บทที่ 4

วิธีการศึกษา

ในบทนี้จะกล่าวถึงวิธีการศึกษาของงานวิจัยนี้ โดยมีขั้นตอนของวิธีการศึกษาดังนี้ ขั้นตอนแรกเป็นการพยากรณ์ปริมาณการขนถ่ายสินค้าของคลังสินค้าการบินไทย ณ ท่าอากาศยานกรุงเทพ โดยการใช้แบบจำลองการพยากรณ์ของบอกซ์และเจนกินส์ ขั้นตอนที่สองเป็นวิธีการพยากรณ์โดยใช้วิธี Decomposition ขั้นตอนที่สามเปรียบเทียบผลของการพยากรณ์โดยวัดค่าของความแม่นยำในการพยากรณ์ ขั้นตอนที่สี่การประมาณการผลผลิตจริงรายเดือน และการหาความสัมพันธ์ระหว่างปริมาณการขนส่งสินค้าส่งออกกับผลิตภัณฑ์มวลรวมภายในประเทศรายเดือน

4.1 การวิเคราะห์อนุกรมเวลาโดยวิธี Box - Jenkins

วิธีของบอกซ์และเจนกินส์มีรูปแบบของข้อมูลอนุกรมเวลาทั้งหมดได้ 3 รูปแบบ คือ รูปแบบอัตตถดถอย (Autoregressive Models : AR) รูปแบบค่าเฉลี่ยเคลื่อนที่ (Moving Average Models : MA) และรูปแบบอัตตถดถอยผสมค่าเฉลี่ยเคลื่อนที่ (Autoregressive Moving Average Models : ARMA)

1.1 รูปแบบอัตตถดถอย (Autoregressive Models : AR) กำหนดให้ ข้อมูล ณ เวลาปัจจุบันสัมพันธ์เชิงถดถอยกับข้อมูลเดียวกันในเวลาอดีต ดังนี้

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \dots + \phi_p Y_{t-p} + e_t$$

เมื่อ Y_t เป็นข้อมูลคงที่ (Stationary) ณ เวลา t

$Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$ เป็นข้อมูลตัวแปรอิสระของข้อมูลตัวแปรตาม ณ เวลาในอดีต p ช่วงเวลา

ϕ_0 เป็นค่าคงที่

$\phi_1, \ldots, \phi_p$ เป็นสัมประสิทธิ์ความถดถอยซึ่งประมาณได้โดยการหาที่ผลบวกของความคลาดเคลื่อนยกกำลังสองน้อยที่สุด

e_t ความคลาดเคลื่อน ณ เวลา t

1.2 รูปแบบค่าเฉลี่ยเคลื่อนที่ (Moving Average Models : MA) กำหนดให้ข้อมูล ณ เวลาปัจจุบันสัมพันธ์กับความคลาดเคลื่อนจากการพยากรณ์ที่เกิดขึ้นในเวลาอดีต ดังนี้

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \theta_3 e_{t-3} - \ldots - \theta_q e_{t-q} - \theta_0$$

เมื่อ Y_t เป็นข้อมูลคงที่ (Stationary) ณ เวลา t

e_t เป็นความคลาดเคลื่อน ณ เวลา t

$e_{t-1}, \ldots, e_{t-q}$ เป็นความคลาดเคลื่อนที่เกิดขึ้นในอดีต q ช่วงเวลา

$\theta_1, \ldots, \theta_q$ เป็นสัมประสิทธิ์หรือน้ำหนักซึ่งประมาณได้โดยการหาผลบวกของความคลาดเคลื่อนยกกำลังสองน้อยที่สุดเช่นกัน

θ_0 เป็นค่าคงที่

1.3 รูปแบบอัตตถดถอยผสมค่าเฉลี่ยเคลื่อนที่ (Autoregressive Moving Average Models : ARMA) กำหนดให้ข้อมูล ณ เวลาปัจจุบันสัมพันธ์กับข้อมูลในอดีตและความคลาดเคลื่อนจากการพยากรณ์ในอดีต ดังนี้

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + .. + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-q} - \theta_0$$

อย่างไรก็ตามจะเห็นว่าถ้ากำหนดให้ $W_t = Y_t - \mu$ เมื่อค่า μ เป็นค่าเฉลี่ยของอนุกรมเวลาคงที่ดังกล่าวจะได้

รูปแบบอัตตถดถอยเป็นดังนี้

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \dots + \phi_p W_{t-p} + e_t$$

ซึ่งจะไม่มีค่าคงที่ ϕ_0 ดังนั้นการหารูปแบบจึงมีสองลักษณะคือ แบบมีค่าคงที่หรือไม่มีและสำหรับรูปแบบที่มีค่าคงที่จะได้ว่า

$$\phi_0 = \mu (1 - \phi_1 - \phi_2 - \dots \phi_p)$$

ทำนองเดียวกันสำหรับรูปแบบค่าเฉลี่ยเคลื่อนที่จะเป็นดังนี้

$$W_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} - \theta_0$$

โดยที่ค่าคงที่สำหรับรูปแบบที่มีค่าคงที่จะมีค่าดังนี้

$$\theta_0 = -\mu$$

สำหรับค่าคงที่สำหรับรูปแบบอัตตถดถอยผสมค่าเฉลี่ยเคลื่อนที่จะเป็นผลรวมของค่าคงที่ของทั้งสองรูปแบบดังกล่าวแล้วรวมกันนั่นเอง

ขั้นตอนวิธีการพยากรณ์

วิธีการพยากรณ์ของบอกซ์และเจนกินส์ มีขั้นตอนหลัก ๆ ดังนี้

1. พิจารณาเลือกรูปแบบที่เหมาะสมทั้ง 3 ดังกล่าวแล้วให้กับข้อมูล Y_t สำหรับหลักเกณฑ์ในการเลือกรูปแบบใช้การพิจารณาจาก อัตตสัมพันธ์ (Autocorrelation) และ อัตตสัมพันธ์เชิงส่วน (Partial Autocorelation) ของข้อมูล ความหมายของอัตตสัมพันธ์ของข้อมูล หมายถึง สหสัมพันธ์ระหว่างข้อมูลนั้น ณ เวลาใดเวลาหนึ่ง กับข้อมูลนั้น ณ เวลาในอดีตช่วงใดช่วงหนึ่ง ตัวอย่างเช่น Y_t สัมพันธ์กับ Y_{t-1} หรือ Y_t สัมพันธ์กับ Y_{t-3} เป็นต้น ดังนั้นอัตตสัมพันธ์จึงสามารถบอกความสัมพันธ์ระหว่างคู่ของข้อมูลนั้น ๆในช่วงเวลาต่างกันได้หลาย ๆช่วง เรียกว่า Lags

2. การตรวจสอบความเหมาะสมสำหรับรูปแบบที่คัดเลือกมาแล้วและได้ทำการประมาณค่าสัมประสิทธิ์ของรูปแบบนั้น ๆเพื่อให้ได้สมการพยากรณ์มา แล้วจึงทำการตรวจสอบโดย คำนวณค่าความคลาดเคลื่อนของการพยากรณ์ แล้วนำข้อมูลความคลาดเคลื่อนมาตรวจสอบด้วยการหาอัตตสัมพันธ์ของความคลาดเคลื่อน ถ้าอัตตสัมพันธ์มีค่าน้อยและไม่มีการเปลี่ยนแปลงที่มีรูปแบบแน่นอน(No Pattern) รูปแบบของข้อมูลที่เลือกมาเป็นอันว่านำไปใช้สำหรับการพยากรณ์ได้ แต่ถ้าอัตตสัมพันธ์มีค่าสูงคือมากกว่าศูนย์อย่างมีนัยสำคัญและมีการเปลี่ยนแปลงที่มีรูปแบบแน่นอน แสดงว่ารูปแบบของข้อมูลที่เลือกมาไม่เหมาะสม ควรต้องกลับไปเลือกหารูปแบบของข้อมูลใหม่

วิธีการของบอกซ์และเจนกินส์

วิธีการของบอกซ์และเจนกินส์แตกต่างจากวิธีการพยากรณ์อื่น ๆที่ว่าจะใช้หลักการของการทำซ้ำหลาย ๆครั้งเพื่อให้ได้รูปแบบ (Model) ที่เหมาะสมที่สุดสำหรับข้อมูล โดยมีขั้นตอนเริ่มแรกจากการที่จะต้องค้นหารูปแบบที่เป็นไปได้สำหรับข้อมูล หลังจากที่ค้นหารูปแบบที่เชื่อว่าเหมาะสมแล้วจะเป็นการตรวจสอบว่าเหมาะสมจริงหรือไม่ ถ้าตรวจแล้วไม่เหมาะสมต้องย้อนกลับไปค้นหารูปแบบที่เหมาะสมใหม่ตั้งแต่ต้น กระบวนการดังกล่าวจะเกิดขึ้นหลายครั้งจนกว่าจะได้รูปแบบที่เหมาะสมที่สุด

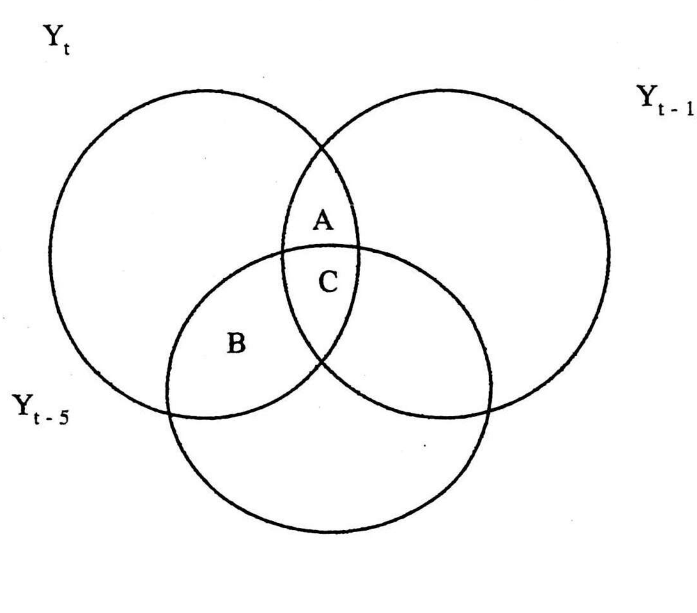
1. การค้นหารูปแบบที่เหมาะสมสำหรับข้อมูล (Model Identification) หลังจากที่พิจารณาแล้วว่าข้อมูลอนุกรมเวลามีการเคลื่อนไหวแบบสุ่ม แบบแนวโน้ม แบบฤดูกาล หรือแบบผสมทั้งแนวโน้มและฤดูกาล และได้มีการเปลี่ยนแปลงข้อมูลอนุกรมเวลาดังกล่าวให้คงที่แล้ว (ไม่มีแนวโน้ม) ขั้นตอนถัดไปคือการค้นหารูปแบบที่เป็นไปได้สำหรับข้อมูลที่มีการเปลี่ยนแปลงแล้ว (ถ้ามีแนวโน้ม) ดังนั้นจะมีข้อมูลที่เปลี่ยนแปลงแล้ว 2 ลักษณะคือ ข้อมูลที่ไม่มีฤดูกาล และข้อมูลที่มีฤดูกาล โดยจะแบ่งการพยากรณ์ของวิธีดังกล่าวออกเป็น 2 หัวข้อตามลักษณะดังกล่าว สำหรับข้อมูลที่มีฤดูกาลจะแยกอธิบายไว้หลังจากส่วนนี้แล้ว สำหรับการค้นหารูปแบบของข้อมูลซึ่งแบ่งออกเป็น 3 ลักษณะดังที่กล่าวมาแล้วข้างต้น คือรูปแบบอัตตถดถอย (Autoregressive Models : AR (p)) รูปแบบค่าเฉลี่ยเคลื่อนที่ (Moving Average : MA (q)) และรูปแบบอัตตถดถอยผสมค่าเฉลี่ยเคลื่อนที่ (Autoregressive

Integrated Moving Average : ARIMA (p,q)) สิ่งที่สำคัญของการค้นหารูปแบบ คือ ค้นหาว่ารูปแบบเป็นแบบใด และเมื่อค้นหาได้แล้วจะต้องค้นหาต่อไปว่ารูปแบบนั้นมีอันดับเท่าไร ตัวอย่างเช่น ถ้ารู้ว่ารูปแบบของข้อมูลเป็นแบบ AR(p) จะต้องค้นหาว่า p เป็นเท่าไรจึงจะเหมาะสมที่สุด เป็นต้น ดังได้กล่าวแล้วว่าเครื่องมือสำหรับการค้นหารูปแบบที่เหมาะสม คือ ฟังก์ชันอัตตสัมพันธ์ของข้อมูล สำหรับการค้นหาว่าอันดับของรูปแบบจะเป็นเท่าไรต้องอาศัยการวิเคราะห์จาก อัตตสัมพันธ์เชิงส่วน (Partial Autocorrelation)

อัตตสัมพันธ์เชิงส่วน (Partial Autocorrelation)

เป็นความสัมพันธ์คู่ระหว่างข้อมูลอนุกรมเวลา Y_t กับ Y_{t-T} ใด ๆ สำหรับ T = 1,2,3 ... โดยที่กำหนดให้ความสัมพันธ์ที่เกิดขึ้นระหว่าง Y_t กับ Y_{t-T} ใด ๆที่เหลือถูกขจัดออกไป จะพบว่าความสัมพันธ์ระหว่าง Y_t กับ Y_{t-1} และ Y_{t-5} แทนด้วยส่วน A+B+C สำหรับอัตตสัมพันธ์เชิงส่วนระหว่าง Y_t กับ Y_{t-1} แทนด้วยส่วน A เรียกว่า r_{11} สำหรับอัตตสัมพันธ์เชิงส่วนระหว่าง Y_t กับ Y_{t-5} แทนด้วยส่วน B เรียกว่า r_{55} ดังนั้นสูตรสำหรับการคำนวณหาค่าอัตตสัมพันธ์เชิงส่วน r_{TT} สำหรับ T = 1,2,3 ... เป็นดังนี้

รูปภาพที่ 4.1 ความสัมพันธ์ระหว่าง Y_t กับ Y_{t-1} และ Y_{t-5}



ที่มา : สุพล ดุรงค์วัฒนา , 2537 หน้า 117

$$
r_{TT} = \begin{cases} r_1 & \text{สำหรับ } T = 1 \\ \dfrac{r_T - \sum_{j=1}^{T-1} r_{T-1.j}\, r_{T-j}}{1 - \sum_{j=1}^{T-1} r_{T-1.j}\, r_j} & \text{สำหรับ } T = 1,2,\ldots,k \end{cases}
$$

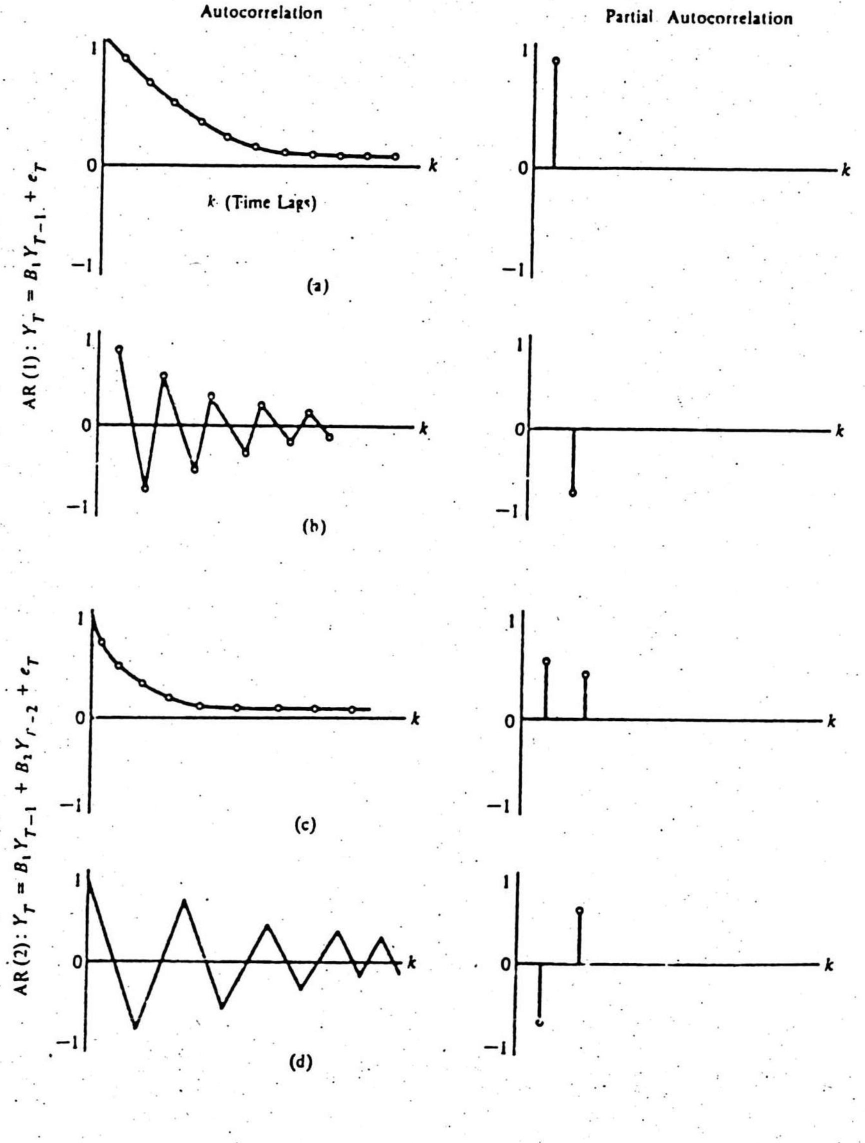
โดยที่มีความแปรปรวนประมาณ

$$
VAR\,(r_{TT}) = 1/n \qquad \text{สำหรับ } T = 1,2,3,\ldots,k
$$

การค้นหารูปแบบและอันดับที่เหมาะสม

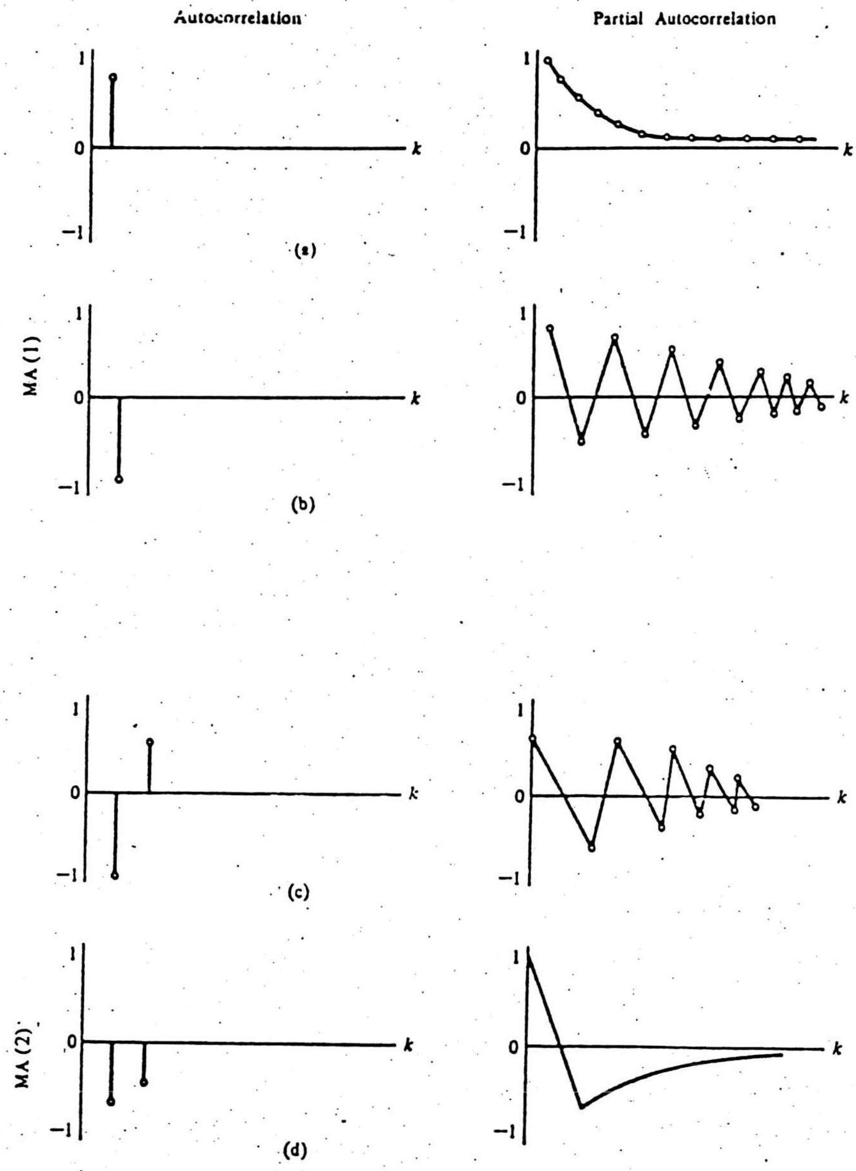
การค้นหารูปแบบและอันดับที่เหมาะสมสำหรับข้อมูล (ที่เปลี่ยนแปลงแล้วถ้าไม่คงที่) เป็นการเปรียบเทียบ SACF และ SPACF กับ ฟังก์ชันอัตตสัมพันธ์จริง (Autocorrelation Function : ACF) กับ ฟังก์ชันอัตตสัมพันธ์เชิงส่วนจริง (Partial Autocorrelation Function : PACF) ของรูปแบบและอันดับต่าง ๆตามลำดับ โดยพยายามเปรียบเทียบว่า SACF และ SPACF จะใกล้เคียง ACF และ PACF ใด ๆมากที่สุด ซึ่งการเปรียบเทียบนี้ต้องอาศัยทั้งความรู้และประสบการณ์ สำหรับภาพของ ACF และ PACF ของรูปแบบและอันดับต่างๆ แสดงไว้ดังรูปภาพที่ 4.2 รูปภาพที่ 4.3 และรูปภาพที่ 4.4 ตามลำดับ

รูปภาพที่ 4.2 ฟังก์ชันอัตตสัมพันธ์และอัตตสัมพันธ์เชิงส่วนจริงของรูปแบบAR(1)และAR(2)



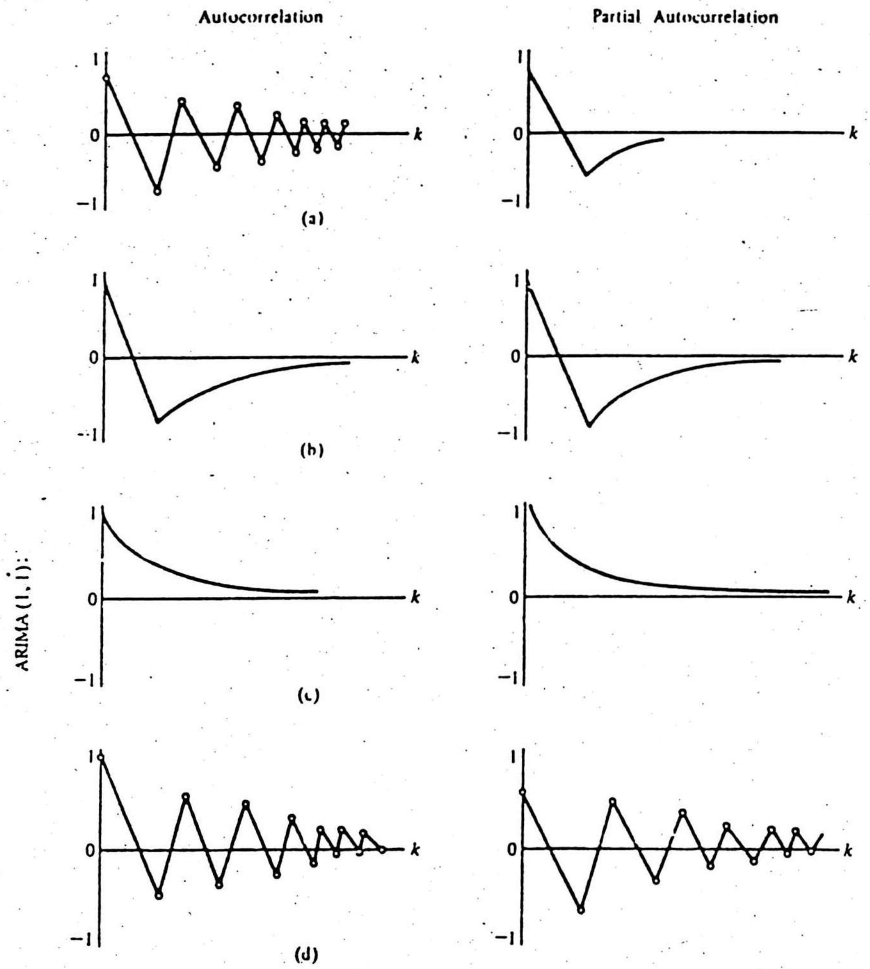
ที่มา : สุพล ดุรงค์วัฒนา , 2537 หน้า 119

รูปภาพที่4.3 ฟังก์ชันอัตตสัมพันธ์และอัตตสัมพันธ์เชิงส่วนจริงของรูปแบบMA(1)และMA(2)



ที่มา : สุพล ดุรงค์วัฒนา , 2537 หน้า 120

รูปภาพที่ 4.4 ฟังก์ชันอัตตสัมพันธ์และอัตตสัมพันธ์เชิงส่วนจริงของรูปแบบ ARMA(1,1)



ที่มา : สุพล ดุรงค์วัฒนา , 2537 หน้า 121

2. การประมาณค่าพารามิเตอร์ของรูปแบบที่ค้นหาได้ (Model Estimation)

หลังจากที่ค้นหารูปแบบและอันดับที่คิดว่าเหมาะสมสำหรับข้อมูลได้แล้ว ขั้นตอนต่อมาคือการประมาณค่าพารามิเตอร์ของรูปแบบที่ค้นหาได้ เพียงแต่กำหนดว่ารูปแบบและอันดับที่ค้นหาได้เป็นอย่างไรเท่านั้น สำหรับขั้นตอนการประมาณค่าพารามิเตอร์ของทั้ง 3 รูปแบบ นั้นต้องอาศัยหลักการของการทำซ้ำหลายครั้ง (Iterative Procedure) เพื่อให้ได้ค่าประมาณที่เหมาะสมที่สุด คือ ค่าที่ทำให้ความคลาดเคลื่อนของการพยากรณ์น้อยที่สุด (ค่า MSE) ซึ่งค่า MSE คำนวณได้จากสูตร

$$\mathrm{MSE} = \sum_{t=1}^{n}(y_t - \hat{y}_t)^2 / n$$

เมื่อค่า y_t เป็นข้อมูลอนุกรมเวลา ณ เวลา t

$\hat{y}_t$ เป็นค่าพยากรณ์ที่คำนวณได้จากรูปแบบที่ประมาณได้ ณ เวลา t

n เป็นจำนวนข้อมูลอนุกรมเวลา

3. การตรวจสอบความเหมาะสมของรูปแบบ (Diagnostic Checking)

ก่อนที่จะนำรูปแบบที่ประมาณได้ไปใช้พยากรณ์ ผู้วิเคราะห์ต้องตรวจสอบเสียก่อนว่ารูปแบบดังกล่าวเหมาะสม โดยการตรวจสอบจากความคลาดเคลื่อน

$$e_t = Y_t - \hat{Y}_t \qquad \text{สำหรับ } t = 1,2,\ldots,n$$

ทั้งนี้ก็ตรวจสอบว่าความคลาดเคลื่อนดังกล่าวเป็นความคลาดเคลื่อนสุ่ม (Random Error) โดยทำการวิเคราะห์อัตตสัมพันธ์ของความคลาดเคลื่อนดังกล่าวว่า อัตตสัมพันธ์ของความคลาดเคลื่อนนั้นมีนัยสำคัญหรือไม่ (เท่ากับ 0) ถ้ามีบางช่วง T ที่ T ต่ำ ๆหรือ T ที่ฤดูกาลใด ๆหรือทวีคูณของฤดูกาล ที่แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่ารูปแบบที่ค้นหามาอาจไม่เหมาะสม จำเป็นต้องกลับไปหารูปแบบที่เหมาะสมและการประมาณค่าพารามิเตอร์ซ้ำอีก สำหรับการตรวจสอบว่ารูปแบบนั้นเหมาะสมหรือไม่อาจตรวจสอบจากการ

ทดสอบสมมติฐานด้วยตัวสถิติไคสแควร์ ที่รู้จักในนามของตัวสถิติ Q ของบอกซ์และเพียซ (Box-Pierce Q Statistic) ซึ่งคำนวณได้ดังนี้

$$Q = (n\text{-}d) \sum_{T=1}^{k} r_T^2$$

ซึ่งมีการแจกแจงประมาณแบบไคสแควร์ ที่องศาความเป็นอิสระเท่ากับ k-p-q

เมื่อ

n เป็นจำนวนข้อมูลอนุกรมเวลา Y_t

d เป็นจำนวนครั้งของการหาผลต่างจนกระทั่งข้อมูลใหม่คงที่

k เป็นจำนวนสัมประสิทธิ์อัตตสัมพันธ์ของความคลาดเคลื่อนที่ใช้ตรวจสอบความเหมาะสม

p เป็นจำนวนอันดับของรูปแบบอัตตถดถอย

q เป็นจำนวนอันดับของรูปแบบค่าเฉลี่ยเคลื่อนที่

r_T เป็นสัมประสิทธิ์อัตตสัมพันธ์ของความคลาดเคลื่อนที่ T = 1,2,3 ...,k

ถ้าค่า Q ที่คำนวณได้จากรูปแบบใด ๆที่ค้นหาได้โตกว่าค่าไคสแควร์ ที่องศาความเป็นอิสระ k-p-q และที่ระดับนัยสำคัญ α แล้วรูปแบบที่ค้นหาได้ไม่น่าจะเหมาะสม ควรย้อนกลับไปหารูปแบบที่เหมาะสมใหม่ การตรวจสอบความเหมาะสมของรูปแบบทั้งสองวิธีนี้ควรใช้ร่วมกัน ทั้งนี้เพราะอัตตสัมพันธ์ของความคลาดเคลื่อนของช่วงที่มีนัยสำคัญอาจเกิดจากเหตุการณ์ที่ผิดปกติ ซึ่งผู้วิเคราะห์ทราบสถานการณ์ดังกล่าว ถ้าเกิดขึ้นในช่วงเวลานั้นเท่านั้น อาจสามารถมองข้ามส่วนดังกล่าวไปได้ อีกประการหนึ่งอาจเป็นไปได้ว่ารูปแบบที่เหมาะสมที่ค้นหามาอาจมีมากกว่า 1 รูปแบบ ข้อแนะนำคือเลือกรูปแบบที่ง่ายที่สุด มีการประมาณค่าพารามิเตอร์ของรูปแบบน้อยที่สุด

4. การนำรูปแบบที่เหมาะสมไปพยากรณ์ (Forecasting with the Model)

หลังจากที่ตรวจสอบหารูปแบบที่เหมาะสมได้แล้ว ขั้นตอนถัดไปคือการนำรูปแบบ (สมการ) ดังกล่าวไปพยากรณ์ค่าที่ต้องการได้ในช่วงเวลาต่าง ๆพร้อมกับสามารถหาช่วงความเชื่อมั่น (Confidence Intervals) สำหรับค่าพยากรณ์เหล่านี้ โดยปกติแล้วช่วงความเชื่อมั่นจะกว้างมากขึ้น ถ้าระยะเวลาของการพยากรณ์ในอนาคตห่างไกลจากปัจจุบันมาก โดยปกติโปรแกรมสำเร็จรูปสำหรับการพยากรณ์ด้วยวิธีบอกซ์และเจนกินส์นั้นจะคำนวณช่วงความเชื่อมั่นไว้ให้เมื่อต้องการสำหรับสูตรการคำนวณหาช่วงความเชื่อมั่นมีดังนี้

$$\sigma_0 = 0$$
$$\sigma_1 = \phi_1 - \theta_1$$
$$\sigma_2 = \phi_1\sigma_1 + \phi_2 - \theta_2$$
$$\sigma_3 = \phi_1\sigma_2 + \phi_2\sigma_1 - \theta_3$$
$$\cdot \quad \cdot \quad \cdot \quad \cdot$$
$$\cdot \quad \cdot \quad \cdot \quad \cdot$$
$$\sigma_j = \phi_1\sigma_{j-1} + \phi_2\sigma_{j-2} + \dots + \phi_p\sigma_{j-p} - \theta_q$$

$$VAR\,(Y_{t+1}) = \left\{1 + \sum_{j=1}^{l-1}\delta_j^2\right\} S_e^2$$ เมื่อ l เป็นระยะช่วงห่างจาก t

และ $S_e^2 = \sum_{t=1}^{n}(y_t - y_t)^2 / (n-1)$ เมื่อ n เป็นจำนวนข้อมูล

และ $(1-\alpha)$ 100 % ช่วงความเชื่อมั่นของค่าพยากรณ์ช่วงที่ 1 คือ

$$Y_{t+1} \pm Z_{\alpha/2}\left\{\sqrt{VAR(y_{t+1})}\right\}$$

ถ้าการคำนวณค่าพยากรณ์ด้วยรูปแบบที่ได้ให้ค่าความคลาดเคลื่อนน้อย แสดงว่ารูปแบบที่ค้นหาได้เหมาะสมแล้วแต่การประมาณค่าพารามิเตอร์อาจต้องมีการประมาณใหม่ทั้งนี้เราทราบแล้วว่าการประมาณเป็นการทำซ้ำจนกว่าได้ค่าประมาณที่เหมาะสม ดังนั้นค่าประมาณ

ที่ได้เมื่อหยุดทำซ้ำอาจไม่ใช่ค่าประมาณที่เหมาะสมที่สุด ดังนั้นอาจต้องย้อนกลับไปประมาณค่าพารามิเตอร์สำหรับรูปแบบดังกล่าวใหม่ แต่ถ้าค่าความคลาดเคลื่อนจากการพยากรณ์สูงมาก แสดงว่ารูปแบบดังกล่าวอาจไม่เหมาะสมต้องกลับไปตั้งต้นใหม่

4.2 การวิเคราะห์อนุกรมเวลาโดยวิธี Decomposition

การวิเคราะห์อนุกรมเวลาโดยวิธี Decomposition นี้ วัตถุประสงค์เบื้องต้นในการวิเคราะห์อนุกรมเวลาเพื่อจะค้นหา วัดความสม่ำเสมอ และแยกลักษณะความเคลื่อนไหวหรือความผันผวนของข้อมูลที่เกิดขึ้นบ่อยๆ ซึ่งอาจมีอยู่ในข้อมูล ลักษณะความเคลื่อนไหวที่ปรากฏจริงๆในระบบเศรษฐกิจ มีอยู่ 4 ลักษณะคือ

1. ลักษณะแนวโน้ม (Secular trend หรือ Long term trend : T) ความโน้มเอียงพื้นฐานของธุรกิจในการเติบโตหรือการลดลงในช่วงปีปกติจะถือว่าเป็น "ความเคลื่อนไหวตลอดไป" (secular movement) หรือ "แนวโน้ม" (trend) ในอนุกรมทิศทางของแนวโน้มจะเห็นได้ชัดจากการวิเคราะห์ดูแผนภาพ ซึ่งจะเห็นว่า อนุกรมกำลังเพิ่มขึ้นหรือลดลงแม้ว่าอัตราเพิ่มขึ้นหรือลดลงจะวัดเป็นตัวเลขแน่นอนไม่ได้ แนวโน้มจะถูกคำนวณออกมาแล้วลากผ่านข้อมูลเดิม ลักษณะเด่นชัดของเส้นก็คือ ความเรียบ (smoothness) ของเส้น เส้นแนวโน้มจะไม่ผิดปกติเหมือนกับเส้นที่แสดงลักษณะของข้อมูลเดิมที่เส้นแนวโน้มผ่าน แนวความคิดเรื่องแนวโน้มต้องเป็นเส้นเรียบเพราะการเคลื่อนไหว หรือเปลี่ยนแปลงเกิดจากแรงซึ่งดำเนินไปเรื่อย ๆตลอดระยะเวลา

2. ลักษณะความเคลื่อนไหวแบบฤดูกาล (Seasonal variation : S) การเคลื่อนไหวแบบฤดูกาลมีสาเหตุ 2 ประการคือ ประการแรก ดินฟ้าอากาศ การเกษตรเป็นตัวอย่างที่เห็นได้ชัด เพราะว่าการเกษตรมีฤดูกาลเพาะปลูกจำกัด ซึ่งทำให้ผลผลิตที่จะเก็บเกี่ยวก็มีระยะเวลาจำกัดและแน่นอน เป็นผลให้รายได้ของกสิกรไม่สม่ำเสมอเท่ากันตลอดปี การใช้จ่ายของกสิกรก็มักจะรวมอยู่ในเดือนที่เขามีการซื้อขายผลิตผลของเขา พ่อค้าท้องถิ่น พ่อค้าในเมืองก็จะกระทบกระเทือนจากผลของฤดูกาลของรายได้ของกสิกร การคมนาคมขนส่งก็อาจจะมากในช่วงที่มีผลผลิตออกสู่ตลาด ธนาคารอาจจะต้องจัดหาเงินไว้มากขึ้นตามฤดูกาล

ระหว่างที่มีการเพาะปลูกและเก็บเกี่ยวสำหรับให้กสิกรกู้ยืม สาเหตุประการที่ 2 คือ ขนบธรรมเนียมประเพณีและนิสัยความเคยชิน ทำให้เกิดฤดูกาลขึ้น เช่น เทศกาลวันปีใหม่ เทศกาลสงกรานต์ วันเข้าพรรษา ออกพรรษา วันเด็กหรือวันแม่ เป็นต้น จะทำให้ธุรกิจการขายสินค้าที่เกี่ยวกับวันเหล่านี้มีการหมุนเวียนมากกว่าในเวลาปกติ

3. ลักษณะการเคลื่อนไหวแบบวงจรหรือวัฏจักร (Cycle variation : C) การเคลื่อนไหวขึ้นลงของข้อมูลที่มีระยะนานกว่า 12 เดือนเรียกว่า การเคลื่อนไหวแบบวงจรหรือวัฏจักร (Cycle variation) เพราะเป็นการเคลื่อนไหวแกว่งไปมา ซึ่งโดยทั่วไปจะเป็นแบบลูกคลื่น แม้ว่าระยะทางจากยอดถึงท้องคลื่นของลูกคลื่นจะไม่เรียบ วัฏจักรเหล่านี้ปกติจะประกอบด้วย 4 ระยะคือ ระยะรุ่งเรือง (prosperity) ระยะถอยหลัง (recession) ระยะตกต่ำ (depression) และระยะฟื้นตัว (ecovery)

4. ลักษณะการเคลื่อนไหวผิดปกติคาดไม่ถึง (Irregular variation : I) ส่วนประกอบลักษณะที่ 4 ของอนุกรมเวลาก็คือ ความเคลื่อนไหวซึ่งเหลืออยู่และไม่ได้จัดอยู่ใน 3 ลักษณะแรก การเคลื่อนไหวนี้ต่างจากการเคลื่อนไหวแบบอื่นคือ ไม่มีแบบหรือแนวว่าจะเกิดขึ้นซ้ำอีกในระยะที่แน่นอน เป็นการเปลี่ยนแปลงที่ไม่มีแนวบอกไว้ล่วงหน้า เช่น แผ่นดินไหว น้ำท่วม พายุ การผันแปรในทางการเมือง การเกิดสงคราม ในทางอุตสาหกรรม เช่น การนัดหยุดงาน ความนิยมอย่างมากชั่วขณะหนึ่ง เหตุการณ์ที่ไม่คาดคิดมาก่อนถือว่าเป็นเหตุการณ์สุดวิสัยทำนายไม่ได้ ส่วนมากมักเป็นระยะสั้น เหตุการณ์เช่นนี้ถ้าพยายามจัดช่วงให้สั้น จะเป็นวิธีการศึกษาได้ผลมาก เป็นการขจัดความผันผวนได้อีกด้วย

ฉะนั้น สรุปได้ว่า การเคลื่อนไหวซึ่งปรากฏอยู่ในอนุกรมเวลาสามารถที่จะอธิบายได้ในลักษณะของแนวโน้ม ฤดูกาล วัฏจักร และความผิดปกติ ลักษณะการเคลื่อนไหว 4 อย่างนี้รวมกันเป็นลักษณะที่สามารถเห็นได้ในอนุกรมของข้อมูลเศรษฐกิจ

แนวโน้ม (Trend)

การวิเคราะห์อนุกรมเวลามุ่งวิเคราะห์ความเคลื่อนไหวประเภทที่ 1 คือ วิเคราะห์แนวโน้ม การวิเคราะห์ก็เพื่อที่จะแก้ไขความไม่แน่นอน หาความสม่ำเสมอของความเคลื่อนไหวของข้อมูลที่เกิดขึ้นเป็นระยะยาวว่าเป็นไปในรูปใด เราทราบแล้วว่าการหาเส้นแนวโน้มต้องมีระยะนานพอสมควร คือ 10 ปีขึ้นไป ฉะนั้นในช่วงเวลา 10 ปีนี้ ก็จะมีความผันผวนประเภทที่ 2 , 3 และ 4 เกิดขึ้นพร้อม ๆ กับแนวโน้มด้วย ดังนั้น การวิเคราะห์ก็อาจจะแก้ไขความไม่แน่นอนประเภทที่ 2 , 3 , 4 ไปด้วย ในการศึกษานี้จะใช้วิธีการประมาณค่าแนวโน้มโดยวิธีลัดโดยอาศัยวิธีกำลังสองน้อยที่สุด (Least Square Method) วิธีนี้จะให้ trend เป็นเส้นตรงที่มีความคลาดเคลื่อนจากกราฟเดิมน้อยที่สุด การที่ให้ชื่อว่ากำลังสองน้อยที่สุดเพราะวิธีการคำนวณจะให้คุณสมบัติทางคณิตศาสตร์สำคัญที่แน่นอน เส้นแนวโน้มที่คำนวณโดยวิธี Least Square จะให้ผลบวกกำลังสองของความคลาดเคลื่อนน้อยที่สุด ซึ่งเขียนได้เป็น $\Sigma\,(Y - Y)^2$ = น้อยที่สุด ด้วยเหตุผลนี้ บางครั้งจึงเรียกเส้นกำลังสองน้อยที่สุดว่า "เส้นที่เหมาะที่สุด" (line of best fit)

สัญญลักษณ์ของเส้นตรงโดยกำลังสองน้อยที่สุดอาจจะเขียนในรูปสมการดังนี้

Y	=	a + bX
Y	=	แนวโน้ม หรือ ค่าคำนวณของแนวโน้ม
a	=	ค่าของเส้นแนวโน้ม ณ จุดเริ่มต้น
b	=	การเติบโต หรือการลดลงในเวลา X หน่วย
X	=	เวลา

แนวโน้มเส้นตรงที่ได้นี้สามารถจะอธิบายการเคลื่อนไหวระยะยาวได้อย่างมีเหตุมีผล และการที่จะได้แนวโน้มตามสมการข้างบน ก็เพียงแต่คำนวณหาค่า a และ b ออกมา โดยใช้หลักว่าด้วยความคลาดเคลื่อนกำลังสองน้อยที่สุด คือ

$$\Sigma\,(Y - \hat{Y})^2 \quad = \quad 0$$ (0 นับว่าน้อยที่สุด)

แต่ $Y = a + bX$ แทนค่า Y ในสมการ $(Y - Y)^2 = 0$ อาศัยหลักอนุพันธ์ (derivative) มุ่งต่อ a ครั้งหนึ่งและต่อ b ครั้งหนึ่ง จะได้ Normal Equation คือ

$$\Sigma Y = na + b\Sigma X \qquad (1)$$

$$\Sigma XY = a\Sigma X + b\Sigma X^2 \qquad (2)$$

Y = ค่าของข้อมูลในอนุกรมเวลา

n = จำนวนของค่า

X = หน่วยของเวลา ซึ่งเรากำหนดเป็นตัวเลขขึ้นมา

วิธีการคำนวณหาสมการแนวโน้มจากข้างต้น ในกรณีที่ข้อมูลเป็นเลขจำนวนมากๆ ค่า X^2 , XY ก็จะยิ่งมาก มีวิธีลัดซึ่งง่ายกว่าและไม่เสียเวลาสามารถทำได้โดยการเลือกปีที่อยู่กลางอนุกรมเป็นจุดเริ่มต้นมากกว่าที่จะเลือกปีแรกของอนุกรม โดยใช้หลักให้ผลรวมของ X เท่ากับ 0 เมื่อนำ X = 0 ไปแทนค่าใน Normal Equation จะทำให้ Normal Equation ลดลงมาเหลือดังนี้

$$\Sigma Y = na$$

$$\Sigma XY = b\Sigma X^2$$

ค่า a จะเขียนได้เป็น

$$a = \frac{\Sigma Y}{n}$$

ซึ่งเป็นสูตรมัชฌิมเลขคณิตของ Y ดังนั้นโดยวิธีลัด a ก็คือค่าเฉลี่ยเลขคณิตของอนุกรม Y

ค่า b จำนวนเปลี่ยนแปลงเฉลี่ยในแนวโน้มต่อเวลา X หน่วยซึ่งจะเขียนได้เป็น

$$b = \frac{\Sigma XY}{\Sigma X^2}$$

การเปลี่ยนแปลงตามฤดูกาล(Seasonal Variation)

ในงานวิจัยนี้จะใช้วิธีการหาอัตราส่วนเทียบกับค่าเฉลี่ยเคลื่อนที่ (Ratio-to-moving-average Method) ในการหาค่าดัชนีของฤดูกาล การหาดัชนีโดยวิธี Ratio-to-moving-average เป็นวิธีที่ได้ผลดีที่สุดในการหาดัชนีแสดงถึงการผันผวนของฤดูกาลหลาย ๆวิธี เพราะเป็นวิธีที่สามารถจะแยกอิทธิพลของแนวโน้มวัฏจักร และความผิดปกติออกได้มากกว่าวิธีอื่น จะเหลือค่าความผันแปรอันเนื่องจากฤดูกาลซึ่งแสดงออกในรูปของดัชนี

หลักการหา Ratio-to-moving-average :

ขั้นแรก คำนวณแนวโน้ม (T) และวัฏจักร (C) โดยวิธีเฉลี่ยเคลื่อนที่ข้อมูล (Moving Average) โดยวิธีนี้จะเป็นการขจัดฤดูกาล (S) และความผิดปกติ (I) ของข้อมูลออกไป

การเฉลี่ยเคลื่อนที่ในงานวิจัยนี้จะใช้การเฉลี่ยเคลื่อนที่ 12 เดือน เนื่องจากรูปแบบของฤดูกาลเกิดขึ้นซ้ำกันทุก 12 เดือน เพราะจะเป็นการแยกฤดูกาลที่มีการเคลื่อนไหวแบบเดียวกัน ทั้งขนาดและระยะเวลาออกจากข้อมูลได้ดีกว่า ผลที่ได้จากค่าเฉลี่ยเคลื่อนที่จะเป็นค่า T และ C

ขั้นที่สอง คำนวณหาอัตราส่วนโดยหารข้อมูลเดิมด้วยค่าเฉลี่ยเคลื่อนที่ที่หาได้ในขั้นแรก ในขั้นนี้จึงได้ชื่อว่า "อัตราส่วนเทียบกับค่าเฉลี่ยเคลื่อนที่" (Ratio-to-moving-average Method) ผลที่ได้จะเป็นความผันแปรจของฤดูกาล (S) และมีความผิดปกติ (I) รวมอยู่ ซึ่งเป็นความผันแปรฤดูกาลเฉพาะ (Specific Seasonal) ของแต่ละปี

$$\frac{TXCXSXI}{TXC} = S X I$$

จะเห็นได้ว่า โดยการเฉลี่ยเคลื่อนที่ (Moving Average) ในขั้นที่ 1 จะให้ค่า T และ C โดยไม่มีอิทธิพลของ S และ I ส่วนอัตราส่วนเทียบกับค่าเฉลี่ยเคลื่อนที่ (Ratio-to-moving-average) ในขั้นที่ 2 จะให้ค่า S และ I โดยไม่มีอิทธิพลของ T และ C

ขั้นที่สาม แยก S X I ในขั้นที่ 2 ออกจากกัน จะได้ดัชนีฤดูกาลโดยประมาณ (Crude Seasonal Index) การแยก I ออกจาก S ทำได้โดยวิธีเฉลี่ยค่า S X I การเฉลี่ย (Average) จะทำให้ความผิดปกติ (I) ต่าง ๆหายไป ตัวอย่างเช่นในอนุกรมเวลาที่มีข้อมูลบางปีใหญ่มาก แต่เมื่อเฉลี่ยแล้วจะมองไม่เห็นข้อมูลที่ใหญ่ผิดปกตินี้ ซึ่งก็จะเป็นการขจัดค่าที่ใหญ่มากผิดปกติ (I) ออกไป การเฉลี่ยนอกจากจะเป็นการขจัดความผิดปกติ (I) แล้ว ยังเป็นการขจัดความผิดพลาดที่อาจจะมีในการคำนวณในขั้นที่ 2

การเฉลี่ยเพื่อให้ค่าผิดปกติหายไปนี้ นอกจากจะใช้วิธีเฉลี่ยจากข้อมูลทั้งหมดแล้ว เราอาจจะตัดค่าที่ผิดปกติที่เห็นชัดอันได้แก่ ค่าที่ใหญ่มากและค่าที่เล็กมากออกไป แล้วเฉลี่ยเฉพาะค่าที่เหลือ ซึ่งเรียกว่า Modified หรือ Positional Means วิธีนี้เป็นวิธีที่นิยมมากกว่าเฉลี่ยจากข้อมูลทั้งหมด ค่าของ Modified หรือ Positional Means ที่ได้จะเป็น ดัชนีโดยประมาณ (Crude Seasonal Index)

ขั้นที่สี่ ปรับค่า (leveling) ดัชนีโดยประมาณที่หาได้จริงให้เป็นดัชนีที่ถูกต้อง (Typical Seasonal Index)

4.3 การสร้างข้อมูลผลผลิตจริงรายเดือน

ในการศึกษาความสัมพันธ์ของตัวแปรต่างๆทางเศรษฐศาสตร์มหภาค ข้อมูลของตัวแปรทางเศรษฐศาสตร์ต่างๆจำเป็นต้องมีช่วงระยะเวลายาวนานพอสมควร เพื่อที่จะทำให้นักเศรษฐศาสตร์สามารถทดสอบสมมติฐานต่างๆที่เกี่ยวกับความสัมพันธ์ของตัวแปรเหล่านั้นได้ บางครั้งข้อมูลทางเศรษฐศาสตร์ที่มีอยู่มีลักษณะเป็นข้อมูลรายปี ในขณะที่นักเศรษฐศาสตร์ต้องการข้อมูลในลักษณะที่เป็นรายเดือน กรณีของประเทศไทยข้อมูลของตัวแปรทางเศรษฐศาสตร์บางตัวมีลักษณะเป็นข้อมูลรายปี แต่ไม่มีข้อมูลที่มีลักษณะเป็นรายเดือน เช่น ข้อมูลผลผลิตจริง (GDP) ดังนั้นเราจึงจำเป็นต้องสร้างข้อมูลของผลผลิตจริงให้เป็นรายเดือนก่อน เพื่อที่จะสามารถทดสอบความสัมพันธ์ระหว่างผลผลิตจริงกับปริมาณการขนส่งสินค้าส่งออกทางอากาศ งานวิจัยนี้ใช้วิธี Chow-Lin ในการสร้างข้อมูลผลผลิตจริงรายเดือน โดยทั่วไปการสร้างข้อมูลทางเศรษฐศาสตร์มหภาครายเดือนโดยวิธีทางสถิติ(Statistical Methods)

นั้น ใช้สมการถดถอยแบบเส้นตรง(Linear regression) กับตัวแปรที่เกี่ยวข้อง ซึ่งมีทั้งข้อมูล
รายปีและข้อมูลรายเดือนที่เกี่ยวข้อง ดังรายละเอียดต่อไปนี้

สมมติให้ p คือ จำนวนตัวแปรที่เกี่ยวข้องทั้งที่จะใช้ในการคำนวณหาค่าประมาณ
การของตัวแปรที่สนใจ ดังนั้นถ้าจำนวนปีทั้งหมดมีอยู่ n ปี Matrix ขนาด n x p จึงแสดง
ถึง ข้อมูลรายปีทั้งหมดของตัวแปรที่เกี่ยวข้อง และ Matrix X ขนาด 12n x p จะแสดงถึง
ข้อมูลรายเดือนทั้งหมดของตัวแปรที่เกี่ยวข้อง แต่สำหรับตัวแปรที่สนใจนั้นให้ Y เป็น
vector ขนาด n x 1 ซึ่งแสดงถึงข้อมูลรายปี n ปีของตัวแปรที่สนใจ ดังนั้นสิ่งที่ต้องการ
คำนวณหาคือข้อมูลรายเดือนจำนวน n ปี ของตัวแปรที่สนใจซึ่งอยู่ในรูปของ Matrix y
ขนาด 12n x 1 นอกจากนี้ยังได้สมมติต่อไปอีกว่า ความสัมพันธ์ระหว่างค่าประมาณการ
ของข้อมูลรายเดือนของตัวแปรที่สนใจกับค่าจริงของข้อมูลรายเดือนของตัวแปรที่เกี่ยวข้อง
เป็นเส้นตรง คือ

$$y = x\beta + U$$

โดย U คือค่า random vector ขนาด 12n x 1 ซึ่งมีค่า mean = 0 และ covariance
matrix $v=E(uu')$

ให้ A เป็น Matrix ขนาด n x 12n ซึ่งจะเป็น Matrix ที่ใช้ในการแปลงข้อมูลราย
เดือนของตัวแปรตัวใดตัวหนึ่งเป็นข้อมูลรายปีของตัวแปรตัวเดียวกัน ผลที่ได้ออกมาคือ

$$Y = Ay = Ax\beta + Au = X\beta + U$$

โดยที่ U ซึ่งเท่ากับ Au นั้น เป็นค่า random vector ขนาด n x 1 และมีค่า mean = 0
และ covariance matrix

$$V = E(UU') = E(Auu'A')$$

จะเห็นได้ว่าการที่สามารถใช้ความสัมพันธ์เชิงเส้นตรงระหว่างตัวแปรที่สนใจกับตัวแปรที่เกี่ยวข้องในการประเมินหาค่าประมาณการของข้อมูลรายเดือนของตัวแปรที่สนใจนั้น ขึ้นอยู่กับว่าค่า β ทั้งในอนุกรมรายปี ($Y = X\beta + U$) และอนุกรมรายเดือน ($y = x\beta + u$) นั้น จะต้องมีค่าเท่ากัน ดังนั้นจึงต้องมีการประมาณหาค่า β โดยใช้วิธี Ordinary least square:OLS กับข้อมูลรายปีของค่าตัวแปร Y และ X และใช้ค่า β ที่ได้ร่วมกับข้อมูลรายเดือนของค่าตัวแปรที่เกี่ยวข้องในการคำนวณหาค่าประมาณการของข้อมูลรายเดือนของตัวแปรที่สนใจต่อไปคือ

$$\overline{y} = x\beta$$

ถ้าความสัมพันธ์ระหว่าง Y และ X ไม่สมบูรณ์ จะทำให้

$$\sum_{i=12k-3}^{12k} y \neq Y_k$$

โดยที่ $k = 1,\ldots,n$

และ n คือ จำนวนปีในอนุกรม

แสดงว่าการใช้ OLS กับตัวแปรที่เกี่ยวข้องกันนั้นไม่ได้รับประกันว่า ผลบวกของข้อมูลรายเดือนในปีใดปีหนึ่งจะมีค่าเท่ากับข้อมูลรายปีของปีนั้นๆ

จาก $Y_k = b_0 + b_1X_k + U_k$ โดยที่ $k = 1,\ldots,n$

ซึ่ง U_k จะมีค่าเท่ากับ $Y_k - \overline{Y}_k$ โดยที่ $\overline{Y}_k$ ถูกกำหนดจากสมการถดถอย

$$\overline{Y}_k = b_0 + b_1X_k \qquad \text{โดยที่ } k = 1,\ldots,n$$

ดังนั้นการประมาณค่าของข้อมูลรายเดือนก็สามารถกระทำได้ง่ายๆโดยใช้ค่า constant และ ค่าสัมประสิทธิ์ที่ได้จากสมการข้างบนกับค่าจริงของข้อมูลรายเดือนของตัวแปรที่เกี่ยวข้อง ดังนี้

$\overline{Y}_k = b_0 + b_1x_1$ โดยที่ i=1,......,12n

จากปัญหาข้างต้นที่ว่า การหาค่าประมาณการของข้อมูลรายเดือนของตัวแปรที่สนใจ โดยใช้ความสัมพันธ์ในสมการถดถอยเชิงเส้นตรง(linear regression) นั้น ไม่ได้ประกันว่าค่าผลรวมของข้อมูลรายเดือนนั้นจะมีค่าเท่ากับค่าจริงของข้อมูลรายปี Chow และ Lin (1971) ได้กล่าวถึงปัญหานี้และได้หาวิธีแก้ไขประกอบกับการใช้ทฤษฎี best linear unbiased estimation (BLUE) ในสมการถดถอยเชิงเส้นตรง

วิธีของ Chow-Lin นั้น อาจแบ่งออกได้เป็น 2 ขั้นตอน โดยเริ่มจากการประมาณค่า y จาก xβ ซึ่งได้จากสมการถดถอยเชิงเส้นตรงและเปลี่ยนแปลงแก้ไขค่าประมาณการของ y โดยคิดเป็นสัดส่วนของผลต่างระหว่างค่าจริงของข้อมูลรายปีและค่าประมาณการของข้อมูลรายปีของตัวแปรที่สนใจ (Y - $\overline{Y}$) อันทำให้ผลรวมของค่าประมาณการของข้อมูลรายเดือนเท่ากับค่าจริงของข้อมูลรายปีของตัวแปรเดียวกัน คือ

$$\sum_{i=12k-3}^{12k} y_I = Y_k \qquad k = 1,........,n$$

ดังนั้นหลังจากหาค่าประมาณการของสัมประสิทธิ์ (β) ของตัวแปรที่เกี่ยวข้องโดย

$$Y = X\beta + U$$
$$\overline{Y} = X\beta$$

Chow-Lin ได้ค่า BLUE ของ Y จาก

$$y = X\,\beta + E[uU']\left[E(UU')\right]^{-1}[Y - Y]$$
$$u = y - x\beta \quad ; \quad U = Y - X\beta$$

จะเห็นได้ว่า ถ้าค่า (Y - Y) = 0 การประมาณค่า y ก็จะได้มาจากผลคูณของค่าสัมประสิทธิ์จากสมการถดถอยกับค่าจริงข้อมูลรายเดือนของตัวแปรที่เกี่ยวข้องเท่านั้น จึงทำ

ให้การปรับค่าไม่จำเป็นและหมายความว่า ค่าประมาณการข้อมูลรายเดือนนั้น จะถูกคำนวณโดยใช้วิธี OLS เท่านั้น แต่ถ้าค่า (Y -Y) $\neq 0$ Matrix ที่ใช้ในการปรับค่า $E(uU')[E(UU)]$ จะถูกผนวกเข้ากับค่า (Y - Y) อันทำให้ค่า y ไม่เท่ากับ x β เสียเลยทีเดียว Matrix ที่ใช้ในการปรับค่านี้มีขนาด 12nxn ซึ่งประกอบด้วยค่าสัมประสิทธิ์ของสมการถดถอยของ u ที่มีต่อ U ซึ่งอาจจะเขียนได้ในอีกลักษณะคือ

$$E(uU')[E(UU')]^{-1} = E(uU'A')V^{-1} = vA'(AvA')^{-1}$$

เนื่องจากว่าค่า u นั้นมาจาก $u = y - x\beta$ แต่ค่า y เป็นค่าที่ไม่ทราบ ดังนั้นค่า Matrix ที่ใช้ในการปรับค่าโดยวิธีของ Chow-Lin นั้น จึงไม่สามารถหาค่าได้ จึงต้องมีการประมาณค่าของ Matrix ที่ใช้ในการปรับค่าขึ้นมา ซึ่ง Chow-Lin ได้สมมติค่า u ดังนี้

ถ้าสมมติว่าค่า u ของ monthly regression ไม่มี serial correlation กับค่า variance,σ^2 ซึ่งคงที่ Matrix ที่ใช้ในการปรับค่าของ Chow-Lin จะอยู่ในลักษณะ

$$v = I_{12nx12n}\,\sigma^2$$

และ $$V = 12I_{nx12}\,\sigma^2$$

ผลที่ได้ออกมา $$vA'v^{-1} = 1/12A'$$

นั่นคือ $$y_i = x_i\beta + 1/12[Y_k - X_k\ \beta \quad ; i=,........,12n$$

แสดงว่าค่าตัวปรับของค่าประมาณของข้อมูลรายเดือนในแต่ละรายเดือนจะมีเป็น 1/12 เท่าของค่าผลต่างระหว่างข้อมูลรายปีที่แท้จริงและข้อมูลรายปีจากสมการถดถอยปีเดียวกัน ซึ่งสามารถใช้ได้ในทุกกรณีไม่ว่า x และ X จะเป็น Univariate หรือ Multivariate

งานวิจัยนี้ได้ใช้แนวความคิดของ Chow-Lin ในการสร้างข้อมูลผลผลิตจริงรายเดือน ณ ราคาปีฐาน พ.ศ. 2531 (1988) โดยมีรายละเอียดดังต่อไปนี้

ขั้นตอนที่ 1 หาตัวแปรที่เป็นตัวกำหนดค่าของผลผลิตจริงรายปี (real gross domestic product: RGDP) และใช้ข้อมูลรายปีของข้อมูลที่เป็นตัวกำหนดผลผลิตจริงกับข้อมูลของผลผลิตจริงรายปีมาทำการประมาณค่าสมการถดถอยโดยวิธี Ordinary Least Square (OLS) ตัวแปรที่ใช้ในการกำหนดตัวแปรผลผลิตจริงในที่นี้ คือ ค่าใช้จ่ายจริงของรัฐบาล (real goverment spending:RGOV) และมูลค่าการส่งออกจริง (real export of goods and services:REXP) โดยตัวแปรทางเศรษฐศาสตร์เหล่านี้ใช้ราคาปี พ.ศ. 2531(1988) เป็นปีฐาน

ขั้นตอนที่ 2 สร้างข้อมูลผลผลิตจริงรายเดือนได้ โดยการประมาณค่าสมการถดถอยของตัวแปรผลผลิตจริงรายเดือน (MRGDP) กับตัวแปรค่าใช้จ่ายจริงของรัฐบาลรายเดือน (MRGOV) และตัวแปรมูลค่าการส่งออกสินค้าและบริการรายเดือน (MREXP)

ขั้นตอนที่ 3 หาผลรวมของค่าผลผลิตจริงรายปี (RGDP) โดยการบวกค่าผลผลิตจริงรายเดือน (MRGDP) และนำเอาค่าผลรวมของผลผลิตจริงรายปีที่ประมาณค่าได้ไปลบออกจากค่าผลผลิตที่แท้จริง (Actual RGDP) เพื่อที่จะได้ค่าคลาดเคลื่อนของผลผลิตจริงรายปี (Annual residual:RES)

$$\mathrm{RES} = \mathrm{RGDP}_t - \sum_{i=1}^{12} MRGDP_{it}$$

ขั้นตอนที่ 4 คำนวณหาค่าเฉลี่ยถ่วงน้ำหนักรายเดือน (weighted monthly residuals:WRES) จาก

$$\mathrm{WRES}_{it} = (\mathrm{MRGDP}_{it} \,/\, \sum_{i=1}^{12} MRGDP_{it})\mathrm{x\ RES}$$

ในขั้นตอนสุดท้าย เราสามารถคำนวณหาค่าประมาณการของผลผลิตจริงรายเดือน (Y) ได้จาก

$$\mathrm{Y}_{it} = \mathrm{MRGDP}_{it} + \mathrm{WRES}_{it}$$