และ MA ที่เหมาะสมกับข้อมูลอนุกรมเวลา โดยพิจารณาจากคอเรโลแกรมของค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเอง (Autocorrelation Function: ACF) และค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเองบางส่วน (Partial Correlation Function: PACF) ของข้อมูลอนุกรมเวลา

2. การประมาณพารามิเตอร์ (parameters estimation): เป็นการประมาณพารามิเตอร์ของรูปแบบที่กำหนด วิธีที่นิยม คือ วิธีการกำลังสองน้อยที่สุด (Ordinary Least Squares: OLS) และวิธีการภาวะความน่าจะเป็นสูงสุด (Maximum Likelihood Estimation: MLE)
3. การตรวจสอบรูปแบบ (diagnostic checking): เป็นการตรวจสอบว่า รูปแบบที่กำหนดมีความเหมาะสมหรือไม่ และค่าคลาดเคลื่อนมีลักษณะ white noise หรือไม่ โดยการพิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเองของค่าคลาดเคลื่อน หรือค่าสถิติ Q (Ljung-Box Statistics: LB) การทดสอบค่าพารามิเตอร์ด้วย t-test และการพิจารณา goodness of fit ของแบบจำลองด้วยค่า stationary R2 และ R2
4. การพยากรณ์ (forecasting): นำแบบจำลองที่ผ่านการตรวจสอบไปพยากรณ์ค่าในอนาคต สามารถทำได้ทั้งการพยากรณ์แบบจุด (point forecast) และการพยากรณ์แบบช่วง (interval forecast)

ง. การทดสอบความแม่นยำของการพยากรณ์

วิธีการตรวจสอบความแม่นยำของแบบจำลองพยากรณ์มีหลายวิธีการ แต่วิธีการที่นิยม ได้แก่ ค่าเฉลี่ยของค่าสัมบูรณ์ของความคลาดเคลื่อน (Mean Absolute Error: MAE) ค่าเฉลี่ยของค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน (Mean Absolute Percentage Error: MAPE) รากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (Root Mean Square Error: RMSE) และรากที่สองของค่าเฉลี่ยของเปอร์เซ็นต์ของความคลาดเคลื่อนกำลังสอง (Root Mean Square Percentage Error: RMSPE) (Witt and Witt, 1995; Lim and McAleer, 2003; Enders, 2004; Li, Song and Witt, 2005; Kim and Moosa, 2005; Song and Li, 2008)

ดังนั้นบทความนี้จึงใช้วิธีการทั้ง 4 ในการทดสอบความแม่นยำของการพยากรณ์รูปทรงการท่องเที่ยวในประเทศไทยภายใต้ลักษณะของข้อมูลที่แตกต่างกัน โดยมีสูตรในการคำนวณของแต่ละวิธีดังนี้ (Gujarati, 1995; Enders, 2004; Lütkepohl and Krätzing, 2004; Quantitative Micro Software, 2008)

$$\text{ค่าเฉลี่ยของค่าสัมบูรณ์ของความคลาดเคลื่อน: MAE} = \frac{\sum_{t=k}^{k+n} \varepsilon_t}{n}$$

- ค่าเฉลี่ยของค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน:

$$MAPE = \frac{\sum_{t=k}^{k+n} \left(\frac{|\epsilon_t|}{Y_t} \right)}{n} \times 100$$

- รากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง:

$$RMSE = \sqrt{\frac{\sum_{t=k}^{k+n} (\epsilon_t)^2}{n}}$$

- รากที่สองของค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนกำลังสอง:

$$RMSPE = \sqrt{\frac{\sum_{t=k}^{k+n} \left(\frac{\epsilon_t}{Y_t} \right)^2}{n}} \times 100$$

3. ผลการศึกษาเชิงประจักษ์

วัตถุประสงค์หลักของบทความนี้ เพื่อตรวจสอบความแม่นยำของการพยากรณ์อุปสงค์การท่องเที่ยวในประเทศไทย ภายใต้ลักษณะของข้อมูลที่แตกต่างกัน ทั้งความแตกต่างของลักษณะเฉพาะของนักท่องเที่ยว (ประเทศที่อาศัย) และความถี่ของข้อมูล (ข้อมูลรายปีและรายเดือน) ด้วยการวิเคราะห์แบบ ex-post forecast โดยข้อมูลที่ใช้ในการสร้างแบบจำลองพยากรณ์เป็นข้อมูลจำนวนนักท่องเที่ยวต่างชาติที่เดินทางมาท่องเที่ยวในประเทศไทยระหว่างปี พ.ศ. 2528-2551 ที่แยกย่อยตามประเทศที่อาศัยจำนวน 22 ประเทศ และเป็นข้อมูลรายเดือนจำนวน 288 เดือน หรือข้อมูลรายปีจำนวน 24 ปี จากรายงานของการท่องเที่ยวแห่งประเทศไทย โดยในการศึกษาใช้ข้อมูลระหว่างปี พ.ศ. 2528-2548 (21 ปี หรือ 252 เดือน) ในการสร้างแบบจำลอง ARIMA และ SARIMA ของนักท่องเที่ยวต่างชาติรายประเทศ และนักท่องเที่ยวต่างชาติรวมทั้งหมด ต่อมานำค่าพยากรณ์ที่ได้จากแบบจำลองดังกล่าวไปเปรียบเทียบกับข้อมูลที่เหลืออีก 3 ปี (พ.ศ. 2549-2551) โดยใช้ MAE MAPE RMSE และ RMSPE ในการทดสอบความแม่นยำของการพยากรณ์

จากการทดสอบความคงที่ของข้อมูลรายปีด้วย KPSS test เพื่อหาจำนวน integration (I) พบว่า ส่วนใหญ่ข้อมูลมีความคงที่ที่ I(1) หรือมีคุณสมบัติคงที่ ณ ผลต่างอันดับที่ 1 ยกเว้นข้อมูลของประเทศเอเชียได้อื่นๆ ที่มีความคงที่ที่ I(0) ส่วนข้อมูลของ

ประเทศสวีเดนและตะวันออกกลางมีความคงที่ที่ I(2) ในขณะที่การทดสอบความคงที่ของข้อมูลรายเดือนด้วย HEGY test พบว่า ส่วนใหญ่ข้อมูลมีความคงที่ที่ I(1,1) หรือมีคุณสมบัติคงที่ ณ ผลต่างอันดับที่ 1 (สำหรับ regular) และอันดับที่ 12 (สำหรับ seasonal) สำหรับประเทศที่มีความคงที่ที่ I(1,0) หรือไม่มี seasonal unit root เช่น ประเทศมาเลเซีย สหรัฐอเมริกา เป็นต้น นอกจากนี้มีบางประเทศมีความคงที่ที่ I(0,1) หรือไม่มี regular unit root เช่น ประเทศฝรั่งเศส เยอรมนี เป็นต้น (ดูรายละเอียดในตารางที่ 1)

จากการทดสอบ unit root และ seasonal unit root ทำให้ทราบอันดับของผลต่าง (d และ D) ของแบบจำลอง ARIMA และ SARIMA ตามลำดับ และเมื่อพิจารณา ACF และ PACF ของข้อมูลรายปี พบว่า ข้อมูลทั้งหมดมีรูปแบบ MA(0) และส่วนใหญ่มีรูปแบบ AR(0) เช่นเดียวกัน ยกเว้น ข้อมูลของประเทศเอเชียตะวันออกอื่นๆ สหราชอาณาจักร อินเดีย และเอเชียได้อื่นๆ ที่มีรูปแบบ AR(1) นอกจากนี้ข้อมูลส่วนใหญ่มีองค์ประกอบของ outlier ดังได้เพิ่มตัวแปร outlier เข้าไปในแบบจำลอง (เฉพาะชุดข้อมูลที่มี outlier เท่านั้น) หลังจากประมาณค่าสัมประสิทธิ์ของแบบจำลองที่กำหนดด้วยวิธีการ OLS พบว่า ค่าสัมประสิทธิ์ต่างๆ ของแบบจำลองมีค่าแตกต่างไปจากศูนย์ที่ระดับนัยสำคัญทางสถิติที่ 0.10 (ทดสอบด้วย t-test) โดยค่า R² มีค่าเข้าใกล้ 1 เป็นส่วนใหญ่ ส่วนค่า stationary R² มีค่าแตกต่างจาก R² ในระดับที่สามารถยอมรับได้ นอกจากนี้จากการทดสอบค่าคลาดเคลื่อนด้วยค่าสถิติ Ljung-Box (LB) พบว่า ค่าสถิติดังกล่าวมีค่าน้อยกว่าค่าวิกฤตของ Chi-square ที่ระดับนัยสำคัญทางสถิติที่ 0.10 ดังนั้นสามารถสรุปได้ว่า รูปแบบของแบบจำลองพยากรณ์ที่กำหนดเป็นรูปแบบที่เหมาะสม และค่าคลาดเคลื่อนมีลักษณะ white noise ตามข้อสมมติเบื้องต้นของวิธีการนี้ (ดูรายละเอียดในตารางที่ 2)

ตารางที่ 1 ผลการทดสอบ Unit Root และ Seasonal Unit Root ของข้อมูลจำนวนนักท่องเที่ยวต่างชาติที่เดินทางมาท่องเที่ยวในประเทศไทยระหว่าง พ.ศ. 2528 - 2548

ประเทศ	Unit Root (KPSS test)		Seasonal Unit Root (HEGY test)				
	KPSS statistic	Integration	$t(\pi_1)$	$F(\pi_{2,\dots,12})$	Lags	LB statistic	Integration
รวมทุกประเทศ	0.1191*	I(1)	-0.6581	1.4119	36	2.4784	I(1,1)
1. เอเชีย	0.0776	I(1)	-0.3592	2.6257***	14	1.4847	I(1,0)
2. ยุโรป	0.1277*	I(1)	-1.4235	0.3011	25	1.2261	I(1,1)

ตารางที่ 1 (ต่อ) ผลการทดสอบ Unit Root และ Seasonal Unit Root ของข้อมูลจำนวนนักท่องเที่ยวชาวต่างชาติที่เดินทางมาท่องเที่ยวในประเทศไทยระหว่าง พ.ศ. 2528 - 2548

ประเทศ	Unit Root (KPSS test)		Seasonal Unit Root (HEGY test)				
	KPSS statistic	Integration	$t(\tau_1)$	$F(\tau_{2,\dots,12})$	Lags	LB statistic	Integration
3. จีน	0.1052	I(1)	-0.3306	2.5625**	43	1.7881	I(1,0)
4. ฮองกง	0.0603	I(1)	-0.5364	0.8977	26	1.2408	I(1,1)
5. ญี่ปุ่น	0.0942	I(1)	-0.4270	0.9512	16	1.8218	I(1,1)
6. เกาหลีใต้	0.0849	I(1)	-0.1481	1.5142	21	0.5745	I(1,1)
7. ไต้หวัน	0.1236	I(1)	-1.5392	2.0762	35	1.5452	I(1,1)
8. เอเชียตะวันออกเฉียง	0.1862**	I(1)	-3.4391***	1.6434	14	1.1069	I(0,1)
9. ฝรั่งเศส	0.1458*	I(1)	-3.9024**	0.6356	13	2.9339	I(0,1)
10. เยอรมนี	0.1915**	I(1)	-2.4380	1.3939	24	0.4726	I(1,1)
11. สวีเดน	0.1332*	I(2)	-1.8089	2.0775*	37	0.8899	I(1,1)
12. สหราชอาณาจักร	0.1831**	I(1)	-3.4070**	1.0163	25	1.8466	I(0,1)
13. ยุโรปอื่นๆ	0.0811	I(1)	-3.6029**	0.9937	37	0.8647	I(0,1)
14. แคนาดา	0.1290*	I(1)	-0.6153	0.3849	25	3.6469	I(1,1)
15. สหรัฐอเมริกา	0.1546**	I(1)	-3.5135**	0.5313	13	3.3919	I(0,1)
16. อเมริกาอื่นๆ	0.0610	I(1)	-0.1543	9.5486***	1	4.6318	I(1,0)
17. อินเดีย	0.2118**	I(1)	-1.4019	1.8922	25	2.5762	I(1,1)
18. เอเชียใต้อื่นๆ	0.2309**	I(0)	-1.9145	5.6962	1	5.5161	I(1,0)
19. ออสเตรเลีย	0.8360	I(1)	-2.0393	2.2019*	13	0.9266	I(1,1)
20. โอเชียเนียอื่นๆ	0.1398*	I(1)	-0.7225	3.255***	17	0.8778	I(1,0)
21. ตะวันออกกลาง	0.1995**	I(2)	-2.5435	4.2521***	30	3.8939	I(1,0)
22. แอฟริกา	0.0812	I(1)	-1.2563	1.3240	26	1.8539	I(1,1)

หมายเหตุ: *** ระดับนัยสำคัญทางสถิติที่ 0.01, ** ระดับนัยสำคัญทางสถิติที่ 0.05 และ * ระดับนัยสำคัญทางสถิติที่ 0.10.

ที่มา: จากการคำนวณ.

สำหรับผลการพิจารณา ACF และ PACF ของข้อมูลรายเดือน พบว่า แบบจำลอง SARIMA มีรูปแบบที่หลากหลาย เช่น SARIMA(0,1,1)(0,1,1)₁₂ SARIMA(1,0,1)(1,1,1)₁₂ เป็นต้น และข้อมูลทุกชุดมีองค์ประกอบของ outlier เนื่องจากข้อมูลรายเดือนสามารถสะท้อนอิทธิพลของเหตุการณ์ความไม่แน่นอนได้ง่ายกว่าข้อมูลรายปี หลังจากประมาณค่าสัมประสิทธิ์ของแบบจำลอง SARIMA(p,d,q)(P,D,Q)₁₂ ที่รวมตัวแปร outlier ด้วยวิธีการ OLS พบว่า ค่าสัมประสิทธิ์ที่ประมาณค่าได้มีค่าแตกต่างไปจากศูนย์ที่ระดับนัยสำคัญทางสถิติที่ 0.10 (ทดสอบด้วย t-test) โดยค่า R2 ส่วนใหญ่มีค่าเกิน 0.90 และค่า stationary R2 มีค่าแตกต่างจาก R2 ในระดับที่น้อยกว่าแบบจำลอง ARIMA นอกจากนี้ จากการทดสอบค่าคลาดเคลื่อนด้วยค่าสถิติ Ljung-Box (LB) พบว่า ค่าสถิติดังกล่าวมีค่าน้อยกว่าค่าวิกฤตของ Chi-square ที่ระดับนัยสำคัญทางสถิติที่ 0.10 ดังนั้นสามารถสรุปได้ว่า รูปแบบของแบบจำลองพยากรณ์ที่กำหนดเป็นรูปแบบที่มีความเหมาะสม และค่าคลาดเคลื่อนจากรูปแบบดังกล่าวมีลักษณะ white noise ตามข้อสมมติของวิธีการนี้ด้วยเช่นกัน (ดูรายละเอียดในตารางที่ 3) ส่วนผลการพิจารณาความแม่นยำของการพยากรณ์โดยใช้ข้อมูลที่มีลักษณะเฉพาะของนักท่องเที่ยวที่แตกต่างกัน พบว่า แบบจำลองที่ใช้ข้อมูลที่มีลักษณะแยกย่อยรายประเทศให้ผลการพยากรณ์ที่แม่นยำมากกว่าแบบจำลองที่ใช้ข้อมูลที่มีลักษณะรวมทั้งหมด โดยตัวชี้วัดทั้ง 4 ตัว ให้ผลลัพธ์ในทิศทางเดียวกัน และแบบจำลองที่ใช้ข้อมูลที่มีลักษณะแยกย่อยรายประเทศ ยังสามารถลดความคลาดเคลื่อนจากการพยากรณ์ได้มากกว่าแบบจำลองที่ใช้ข้อมูลที่มีลักษณะรวมทั้งหมดประมาณ 6-9% สำหรับผลการพิจารณาความแม่นยำของการพยากรณ์ที่ใช้ข้อมูลที่มีความถี่ที่แตกต่างกัน พบว่า แบบจำลองที่ใช้ข้อมูลรายเดือนหรือข้อมูลที่มีความถี่ที่สูงกว่า (higher frequency data) ให้ผลการพยากรณ์ที่มีความแม่นยำมากกว่าแบบจำลองที่ใช้ข้อมูลรายปีหรือข้อมูลความถี่ที่ต่ำกว่า (lower frequency data) โดยแบบจำลองที่ใช้ข้อมูลรายเดือนสามารถลดความคลาดเคลื่อนจากการพยากรณ์ได้มากกว่าแบบจำลองที่ใช้ข้อมูลรายปีสูงถึง 113-122% นอกจากนี้เมื่อมีการผสม (combination) ลักษณะของข้อมูลเข้าด้วยกัน พบว่า แบบจำลองพยากรณ์ที่ใช้ข้อมูลที่มีลักษณะแยกย่อยรายประเทศและเป็นข้อมูลรายเดือนให้ผลการพยากรณ์ที่มีความแม่นยำมากกว่าแบบจำลองพยากรณ์ที่ใช้ข้อมูลในลักษณะอื่นๆ ทั้งหมด ดังแสดงในตารางที่ 4

ตารางที่ 2 รูปแบบและผลการตรวจสอบความเหมาะสมของแบบจำลอง ARIMA (p,d,q)

ประเทศ	แบบจำลอง ARIMA	จำนวน Outlier	Stationary R ²	R ²	Ljung-Box statistic (LB)
รวมทุกประเทศ	(0,1,0)	1	0.614	0.986	19.661 (DF=18, P-value=0.389)
1. มาเลเซีย	(0,1,0)	0	0.412	0.873	28.631 (DF=18, P-value=0.063)
2. สิงคโปร์	(0,1,0)	1	0.514	0.979	15.534 (DF=18, P-value=0.624)
3. จีน	(0,1,0)	2	0.636	0.972	14.002 (DF=18, P-value=0.728)
4. ฮองกง	(0,1,0)	1	0.516	0.818	20.648 (DF=18, P-value=0.297)
5. ญี่ปุ่น	(0,1,0)	1	0.505	0.960	17.381 (DF=18, P-value=0.497)
6. เกาหลีใต้	(0,1,0)	2	0.658	0.956	11.339 (DF=18, P-value=0.879)
7. ไต้หวัน	(0,1,0)	0	0.356	0.631	19.391 (DF=18, P-value=0.368)
8. เอเชียตะวันออกเฉียง	(1,1,0)	1	0.443	0.977	22.506 (DF=17, P-value=0.168)
9. ฝรั่งเศส	(0,1,0)	1	0.340	0.864	9.7089 (DF=18, P-value=0.940)
10. เยอรมนี	(0,1,0)	1	0.445	0.967	15.212 (DF=18, P-value=0.647)
11. สวีเดน	(0,2,0)	2	0.660	0.988	14.164 (DF=18, P-value=0.718)
12. สหราชอาณาจักร	(1,1,0)	0	0.581	0.967	15.227 (DF=18, P-value=0.646)
13. ยุโรปอื่นๆ	(0,1,0)	2	0.800	0.990	27.559 (DF=18, P-value=0.069)
14. แคนาดา	(0,2,0)	0	0.312	0.910	24.976 (DF=18, P-value=0.126)
15. สหรัฐอเมริกา	(0,1,0)	1	0.583	0.969	12.130 (DF=18, P-value=0.840)
16. อเมริกาอื่นๆ	(0,1,0)	4	0.916	0.964	17.193 (DF=18, P-value=0.509)
17. อินเดีย	(1,1,0)	3	0.766	0.975	13.857 (DF=17, P-value=0.677)
18. เอเชียใต้อื่นๆ	(1,0,0)	0	0.274	0.274	28.839 (DF=17, P-value=0.038)
19. ออสเตรเลีย	(0,1,0)	1	0.562	0.907	9.958 (DF=18, P-value=0.933)
20. โอเชียเนียอื่นๆ	(0,1,0)	0	0.486	0.927	19.558 (DF=18, P-value=0.358)
21. ตะวันออกกลาง	(0,2,0)	1	0.874	0.934	9.929 (DF=18, P-value=0.934)
22. แอฟริกา	(0,1,0)	1	0.464	0.926	12.118 (DF=18, P-value=0.841)

ที่มา: จากการคำนวณ.

ตารางที่ 3 รูปแบบและผลการตรวจสอบความเหมาะสมของแบบจำลอง SARIMA (p,d,q)(P,D,Q)12

ประเทศ	แบบจำลอง SARIMA	จำนวน outlier	stationary R ²	R ²	Ljung-Box statistic (LB)
รวมทุกประเทศ	(2,1,1)(0,1,1) ₁₂	7	0.787	0.981	26.623 (DF=13, P-value=0.140)
มาเลเซีย	(0,1,1)(0,0,1) ₁₂	5	0.694	0.766	17.676 (DF=16, P-value=0.343)
สิงคโปร์	(0,1,1)(0,1,1) ₁₂	4	0.687	0.946	27.320 (DF=16, P-value=0.038)
จีน	(0,1,1)(1,0,1) ₁₂	8	0.532	0.889	21.725 (DF=15, P-value=0.115)
ฮ่องกง	(0,1,1)(0,1,1) ₁₂	10	0.832	0.814	31.198 (DF=16, P-value=0.013)
ญี่ปุ่น	(0,1,3)(0,1,1) ₁₂	6	0.537	0.961	22.948 (DF=16, P-value=0.115)
เกาหลีใต้	(0,1,1)(0,1,1) ₁₂	8	0.729	0.945	23.945 (DF=16, P-value=0.091)
ไต้หวัน	(0,1,1)(0,1,1) ₁₂	4	0.646	0.855	16.464 (DF=16, P-value=0.421)
เอเชียตะวันออกเฉียง	(1,0,1)(1,1,0) ₁₂	7	0.606	0.968	12.669 (DF=15, P-value=0.628)
ฝรั่งเศส	(0,0,3)(1,1,0) ₁₂	7	0.642	0.942	27.930 (DF=14, P-value=0.015)
เยอรมนี	(0,1,2)(0,1,1) ₁₂	4	0.596	0.966	27.598 (DF=15, P-value=0.024)
สวีเดน	(1,1,2)(1,1,0) ₁₂	4	0.565	0.984	23.035 (DF=15, P-value=0.105)
สหราชอาณาจักร	(1,0,1)(0,1,1) ₁₂	2	0.566	0.975	12.387 (DF=15, P-value=0.650)
ยุโรปอื่นๆ	(1,0,0)(0,1,1) ₁₂	3	0.611	0.974	18.595 (DF=16, P-value=0.290)
แคนาดา	(0,1,2)(1,1,0) ₁₂	5	0.507	0.958	5.694 (DF=15, P-value=0.984)
สหรัฐอเมริกา	(1,0,0)(0,1,1) ₁₂	5	0.774	0.969	14.081 (DF=16, P-value=0.593)
อเมริกาอื่นๆ	(1,1,1)(0,0,1) ₁₂	3	0.697	0.785	20.311 (DF=15, P-value=0.160)
อินเดีย	(0,1,2)(0,1,1) ₁₂	9	0.791	0.965	28.509 (DF=15, P-value=0.019)
เอเชียใต้อื่นๆ	(2,1,0)(1,0,1) ₁₂	4	0.735	0.759	8.039 (DF=14, P-value=0.887)
ออสเตรเลีย	(0,1,2)(0,1,1) ₁₂	2	0.440	0.950	20.200 (DF=15, P-value=0.164)
โอเชียเนียอื่นๆ	(0,1,1)(0,0,1) ₁₂	1	0.574	0.924	12.554 (DF=16, P-value=0.705)
ตะวันออกกลาง	(1,1,1)(1,0,0) ₁₂	2	0.619	0.911	20.037 (DF=15, P-value=0.171)
แอฟริกา	(0,1,1)(0,1,1) ₁₂	9	0.750	0.958	13.858 (DF=16, P-value=0.609)

ที่มา: จากการคำนวณ.

ตารางที่ 4 ขนาดของความคลาดเคลื่อนที่เกิดจากการพยากรณ์ภายใต้คุณลักษณะของข้อมูลที่แตกต่างกัน

คุณลักษณะของข้อมูล	MAE	MAPE	RMSE	RMSPE
รวมทั้งหมด (เฉลี่ย) [aggregated]	916,789	11.10	944,038	13.08
แยกประเทศ (เฉลี่ย) [disaggregated]	840,516	10.48	874,435	12.36
อัตราส่วนของความแตกต่าง	1.09	1.06	1.08	1.06
รายปี (เฉลี่ย) [annual]	1,640,666	11.56	1,665,493	11.74
รายเดือน (เฉลี่ย) [monthly]	737,838	5.21	780,497	5.62
อัตราส่วนของความแตกต่าง	2.22	2.22	2.13	2.13
รวมทั้งหมด&รายปี [aggregated & annual]	1,717,093	12.09	1,735,279	12.22
รวมทั้งหมด&รายเดือน [aggregated & monthly]	752,401	5.33	802,533	5.70
อัตราส่วนของความแตกต่าง	2.28	2.27	2.16	2.14
แยกประเทศ&รายปี [disaggregated & annual]	1,564,240	11.03	1,595,707	11.26
แยกประเทศ&รายเดือน [disaggregated & month]	723,275	5.10	758,461	5.34
อัตราส่วนของความแตกต่าง	2.16	2.16	2.10	2.11
รวมทั้งหมด&รายปี [aggregated & annual]	1,717,093	12.09	1,735,279	12.22
แยกประเทศ&รายปี [disaggregated & annual]	1,564,240	11.03	1,595,707	11.26
อัตราส่วนของความแตกต่าง	1.10	1.10	1.09	1.09
รวมทั้งหมด&รายเดือน [aggregated & monthly]	752,401	5.33	802,533	5.70
แยกประเทศ&รายเดือน [disaggregated & month]	723,275	5.10	758,461	5.34
อัตราส่วนของความแตกต่าง	1.04	1.05	1.06	1.07

ที่มา: จากการคำนวณ.

การอภิปรายผลการศึกษา

ผลการศึกษาเชิงประจักษ์แสดงให้เห็นว่า ลักษณะของข้อมูลอนุกรมเวลาที่แตกต่างกันมีผลต่อความแม่นยำในการพยากรณ์อุปสงค์การท่องเที่ยวในประเทศไทย โดยข้อมูลอนุกรมเวลาที่มีลักษณะแยกย่อยให้ผลการพยากรณ์ที่แม่นยำมากกว่าข้อมูลอนุกรมเวลาที่มีลักษณะรวมกัน ซึ่งสอดคล้องกับงานศึกษาของ Kim and Moosa (2005) แต่ขัดแย้งกับงานศึกษาของ Vu and Turner (2005) และ Kon and Turner (2005) ที่แสดงให้เห็นว่า การใช้ข้อมูลอนุกรมเวลาที่มีลักษณะรวมกันในการสร้างแบบจำลองพยากรณ์ให้ผลการพยากรณ์ที่แม่นยำมากกว่าการใช้ข้อมูลอนุกรมเวลาที่มีลักษณะแยกย่อย เนื่องจากในกรณีทั้งสองใช้ข้อมูลอนุกรมเวลาของประเทศเกาหลีใต้และสิงคโปร์ ซึ่งมีสิ่งดึงดูดใจทางด้านการท่องเที่ยวอย่างประเทศไทย และอุปสงค์การท่องเที่ยวมีลักษณะ homogeneous มากกว่าประเทศไทย ดังนั้นกรณีของประเทศไทยซึ่งมีสิ่งดึงดูดใจทางด้านการท่องเที่ยวที่หลากหลายกว่าและอุปสงค์การท่องเที่ยวมีลักษณะ heterogeneous การใช้ข้อมูลอนุกรมเวลาที่มีลักษณะแยกย่อยในการสร้างแบบจำลองพยากรณ์จะให้ผลการพยากรณ์ที่มีความแม่นยำมากกว่าเนื่องจากข้อมูลที่มีลักษณะแยกย่อย ได้จำแนกตามลักษณะเฉพาะที่แตกต่างกันของนักท่องเที่ยว จึงทำให้การพยากรณ์ภายใต้คุณลักษณะเฉพาะที่แตกต่างกันสามารถลดความคลาดเคลื่อนจากการพยากรณ์ได้มากกว่ากรณีที่ใช้ข้อมูลอนุกรมเวลาที่มีลักษณะรวมกันที่เป็นการรวมเอาความแตกต่างทั้งหมดของนักท่องเที่ยวไว้ในชุดของข้อมูลอนุกรมเวลาเดียวกัน

สำหรับการใช้ข้อมูลอนุกรมเวลาที่มีความถี่ที่สูงกว่าทำให้ผลการพยากรณ์มีความแม่นยำมากกว่า โดยเฉพาะกรณีของอุปสงค์การท่องเที่ยวในประเทศไทย ซึ่งสอดคล้องกับข้อสังเกตของ Lim and McAleer (2003) และ Song and Li (2008) ดังนั้นการใช้ข้อมูลรายเดือนจึงให้ผลการพยากรณ์ที่แม่นยำมากกว่าการใช้ข้อมูลรายปี เนื่องจากข้อมูลรายเดือนประกอบด้วยอิทธิพลของแนวโน้ม (trend) ฤดูกาล (seasonal) วัฏจักร (cycle) และเหตุการณ์ความไม่แน่นอน (irregular) ในขณะที่ข้อมูลรายปีไม่มีอิทธิพลของฤดูกาล ซึ่งอิทธิพลของฤดูกาลถือเป็นหนึ่งในอิทธิพลที่สำคัญของอุปสงค์การท่องเที่ยว (Butler, 2001) และนักท่องเที่ยวในแต่ละตลาดหรือในแต่ละประเทศมีฤดูกาลท่องเที่ยวที่แตกต่างกัน (Jeffrey and Barden, 1999; Koenig-Lewis and Bischoff, 2004; Fernandez-Morales and Mayorga-Toledano, 2008; Akarapong and Mingsarn, 2009; อัครพงศ์ อันทอง และ บัณฑิต ขาวสะอาด, 2552) ดังนั้นการสร้างแบบจำลองพยากรณ์อุปสงค์การท่องเที่ยวในประเทศไทยโดยใช้ข้อมูลอนุกรมเวลาที่ปราศจากอิทธิพลของฤดูกาล ก่อให้เกิดความคลาดเคลื่อนในการพยากรณ์มากกว่าการสร้างแบบจำลองโดยใช้ข้อมูลอนุกรมเวลาที่มี

อิทธิพลของฤดูกาล นอกจากนี้ข้อมูลอนุกรมเวลาที่มีความถี่สูงหรือเป็นข้อมูลรายเดือน ส่วนใหญ่สามารถสะท้อนความผิดปกติหรืออิทธิพลของเหตุการณ์ความไม่แน่นอนได้มากกว่า โดยอิทธิพลของเหตุการณ์ความไม่แน่นอนมีผลต่ออุปสงค์การท่องเที่ยวในแต่ละประเทศที่แตกต่างกัน เนื่องจากนักท่องเที่ยวแต่ละประเทศได้รับอิทธิพลและอ่อนไหวต่อเหตุการณ์ความไม่แน่นอนแตกต่างกัน (Akarapong et al., 2006; อัครพงศ์ อ้นทอง และปวีณา คำพุก 2552)

จากกรณีศึกษาอุปสงค์ของการท่องเที่ยวในประเทศไทย ทำให้ได้ข้อสังเกตเพิ่มเติมว่า ในการสร้างแบบจำลองพยากรณ์อุปสงค์การท่องเที่ยวควรมีการตรวจสอบข้อมูลอนุกรมเวลาก่อนนำมาใช้ เพื่อให้ผลการพยากรณ์มีความแม่นยำมากขึ้น ถ้าหากสามารถแยกย่อยข้อมูลอนุกรมเวลาตามลักษณะเฉพาะของนักท่องเที่ยว และใช้ข้อมูลที่มีค่าความถี่ที่สูงกว่าได้ จะยังทำให้แบบจำลองที่สร้างขึ้นสามารถพยากรณ์ได้แม่นยำมากขึ้น แต่อย่างไรก็ตามก่อนที่จะนำข้อมูลอนุกรมเวลามาใช้ ผู้ใช้ควรมีการตรวจสอบข้อมูลอนุกรมเวลาในเบื้องต้นดังนี้

1. ควรพิจารณาข้อมูลอนุกรมเวลาที่นำมาใช้ด้วยการเขียน (plot) กราฟของข้อมูลอนุกรมเวลาแต่ละชุด (โดยปกติจะใช้กราฟเส้น) เพื่อพิจารณาการเคลื่อนไหวของข้อมูลอนุกรมเวลา และพิจารณาว่า ข้อมูลอนุกรมเวลามีลักษณะของรายละเอียดและมีลักษณะการเคลื่อนไหวที่แตกต่างหรือไม่ รวมทั้งพิจารณาแบบแผนการเคลื่อนไหวของข้อมูลอนุกรมเวลาในเบื้องต้นว่า มีอิทธิพลของแนวโน้ม ฤดูกาล วัฏจักร และเหตุการณ์ความไม่แน่นอนหรือไม่ (เป็นการพิจารณาองค์ประกอบของข้อมูลอนุกรมเวลาในเบื้องต้น) นอกจากนี้ผู้ใช้ต้องให้ความสำคัญกับค่าสถิติพื้นฐาน เช่น ค่าเฉลี่ย ค่าความแปรปรวน ค่าสูงสุด ค่าต่ำสุด ขนาดตัวอย่าง เพื่อทราบว่า ข้อมูลอนุกรมเวลาที่นำมาใช้มีลักษณะอย่างไร มีการแจกแจงแบบปกติหรือไม่ มีลักษณะใดๆ เบี่ยงเบนหรือเบี่ยงน้อยเพียงใด และมีค่าสุดโต่งหรือไม่
2. ควรตรวจสอบองค์ประกอบของข้อมูลอนุกรมเวลาด้วยวิธีการแยกองค์ประกอบ เพื่อทราบว่า ข้อมูลอนุกรมเวลานั้นๆ มีอิทธิพลของแนวโน้ม ฤดูกาล วัฏจักร และเหตุการณ์ความไม่แน่นอนมากน้อยแตกต่างกันอย่างไร วิธีการแยกองค์ประกอบที่นิยมใช้มีอยู่ 3 วิธี ได้แก่ วิธีค่าเฉลี่ยอย่างง่าย วิธีค่าเฉลี่ยเคลื่อนที่ และวิธี Census II แต่ในกรณีที่ข้อมูลอนุกรมเวลาของอุปสงค์การท่องเที่ยวมีความผันผวนสูง ควรใช้วิธี Census II ที่เรียกว่า X-12-ARIMA เนื่องจากเป็นวิธีการที่เหมาะสมในกรณีที่ไม่สามารถกำหนดช่วงเวลาของการหาค่าเฉลี่ยได้

แน่นอน และในกรณีที่ข้อมูลมีความผันผวนสูง (Akarapong et al., 2006; Akarapong and Mingsarn, 2009; อัครพงศ์ อ้นทอง และมิ่งสรรพ ขาวสะอาด, 2552) นอกจากนี้การทราบองค์ประกอบของข้อมูลอนุกรมเวลาทำให้ผู้ใช้สามารถกำหนดรูปแบบและตัวแปรในแบบจำลองพยากรณ์ได้เหมาะสมและถูกต้อง

3. ควรตรวจสอบคุณสมบัติความคงที่ของข้อมูลอนุกรมเวลา เพื่อทราบว่า ข้อมูลอนุกรมเวลาที่นำมาใช้มีสถานะสมดุลเชิงสถิติ หรือข้อมูลอนุกรมเวลาไม่มีการเปลี่ยนแปลงคุณสมบัติทางสถิติเมื่อเวลามีการเปลี่ยนแปลง วิธีการตรวจสอบความคงที่ของข้อมูลอนุกรมเวลาที่นิยมใช้ในปัจจุบันได้แก่ การทดสอบ unit root ผู้ใช้ควรเลือกวิธีการทดสอบ unit root ให้สอดคล้องกับลักษณะของข้อมูลอนุกรมเวลา กล่าวคือ ถ้าหากข้อมูลอนุกรมเวลาเป็นข้อมูลที่ไม่มีความถี่ของฤดูกาล ผู้ใช้ควรเลือกวิธีการ unit root เช่น ADF test, PP test, KPSS test เป็นต้น แต่ถ้าข้อมูลอนุกรมเวลามีอิทธิพลของฤดูกาล ผู้ใช้ควรเลือกวิธีการทดสอบ seasonal unit root ที่เรียกว่า HEGY test เช่น วิธีการที่เสนอโดย Franses (1991) Beaulieu and Miron (1993) เป็นต้น

หลังจากการตรวจสอบข้อมูลอนุกรมเวลาในเบื้องต้น ผู้ใช้จะทราบถึงลักษณะพื้นฐานของข้อมูลอนุกรมเวลาที่นำมาใช้ เช่น องค์ประกอบ ความคงที่ เป็นต้น ซึ่งถ้าหากข้อมูลอนุกรมเวลามีอิทธิพลของฤดูกาล ผู้ใช้ควรเลือกแบบจำลอง SARIMA แทนแบบจำลอง ARIMA และถ้าหากว่าข้อมูลอนุกรมเวลามีอิทธิพลของเหตุการณ์ความไม่แน่นอน หรือมี outlier ผู้ใช้ควรเลือกแบบจำลอง intervention (อัครพงศ์ อ้นทอง และปวีณา คำพุก 2552) หรือเพิ่มตัวแปร outlier ทั้งที่เป็น additive outlier หรือ level shifts เข้าไปในแบบจำลอง เป็นต้น (Akarapong et al., 2006) นอกจากนี้หากข้อมูลอนุกรมเวลาที่ใช้ไม่มีความคงที่ ผู้ใช้ต้องแปลงข้อมูลอนุกรมเวลาให้มีความคงที่ก่อนนำมาใช้ในการสร้างแบบจำลองพยากรณ์ โดยเฉพาะแบบจำลองพยากรณ์ที่เป็นสมการถดถอยที่ประมาณค่าสัมประสิทธิ์ด้วยวิธีการ OLS เนื่องจากการใช้ข้อมูลอนุกรมเวลาที่ไม่มีความคงที่อาจทำให้เกิดความสัมพันธ์ที่ไม่แท้จริง (spurious regression) (Granger and Newbold, 1974) ซึ่งทำให้ค่าสถิติ t และค่า R² ที่คำนวณได้มีค่าสูง แต่ค่า Durbin-Watson (D.W.) มีค่าต่ำ โดย Granger and Newbold ได้ตั้งข้อสังเกตว่า ในกรณีดังกล่าวค่า R² > D.W. (Enders, 2004) นอกจากนี้ความคงที่ของข้อมูลอนุกรมเวลาถือได้ว่าเป็นข้อสมมติเบื้องต้นที่สำคัญเมื่อใช้วิธีการทางเศรษฐมิติในการวิเคราะห์ (Engle and Granger, 1987) โดยทั่วไปไม่ใช้การหาผลต่าง

ของข้อมูลอนุกรมเวลาในการแปลงข้อมูลอนุกรมเวลาให้มีความคงที่ สามารถทำได้ทั้ง regular difference (สำหรับข้อมูลอนุกรมเวลาที่มีอิทธิพลของฤดูกาล) และ seasonal difference (สำหรับข้อมูลอนุกรมเวลาที่มีอิทธิพลของฤดูกาล) (อัครพงษ์ อ้นทอง และปวีณา คำพุทะ, 2552) นอกจากนี้ถ้าหากข้อมูลอนุกรมเวลาที่มีความแปรปรวนไม่คงที่หรือมีความแปรปรวนสูง ผู้ใช้สามารถทำ natural logarithm ข้อมูลอนุกรมเวลาก่อนนำมาใช้ ซึ่งเป็นวิธีการหนึ่งที่นักเศรษฐศาสตร์นิยมนำมาใช้ลดความแปรปรวนของข้อมูลที่ใช้ในการศึกษา (Studenmund, 2006; อัครพงษ์ อ้นทอง, 2552; อัครพงษ์ อ้นทอง และมิ่งสรรพ ขาวสะอาด, 2552)

5. สรุปและข้อเสนอแนะ

บทความนี้ตรวจสอบถึงความแม่นยำของแบบจำลองพยากรณ์อุปสงค์การท่องเที่ยวในประเทศไทยภายใต้ลักษณะของข้อมูลที่แตกต่างกัน โดยใช้การวิเคราะห์แบบ ex-post forecast และใช้ข้อมูลจำนวนนักท่องเที่ยวชาวต่างชาติที่เดินทางมาท่องเที่ยวในประเทศไทยระหว่างปี พ.ศ. 2528-2551 ในการสร้างแบบจำลอง ARIMA และ SARIMA เพื่อการพยากรณ์ และใช้ MAE MAPE RMSE และ RMSPE ในการทดสอบความแม่นยำของการพยากรณ์ ผลการศึกษาพบว่า ความแม่นยำในการพยากรณ์อุปสงค์การท่องเที่ยวในประเทศไทย ส่วนหนึ่งขึ้นอยู่กับลักษณะของข้อมูล โดยข้อมูลอนุกรมเวลาที่มีลักษณะแยกย่อยรายประเทศให้ผลการพยากรณ์ที่แม่นยำมากกว่าข้อมูลอนุกรมเวลาที่มีลักษณะรวมทั้งหมด และข้อมูลรายเดือนที่มีค่าความถี่ที่สูงกว่าให้ผลการพยากรณ์ที่มีความแม่นยำมากกว่าการใช้ข้อมูลรายปีที่มีค่าความถี่ที่ต่ำกว่า เนื่องจากอุปสงค์การท่องเที่ยวในประเทศไทยมีอิทธิพลของฤดูกาล และเหตุการณ์ความไม่แน่นอน จากกรณีศึกษาอุปสงค์การท่องเที่ยวในประเทศไทย ทำให้ได้ข้อสังเกตเพิ่มเติมว่า ก่อนการใช้ข้อมูลอนุกรมเวลาในการสร้างแบบจำลองพยากรณ์อุปสงค์การท่องเที่ยว ผู้ใช้ควรพิจารณาแบบแผนการเคลื่อนไหว องค์ประกอบ และความคงที่ของข้อมูลอนุกรมเวลาก่อนนำมาสร้างแบบจำลองพยากรณ์ เนื่องจากลักษณะพื้นฐานของข้อมูลอนุกรมเวลามีผลต่อการตัดสินใจเลือกวิธีการหรือแบบจำลองพยากรณ์ นอกจากนี้หากข้อมูลอนุกรมเวลามีคุณสมบัติไม่คงที่ ผู้ใช้ต้องแปลงข้อมูลอนุกรมเวลาให้มีความคงที่ก่อนนำมาใช้ เพราะความคงที่ของข้อมูลอนุกรมเวลาเป็นข้อสมมติเบื้องต้นที่สำคัญในการใช้วิธีการพยากรณ์ และเศรษฐมิติ ดังนั้นหากผู้ใช้ละเลยการตรวจสอบลักษณะของข้อมูลอนุกรมเวลาต่างๆ เหล่านี้ อาจนำมาสู่การตัดสินใจเลือกวิธีการและแบบจำลองพยากรณ์ที่ไม่เหมาะสมหรือผิดพลาดได้ และส่งผลต่อความแม่นยำในการพยากรณ์

8 เอกสารอ้างอิง

- อัครพงษ์ อ้นทอง และปวีณา คำพุทะ. 2552. "การพยากรณ์จำนวนนักท่องเที่ยวต่างชาติที่เดินทางมาท่องเที่ยวในประเทศไทยโดยใช้แบบจำลอง SARIMA Intervention." *วารสารวิชาการ มหาวิทยาลัยอุบลราชธานี* 11 (1): 196-214.
- อัครพงษ์ อ้นทอง และมิ่งสรรพ ขาวสะอาด. 2552. "ความเป็นฤดูกาลของการท่องเที่ยวในจังหวัดเชียงใหม่." *วารสารเศรษฐศาสตร์ มหาวิทยาลัยเกษตรศาสตร์* 16 (2): 26-38.
- อัครพงษ์ อ้นทอง. 2552. "ความเอนเอียงของวิธีการดีอีเอแบบสองขั้นตอน." *วารสารเศรษฐศาสตร์ มหาวิทยาลัยเกษตรศาสตร์* 16 (1): 39-53.
- Akarapong Untong and Mingsarn Kaosa-ard. 2009. "Seasonality analysis of tourist arrivals at accommodation establishments in Chiang Mai, Thailand." *Proceeding 15th Asia Pacific Tourism Association Annual Conference*. July 9-12 2009. Incheon, Korea.
- Akarapong Untong, Pairach Piboonrungraj and Mingsarn Kaosa-ard. 2006. "The impact of world disasters on the number of international tourist arrivals to Thailand." *Proceeding 12th Asia Pacific Tourism Association and 4th APacCHRIE Joint Conference*. June 26-29 2006. Hualien, Taiwan.
- Beaulieu, J.J. and Miron, J.A. 1993. "Seasonal unit roots in aggregate US data." *Journal of Econometrics* 55 (1-2): 305-328.
- Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. 1994. *Time series analysis: Forecasting and control*. 3rd ed. Englewood Cliffs, NJ: Prentice-Hall.
- Butler, R.W. 2001. "Seasonality in tourism: Issues and implications." In Baum, T. and Lundtorp, S. (Ed.), *Seasonality in Tourism*: 5-22. Pergamon: Oxford.
- Crouch, G.I. 1994. "The study of international tourism demand: A review of findings." *Journal of Travel Research* 33 (1): 41-54.
- Davidson, R. and MacKinnon, J.G. 1993. *Estimation and Inference in Econometrics*. New York: Oxford University Press.
- Enders, W. 2004. *Applied Econometric Time Series*. 2nd ed. New York: John Wiley & Sons.
- Engle R.F. and Granger, C.W.J. 1987. "Co-integration and Error Correction: Representation, estimation, and testing." *Econometrica* 55 (2): 251-276.

- Fernandez-Morales, A. and Mayorga-Toledano M.C. 2008. "Seasonal concentration of the hotel demand in Costa del Sol: A decomposition by nationalities." **Tourism Management** 29 (5): 940-949.
- Franses, P.H. 1991. "Seasonality, non-stationarity and the forecasting of monthly time series." **International Journal of Forecasting** 7 (2): 227-208.
- Franses, P.H. and Hobijn, B. 1997. "Critical values for unit root tests in seasonal time series." **Journal of Applied Statistics** 24 (1): 25-48.
- Granger, C.W.J. and Newbold, P. 1974. "Spurious regressions in econometrics." **Journal of Econometrics** 2 (2): 111-120.
- Greene, W.H. 2003. **Econometric Analysis**. 5th ed. Prentice Hall.
- Gujarati, D. 1995. **Basic Econometrics**. 3rd ed. Boston: McGraw-Hill.
- Hair, J.F., Anderson, R.E., Tatham, R.L. and Black, W. 1998. **Multivariate Data Analysis**. 5th ed. London: Prentice-Hall International Inc.
- Hamilton, J.D. 1994. **Time Series Analysis**. New Jersey: Princeton University Press.
- Hayashi, F. 2000. **Econometrics**. New Jersey: Princeton University Press.
- Hylleberg, S., Engle, R.F., Granger, C.W.J. and Yoo, B.S. 1990. "Seasonal integration and cointegration." **Journal of Econometrics** 44 (1-2): 215-238.
- Jeffrey, D.J. and Barden, R.R.D. 1999. "An analysis of the nature, causes and marketing implications of seasonality in the occupancy performance of English hotels." **Tourism Economics** 5 (1): 69-91
- Kim, J.H. and Moosa, I.A. 2005. "Forecasting international tourist flows to Australia: a comparison between the direct and indirect methods." **Tourism Management** 26 (1): 69-78.
- Koenig-Lewis, N. and Bischoff, E.E. 2004. "Analyzing seasonality in Welsh room occupancy data." **Annals of Tourism Research** 31 (2): 374-392.
- Kon, S.C. and Turner, L. 2005. "Neural network forecasting of tourism demand." **Tourism Economics** 11 (3): 301-328.
- Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. and Shin, Y. 1992. "Testing the null hypothesis of stationarity against the alternative of a unit root." **Journal of Econometrics** 54 (1-3): 159-178.

- Li, G., Song, H. and Witt, S.F. 2005. "Recent developments in econometric modeling and forecasting." **Journal of Travel Research** 44 (1): 82-99.
- Lim, C. 1997. "Review of international tourism demand models." **Annals of Tourism Research** 24 (4): 835-849.
- Lim, C. and McAleer, M. 2000. "A seasonal analysis of Asian tourist arrivals to Australia." **Applied Economics** 32 (4): 499-509.
- Lim, C. and McAleer, M. 2003. "Modeling international travel demand from Singapore to Australia." **CIRJE Discussion Papers** CIRJE-F-214. <http://www.e.u-tokyo.ac.jp/cirje/research/03research02dp.html>, January 13, 2009.
- Lütkepohl, H. and Krätzing, M. 2004. **Applied Time Series Econometrics**. Cambridge: Cambridge University Press.
- Quantitative Micro Software, LLC. 2008. **EViews 6.0 User's Guide**.
- Song, H. and Li, G. 2008. "Tourism demand modelling and forecasting—A review of recent research." **Tourism Management** 29 (2): 203-220.
- Bludenmund, A.H. 2006. **Using Econometrics: A Practical Guide**. 5th ed. Boston: Pearson Education, Inc.
- Vu, J.C. and Turner, L. 2005. "Data disaggregation in demand forecasting." **Tourism and Hospitality Research** 6 (1): 38-52.
- Witt, B.F. and Witt, C.A. 1995. "Forecasting tourism demand: A review of empirical research." **International Journal of Forecasting** 22 (3): 447-475.