บทที่ 2

ทฤษฎีสถิติที่เกี่ยวข้องกับงานวิจัย

การวิจัยเรื่องการเปรียบเทียบวิธีการพยากรณ์ความต้องการใช้ไฟฟ้าในส่วนภูมิภาคนี้ เป็นการศึกษาเพื่อเปรียบเทียบวิธีการพยากรณ์ต่าง ๆ ที่ใช้สำหรับการพยากรณ์ข้อมูลของปริมาณความต้องการใช้ไฟฟ้าในแต่ละประเภท โดยใช้เทคนิคการพยากรณ์เชิงสถิติที่ประกอบด้วย การวิเคราะห์อนุกรมเวลา (Time Series Analysis) การวิเคราะห์การถดถอย (Regression Analysis) และการพยากรณ์ของการไฟฟ้าส่วนภูมิภาค ซึ่งมีวิธีการดังต่อไปนี้

2.1 การวิเคราะห์อนุกรมเวลา (Time Series Analysis)

การวิเคราะห์อนุกรมเวลา เป็นการศึกษาหารูปแบบการเปลี่ยนแปลงของตัวแปร ที่เปลี่ยนไปตามเวลาในอดีตจนถึงปัจจุบัน แล้วนำรูปแบบนั้นมาวิเคราะห์ เพื่อพยากรณ์ค่าของตัวแปรนั้นในอนาคต โดยเทคนิคหรือวิธีที่ใช้ในการวิเคราะห์ข้อมูลอนุกรมเวลามีหลายวิธี สำหรับวิธีการพยากรณ์ที่ใช้ในการศึกษาครั้งนี้ประกอบด้วย วิธีการบอกซ์ – เจนกินส์ (Box – Jenkins Methods) วิธีการปรับให้เรียบแบบเลขชี้กำลัง (Exponential Smoothing Methods) และวิธีการแยกองค์ประกอบ (Decomposition Method) ซึ่งมีวิธีการดังนี้

2.1.1 วิธีการบอกซ์ – เจนกินส์ (Box – Jenkins Methods)

วิธีการพยากรณ์ข้อมูลอนุกรมเวลาโดยทั่วไป จะมีข้อสมมติพื้นฐานข้อหนึ่งคือ อนุกรมเวลา $\{ \ldots , Y_{t-1} , Y_t , Y_{t+1} , \ldots \}$ ไม่มีสหสัมพันธ์ แต่ข้อสมมติดังกล่าว บ่อยครั้งพบว่าไม่เป็นจริง นั่นคือมีหลายกรณีที่อนุกรมเวลา $\{ \ldots , Y_{t-1} , Y_t , Y_{t+1} , \ldots \}$ มีสหสัมพันธ์ ถ้าเป็นเช่นนี้การใช้วิธีที่มีข้อสมมติว่า ตัวแปรอนุกรมเวลาไม่มีสหสัมพันธ์อาจจะไม่เหมาะสม ทั้งนี้เพราะวิธีการนี้ไม่ได้นำสหสัมพันธ์ที่ปรากฏไปใช้ประโยชน์ในการสร้างตัวแบบพยากรณ์ ฉะนั้นวิธีการสร้างตัวแบบพยากรณ์สำหรับอนุกรมเวลา ที่ได้นำเอาสหสัมพันธ์ที่ปรากฏไปวิเคราะห์ใช้ประโยชน์ จะเป็นวิธีที่ให้ผลการพยากรณ์ดีกว่า ซึ่งวิธีหนึ่งที่เป็นที่รู้จักและใช้กันมากคือ วิธีการบอกซ์ – เจนกินส์

โดยวิธีการบอกซ์ – เจนกินส์ จะหาตัวแบบอนุกรมเวลาโดยพิจารณาสหสัมพันธ์ระหว่าง Y ที่ตำแหน่งเวลา หรือคาบเวลา t (Y_t) และ Y ที่ตำแหน่งเวลา หรือคาบเวลาต่าง ๆ ที่ผ่านมา (Y_{t-1} ,

$Y_{t-2}, \ldots$) เมื่อได้ตัวแบบแล้ว ตัวแบบนี้จะแสดงความสัมพันธ์ระหว่าง Y_t กับ $Y_{t-1}, Y_{t-2}, \ldots$ และจะใช้ตัวแบบนี้พยากรณ์ค่า $Y_{t+1}, Y_{t+2}, \ldots$ ในอนาคต ตัวแบบบอกซ์ – เจนกินส์โดยทั่วไปจะใช้พยากรณ์ค่าในช่วงเวลาข้างหน้าที่เป็นระยะสั้น หรือปานกลาง ทั้งนี้เพราะตัวแบบโดยทั่วไปจะให้ความสำคัญ หรือน้ำหนักกับข้อมูลในอดีตระยะใกล้มากกว่าข้อมูลในอดีตที่ห่างไกลออกไปมาก ๆ

วิธีการบอกซ์ – เจนกินส์นี้ ใช้ได้กับข้อมูลอนุกรมเวลาที่มีองค์ประกอบใด ๆ เช่น แนวโน้ม และวัฏจักร หรือฤดูกาล ซึ่งอาจจะมีองค์ประกอบมากกว่าหนึ่งองค์ประกอบในอนุกรมชุดเดียวกัน

ลักษณะของตัวแบบบอกซ์ – เจนกินส์ มีแนวคิดมาจากการศึกษาวิเคราะห์กระบวนการเชิงเส้น หรือตัวกรองเชิงเส้น (linear filter) :

$$Y_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \ldots$$

นั่นคือ พิจารณาอนุกรมเวลาหรือค่าสังเกต Y_t เกิดจากผลบวกเชิงเส้นของตัวแปรสุ่ม a_t, $a_{t-1}, \ldots$ ที่ไม่มีสหสัมพันธ์กัน เรียกตัวแปรสุ่ม $a_t, a_{t-1}, \ldots$ ว่า ค่าผิดพลาดสุ่ม หรือเรียกว่า กระตุกสุ่ม (random shocks) โดยสมมติว่าแต่ละตัวมีค่าเฉลี่ยเท่ากับศูนย์ และมีความแปรปรวนคงที่ และโดยทั่วไปสมมติด้วยว่ามีการแจกแจงปกติ

จากตัวกรองเชิงเส้นหรือตัวแบบเชิงเส้น พารามิเตอร์ μ คือค่าระดับค่าเฉลี่ยของ Y_t เมื่ออนุกรมเวลาอยู่ในสภาพคงที่ และพารามิเตอร์ $\psi_1, \psi_2, \ldots$ เป็นน้ำหนักที่ให้กับตัวแปรสุ่ม a_{t-1}, $a_{t-2}, \ldots$ กระบวนการหรือตัวแบบเชิงเส้นนี้จะไม่ให้ประโยชน์ ถ้ามีพารามิเตอร์จำนวนอนันต์ (จำนวนไม่รู้จบ) เพราะฉะนั้นจะสร้างตัวแบบที่ประกอบด้วยพารามิเตอร์จำนวนจำกัด และเพียงพอที่จะอธิบายอนุกรมเวลาที่พิจารณา

1. ***ตัวแบบภายใต้ภาวะคงที่*** **(Stationary Models)**

1.1 ตัวแบบอัตถดถอย [Autoregressive (AR) Models]

จากตัวแบบเชิงเส้น พัฒนาตัวแบบเฉพาะกลุ่มหนึ่งเรียกว่า "ตัวแบบอัตถดถอยอันดับ p" หรือ "กระบวนการอัตถดถอยอันดับ p" (autoregressive process of order p) แทนด้วยตัวอักษรย่อ AR(p) และมีรูปแบบทั่วไปดังนี้

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \ldots + \phi_p(Y_{t-p} - \mu) + a_t$$

หรือ $Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \ldots + \phi_p Z_{t-p} + a_t$

หรือ $Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + a_t$

โดยที่ $Z_t = Y_t - \mu$, $Z_{t-1} = Y_{t-1} - \mu$, ... และ $c = \mu(1 - \phi_1 - \phi_2 - \ldots - \phi_p)$

และ ϕ_1 , ϕ_2 , ... , μ เป็นพารามิเตอร์ที่โดยทั่วไปไม่ทราบค่า จะประมาณค่าจากข้อมูล

จากตัวแบบ AR(p) ข้างต้น อาจเขียนในรูปแบบสั้น ๆ :

$$\phi_p(B) Z_t = a_t$$

โดยที่ $\phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^P)$

และ $BZ_t = Z_{t-1}$, $B^2 Z_t = Z_{t-2}$, ... , $B^P Z_t = Z_{t-p}$

เรียก B ว่า "ตัวถอยหลังเวลา" (backward – shift operator)

โดยทั่วไปในทางปฏิบัติ อันดับ p จะไม่สูงมาก เช่น 1, 2 หรือ 3 ถ้า p = 1 , 2 เขียนกระบวนการคงที่หรือตัวแบบคงที่ AR(1) และ AR(2) ได้ดังนี้

ตัวแบบ AR(1) : $Y_t = c + \phi_1 Y_{t-1} + a_t$

โดยมีเงื่อนไขของการเป็นตัวแบบคงที่ ดังนี้ $-1 < \phi_1 < 1$

ตัวแบบ AR(2) : $Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + a_t$

โดยมีเงื่อนไขของการเป็นตัวแบบคงที่ ดังนี้ $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$ และ $-1 < \phi_1 < 1$

1.2 ตัวแบบค่าเฉลี่ยเคลื่อนที่ [Moving Average (MA) Models]

จากตัวแบบเชิงเส้น พัฒนาตัวแบบเฉพาะอีกกลุ่มหนึ่งเรียกว่า "ตัวแบบค่าเฉลี่ยเคลื่อนที่อันดับ q" หรือ "กระบวนการค่าเฉลี่ยเคลื่อนที่อันดับ q" (moving average process of order q) แทนด้วยตัวอักษรย่อ MA(q) และมีรูปแบบทั่วไปดังนี้

$$Y_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q}$$

หรือ $$Z_t = \theta_q(B)\, a_t$$

โดยที่ $Z_t = Y_t - \mu$ และ $\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q)$

ตัวแบบบอกซ์ – เจนกินส์ ยังมีเงื่อนไขที่ต้องสอดคล้องอีกเงื่อนไขหนึ่งนอกเหนือจากเงื่อนไขคงที่ (stationarity) คือเงื่อนไขผกผันได้ (invertibility) ซึ่งพบว่าตัวแบบ AR(p), $p < \infty$ มีคุณสมบัติผกผันได้เสมอแต่อาจไม่คงที่ ขณะที่ตัวแบบ MA(q), $q < \infty$ มีคุณสมบัติคงที่เสมอแต่อาจจะผกผันไม่ได้ เพราะฉะนั้น ต้องตรวจสอบคุณสมบัติคงที่ในตัวแบบ AR และตรวจสอบคุณสมบัติการผกผันได้ในตัวแบบ MA

ทำนองเดียวกันกับตัวแบบ AR ในทางปฏิบัติจะมีอันดับ q ไม่สูงมาก เช่น 1, 2 หรือ 3 ถ้า $q = 1, 2$ จะได้ตัวแบบ MA(1) และ MA(2) ดังนี้

ตัวแบบ MA(1) : $$Y_t = \mu + a_t - \theta_1 a_{t-1}$$

โดยมีเงื่อนไขของการผกผันได้ ดังนี้ $-1 < \theta_1 < 1$

ตัวแบบ MA(2) : $$Y_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$$

โดยมีเงื่อนไขของการผกผันได้ ดังนี้ $\theta_1 + \theta_2 < 1$, $\theta_2 - \theta_1 < 1$ และ $-1 < \theta_1 < 1$

1.3 ตัวแบบผสมอัตถดถอย – ค่าเฉลี่ยเคลื่อนที่ [Autoregressive – Moving Average (ARMA) Models]

ในบางกรณีการใช้ตัวแบบ หรือกระบวนการผสมระหว่างตัวแบบ AR และ MA จะเป็นตัวแบบที่ประหยัด แทนการใช้ตัวแบบ AR อันดับสูง ๆ ตัวแบบเดียว หรือแทนการใช้ตัวแบบ MA อันดับสูง ๆ ตัวแบบเดียว โดยจะใช้สัญลักษณ์ ARMA(p, q) หมายถึง ตัวแบบผสมอัตถดถอยอันดับ p และค่าเฉลี่ยเคลื่อนที่อันดับ q และมีรูปแบบทั่วไปดังนี้

$$Y_t = \mu + \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \ldots + \phi_p(Y_{t-p} - \mu)$$
$$+ a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q}$$

หรือ $\phi_p(B)\,Z_t \;=\; \theta_q(B)\,a_t$

อันดับ p และ q สำหรับตัวแบบ ARMA จะไม่สูงมากในทางปฏิบัติ ถ้า p = 1 และ q = 1 จะได้ตัวแบบ ARMA(1,1) :

$$Y_t \;=\; \mu + \phi_1(Y_{t-1} - \mu) + a_t - \theta_1 a_{t-1}$$

หรือ $$Y_t \;=\; c + \phi_1 Y_{t-1} + a_t - \theta_1 a_{t-1}$$

โดยมีเงื่อนไขของความคงที่และผกผันได้ ดังนี้ $-1 < \phi_1 < 1$ และ $-1 < \theta_1 < 1$

2. **ตัวแบบภายใต้ภาวะไม่คงที่** (Nonstationary Models) **และตัวแบบ** ARIMA

ถ้าข้อมูลอนุกรมเวลาหรือกระบวนการ Y_t ไม่อยู่ในสภาพคงที่ในค่าเฉลี่ย และ / หรือความแปรปรวน นักพยากรณ์จะต้องแปลงข้อมูลนั้นให้อยู่ในสภาพคงที่ก่อนพิจารณากำหนดตัวแบบ

การแปลงข้อมูลอนุกรมเวลาที่ไม่คงที่ในค่าเฉลี่ย จะใช้วิธีการทำผลต่างโดยนำข้อมูลมาลบกันได้เป็นข้อมูลชุดใหม่ นอกจากนี้การทำผลต่างดังกล่าว อาจจะต้องทำมากกว่าหนึ่งครั้งจึงจะทำให้อนุกรมเวลามีค่าเฉลี่ยคงที่ ซึ่งโดยทั่วไปถ้าอนุกรมเวลามีแนวโน้ม มักจะทำผลต่างสองครั้งจึงจะคงที่ การทำผลต่างไม่ควรจะทำหลายครั้งมากเกินความจำเป็น เพราะจะมีผลทำให้ค่าพยากรณ์มีความคลาดเคลื่อนสูง

ในกรณีที่อนุกรมเวลาไม่คงที่ในความแปรปรวน หรือมีการเคลื่อนไหวเป็นเส้นโค้ง วิธีการแปลงข้อมูลที่ใช้กันมากคือ ใส่ ln ในอนุกรม Y_t ได้เป็นข้อมูลอนุกรมใหม่ $X_t = \ln Y_t$ ซึ่ง $Y_t > 0$, $t = 1, 2, \ldots, n$ วิธีนี้มักจะใช้เมื่อความแปรปรวนแปรผันตามระดับค่าเฉลี่ย เช่น ความแปรปรวนมากขึ้นขณะที่อนุกรมมีแนวโน้มเพิ่มขึ้น หรือมีระดับค่าเฉลี่ยสูงขึ้น

ถ้าอนุกรมเวลาไม่คงที่ทั้งในค่าเฉลี่ย และความแปรปรวน ในกรณีนี้จะต้องแปลงข้อมูลให้คงที่ทั้งในค่าเฉลี่ย และความแปรปรวน และโดยทั่วไปจะแปลงข้อมูลในเรื่องความแปรปรวนก่อนแล้วจึงทำผลต่างในข้อมูลที่ถูกแปลงแล้ว ทั้งนี้เพราะถ้าทำผลต่างก่อนอาจแปลงแก้ความแปรปรวนไม่ได้ เช่น ถ้าผลต่างมีค่าเป็นลบ และถ้าใช้วิธีหาค่า ln จะหาค่า ln ไม่ได้

เมื่ออนุกรมเวลามีสภาพไม่คงที่ หรือไม่เคลื่อนไหวรอบค่าเฉลี่ยคงที่ค่าหนึ่งค่าเดียว จะต้องแปลงข้อมูลนั้นให้อยู่ในสภาพคงที่ ฉะนั้นถ้ามีการทำผลต่าง d ครั้ง จะเขียนตัวแบบผสมเป็น ARIMA (Autoregressive Integrated Moving Average Models) ด้วยอันดับ (p,d,q) มีรูปแบบทั่วไปดังนี้

ตัวแบบ ARIMA(p,d,q) :

$$\phi_p(B)(1-B)^d Y_t = \delta + \theta_q(B)a_t$$

หรือ $$\phi_p(B)W_t = \delta + \theta_q(B)a_t$$

โดยที่ $W_t = (1-B)^d Y_t$ และ δ (อาจมีค่าเท่ากับศูนย์) เป็นพารามิเตอร์แสดงระดับค่าเฉลี่ยคงที่ของอนุกรม W_t และ Y_t เป็นอนุกรมที่ถูกแปลงให้มีความแปรปรวนคงที่แล้ว ถ้าอนุกรมแรกเริ่มไม่คงที่ในความแปรปรวน

3. **ตัวแบบ ARIMA เมื่อมีองค์ประกอบฤดูกาล**

องค์ประกอบที่เป็นฤดูกาลในที่นี่ จะไม่จำกัดเฉพาะการแปรผันของข้อมูลคล้ายคลึงกันในระยะเวลาห่างกัน 1 ปี (คล้ายคลึงกันจากปีหนึ่งไปยังอีกปีหนึ่ง) แต่จะรวมถึงความแปรผันคล้ายคลึงกันในช่วงเวลาห่างอื่น ๆ ด้วย

สำหรับอนุกรมเวลาที่มีองค์ประกอบฤดูกาล โดยมีคาบเวลาของฤดูกาล $S > 1$ ตัวแบบอนุกรมเวลาในส่วนที่เป็นฤดูกาล จะมีโครงสร้างเป็นไปได้เหมือนกับองค์ประกอบที่ไม่ใช่ฤดูกาล นั่นคือ จะมีตัวแบบ ARIMA ด้วยอันดับ $(P,D,Q)_S$ ซึ่ง P คืออันดับในส่วนของกระบวนการ AR , Q คืออันดับในส่วนของกระบวนการ MA และ D คือจำนวนครั้งทำผลต่างอนุกรมเวลาห่างกัน S คาบเวลา

เมื่อนำองค์ประกอบในส่วนที่ไม่ใช่ฤดูกาล และส่วนที่เป็นฤดูกาลมาผนวกเข้าด้วยกัน จะได้ตัวแบบ ARIMA ที่แสดงส่วนประกอบทั้งสอง และตัวแบบทั่วไปตัวแบบหนึ่ง คือ ตัวแบบในรูปผลคูณ ARIMA$(p,d,q)(P,D,Q)_S$ มีรูปแบบดังนี้

ตัวแบบ ARIMA$(p,d,q)(P,D,Q)_S$:

$$\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D Y_t = \delta + \theta_q(B)\Theta_Q(B^S)a_t$$

โดยที่ $\Phi_P(B^S) = (1 - \Phi_S B^S - \Phi_{2S} B^{2S} - \ldots - \Phi_{PS} B^{PS})$

$\Theta_Q(B^S) = (1 - \Theta_S B^S - \Theta_{2S} B^{2S} - \ldots - \Theta_{PS} B^{PS})$

การพิจารณากำหนดอันดับ (p, d, q) และ $(P, D, Q)_S$ จะพิจารณาแยกจากกัน แต่ใช้หลักการพิจารณาเหมือนกัน เมื่ออันดับ p และ q คือ อันดับของกระบวนการ AR และ MA ในส่วนที่ไม่ใช่ฤดูกาล และอันดับ P และ Q คือ อันดับของกระบวนการ AR และ MA ในส่วนที่เป็นฤดูกาล สำหรับ d คือ จำนวนครั้งที่ทำผลต่างอนุกรมเวลา เมื่ออนุกรมเวลาในส่วนที่ไม่ใช่ฤดูกาลมีสภาพไม่คงที่ในค่าเฉลี่ย

กระบวนการ AR และ MA ต่างมีรูปแบบโครงสร้างเฉพาะสำหรับอันดับ p และ q ของฟังก์ชันอัตสหสัมพันธ์ ACF (Autocorrelation Function) แทนด้วย ρ_k และโครงสร้างของฟังก์ชันอัตสหสัมพันธ์ย่อย PACF (Partial Autocorrelation Function) แทนด้วย ϕ_{kk} ซึ่ง k หมายถึงคาบเวลาห่างระหว่างอนุกรม และเรียกคาบเวลานี้ว่า "แล็ก k" (lag k) ฉะนั้น ρ_1 คือ อัตสหสัมพันธ์ที่แล็ก 1 หรืออัตสหสัมพันธ์ของอนุกรมเวลาที่ห่างกัน 1 หน่วย หรือ 1 คาบเวลา (Y_t, Y_{t+1}), t = 1, 2, . . . ซึ่งวัดความสัมพันธ์เชิงเส้นระหว่างอนุกรมเวลาที่ห่างกัน 1 คาบเวลา สำหรับ ϕ_{kk} เป็นอัตสหสัมพันธ์ที่วัดความสัมพันธ์ ระหว่างอนุกรมเวลาที่ห่างกัน k คาบเวลา (Y_t, Y_{t+k}) โดยพิจารณาจากผลกระทบจากอนุกรมเวลา $Y_{t+1}, Y_{t+2}, \ldots, Y_{t+k-1}$ เข้ามาด้วย ค่าของ ρ_k และ ϕ_{kk} ต่างมีค่าอยู่ระหว่าง -1 และ 1

เพราะฉะนั้นการกำหนดอันดับ จะประมาณค่า ρ_k และ ϕ_{kk} โดยใช้ข้อมูลอนุกรมเวลาที่นำมาวิเคราะห์ แทนค่าประมาณด้วย $\hat{\rho}_k$ และ $\hat{\phi}_{kk}$ และเรียกว่า "ฟังก์ชันอัตสหสัมพันธ์ตัวอย่าง" SACF (Sample Autocorrelation Function) และ "ฟังก์ชันอัตสหสัมพันธ์ย่อยตัวอย่าง" SPACF (Sample Partial Autocorrelation Function) ค่า $\hat{\rho}_k$ และ $\hat{\phi}_{kk}$ ที่แล็ก k = 1, 2, . . . , m ควรให้มีจำนวนค่ามากพอ ที่จะพิจารณาโครงสร้างการแปรผันของอัตสหสัมพันธ์ได้ง่าย ซึ่งจะพิจารณาค่า $\hat{\rho}_k$ และ $\hat{\phi}_{kk}$ แต่ละประเภท จำนวน m = n/4 โดยประมาณ จากค่าประมาณเหล่านี้ จะเปรียบเทียบลักษณะแปรผันกับโครงสร้างแปรผันของ $\hat{\rho}_k$ และ $\hat{\phi}_{kk}$ ทางทฤษฎี และเลือกตัวแบบ ARIMA ที่โครงสร้างของ $\hat{\rho}_k$ และ $\hat{\phi}_{kk}$ เข้ากับโครงสร้าง ρ_k และ ϕ_{kk} ของ ARIMA นั้นมากที่สุด ด้วยวิธีการเลือกตัวแบบ ARIMA โดยการเปรียบเทียบโครงสร้างแปรผันของค่าตัวอย่าง $\hat{\rho}_k$ และ $\hat{\phi}_{kk}$ กับ

โครงสร้างแปรผันของค่าทางทฤษฎี ρ_k และ ϕ_{kk} จึงจำเป็นที่นักพยากรณ์จะต้องทราบลักษณะโครงสร้างเชิงทฤษฎีของ ρ_k และ ϕ_{kk} ของกระบวนการ AR , MA และ ARMA ที่อันดับต่าง ๆ เพื่อจะได้เลือกกระบวนการ และอันดับ เป็นตัวแบบ ARIMA ทดลอง สำหรับข้อมูลอนุกรมเวลาที่พิจารณา

สรุปลักษณะแปรผันของ ACF และ PACF ของกระบวนการอนุกรมเวลาคงที่ สำหรับกระบวนการพื้นฐานดังนี้

กระบวนการ	ACF	PACF
AR(1)	ค่า ρ_k ลดลงอย่างรวดเร็ว ขณะที่ $k > 1$	ค่า ϕ_{kk} จะมีค่าสูงที่ $k = 1$ และเท่ากับ 0 เมื่อ $k > 1$
AR(2)	ค่า ρ_k ลดลงอย่างรวดเร็ว ขณะที่ $k > 1$	ค่า ϕ_{kk} จะมีค่าสูงที่ $k = 1,2$ และเท่ากับ 0 เมื่อ $k > 2$
MA(1)	ค่า ρ_k จะมีค่าสูงที่ $k = 1$ และเท่ากับ 0 เมื่อ $k > 1$	ค่า ϕ_{kk} ลดลงอย่างรวดเร็ว ขณะที่ $k > 1$
MA(2)	ค่า ρ_k จะมีค่าสูงที่ $k = 1,2$ และเท่ากับ 0 เมื่อ $k > 2$	ค่า ϕ_{kk} ลดลงอย่างรวดเร็ว ขณะที่ $k > 1$
ARMA(1,1)	ค่า ρ_k ลดลงอย่างรวดเร็ว หลังจากแล็ก $k = 1$	ค่า ϕ_{kk} ลดลงอย่างรวดเร็ว หลังจากแล็ก $k = 1$

สำหรับกระบวนการ AR , MA และ ARMA ที่อันดับอื่น ๆ พิจารณาได้ในทำนองเดียวกัน

ในองค์ประกอบที่เป็นฤดูกาลมีคาบเวลา S การกำหนดอันดับ P และ Q พิจารณาทำนองเดียวกันกับองค์ประกอบที่ไม่ใช่ฤดูกาล โดยพิจารณาโครงสร้างแปรผันของอัตสหสัมพันธ์ $\hat{\rho}_k$ และ $\hat{\phi}_{kk}$ ที่แล็ก ฤดูกาล S , 2S , 3S , . . . เปรียบเทียบกับโครงสร้างของ ρ_k และ ϕ_{kk} ทางทฤษฎี ซึ่งมีลักษณะตามที่กล่าวไปแล้วข้างต้น

การวินิจฉัยตัวแบบ ARIMA จะกระทำการตรวจสอบคุณสมบัติเชิงสถิติของค่าผิดพลาดสุ่ม $\{a_t\}$ และการทดสอบว่าค่าผิดพลาดสุ่มนั้นมีอัตสหสัมพันธ์หรือไม่ จะเป็นการตรวจสอบที่สำคัญมากที่สุดในการวินิจฉัยความเพียงพอในเชิงสถิติของตัวแบบ ARIMA ฉะนั้นการทดสอบจะคำนวณค่า SACF และ SPACF ของค่าเศษเหลือตกค้าง $e_t = y_t - \hat{y}_t$ ซึ่งเป็นค่าประมาณของ a_t ที่แล็ก k ต่าง ๆ และทดสอบด้วยค่าของตัวสถิติที (t) สำหรับทดสอบว่า ค่าผิดพลาดสุ่มมีอัตสหสัมพันธ์หรือไม่ ที่แต่ละแล็ก $k = 1, 2, \ldots, m$ และทดสอบอัตสหสัมพันธ์รวม หรือพร้อมกัน k แล็ก ด้วยตัวสถิติไคกำลังสอง (chi – squared test) ว่าค่าผิดพลาดไม่มีอัตสหสัมพันธ์ k แล็กแรก

นอกจากการวินิจฉัยตัวแบบ ด้วยการทดสอบเชิงสถิติแล้ว นักพยากรณ์อาจตรวจสอบด้วยวิธีการอื่น ๆ ด้วย เช่น การเขียนกราฟของค่าเศษเหลือตกค้างกับแกนเวลา ถ้าพบว่าค่าของเศษเหลือตกค้าง กระจายเป็นแนวในลักษณะขนานรอบค่าเฉลี่ยศูนย์ แสดงเหตุผลได้ว่า ค่าผิดพลาดมีค่าเฉลี่ยศูนย์ และมีความแปรปรวนคงที่ แต่ถ้าการกระจายของค่าเศษเหลือตกค้าง มีรูปแบบที่ต่างไปจากแนวขนาน ควรพิจารณาปรับปรุงแก้ไขตัวแบบ ซึ่งอาจจะพบว่าความแปรปรวนยังไม่คงที่ (ถ้าความแปรปรวนแปรเปลี่ยนตามเวลา) ต้องปรับให้คงที่ด้วยวิธีการแปลงข้อมูล เป็นต้น

ผลจากการวินิจฉัยตัวแบบ นอกจากจะช่วยตรวจสอบว่า ตัวแบบที่กำลังพิจารณาเหมาะสมเพียงพอในเชิงสถิติหรือไม่แล้ว ยังให้แนวทางในการปรับปรุงแก้ไขตัวแบบด้วย ถ้าพบว่าตัวแบบยังไม่เหมาะสม กล่าวคือจากลักษณะของ SACF และ SPACF ของเศษเหลือตกค้าง อาจพบว่าควรเพิ่มองค์ประกอบ MA เข้าในตัวแบบ ถ้ายังไม่มีองค์ประกอบ MA หรือเพิ่มอันดับของ MA ให้มากขึ้น หรืออาจพบว่าควรเพิ่มองค์ประกอบ AR หรืออันดับของ AR ในตัวแบบ เป็นต้น

2.1.2 วิธีการปรับให้เรียบแบบเลขชี้กำลัง (Exponential Smoothing Methods)

วิธีการปรับให้เรียบแบบเลขชี้กำลัง เป็นวิธีการพยากรณ์ข้อมูลอนุกรมเวลา โดยการกำจัดอิทธิพลของความผันแปรที่ไม่แน่นอนออกไป เพื่อที่จะทำข้อมูลอนุกรมเวลาให้เรียบขึ้น ซึ่งจะทำให้สามารถพยากรณ์หรือประมาณค่าตัวแปรในอนาคตได้ นอกจากนั้นวิธีนี้ยังเป็นวิธีที่แก้ข้อเสียของวิธีการพยากรณ์โดยใช้ค่าเฉลี่ยเคลื่อนที่ เนื่องจากวิธีค่าเฉลี่ยเคลื่อนที่ เป็นวิธีที่ให้น้ำหนักกับข้อมูลทุกค่าเท่ากันหมด แต่วิธีนี้มีหลักการว่าจะให้ความสำคัญหรือน้ำหนักกับข้อมูลไม่เท่ากัน วิธีการปรับให้เรียบแบบเลขชี้กำลังมีหลายวิธี ซึ่งแต่ละวิธีจะเหมาะสมกับข้อมูลแต่ละประเภท ในการทำวิจัยครั้งนี้จะกล่าวถึง วิธีการพยากรณ์ของวินเตอร์ (Winters ' Forecast Method) ดังนี้

วิธีการของวินเตอร์ เหมาะสำหรับใช้ในการพยากรณ์ข้อมูลที่มีแนวโน้ม และมีอิทธิพลของฤดูกาล (trend – season data) และใช้พยากรณ์ระยะสั้นจนถึงปานกลาง ข้อมูลไม่ควรเป็นรายปี เพราะจะทำให้ไม่สามารถแยกอิทธิพลของฤดูกาลได้ ข้อมูลควรอยู่ในรูปรายเดือน รายสัปดาห์ รายไตรมาส เป็นต้น วิธีการของวินเตอร์นี้ ยังคงใช้หลักการปรับให้เรียบแบบเลขชี้กำลัง คือให้น้ำหนักกับข้อมูลไม่เท่ากัน และมีค่าคงที่ปรับให้เรียบ 3 ค่า คือ

α (Alpha) = ค่าคงที่ปรับให้เรียบสำหรับระดับ , $0 \leq \alpha \leq 1$

γ (Gamma) = ค่าคงที่ปรับให้เรียบสำหรับแนวโน้ม , $0 \leq \gamma \leq 1$

δ (Delta) = ค่าคงที่ปรับให้เรียบสำหรับฤดูกาล , $0 \leq \delta \leq 1$

ตัวแบบอนุกรมเวลาตัวแบบหนึ่งของวินเตอร์ ซึ่งประกอบด้วยองค์ประกอบแนวโน้ม และฤดูกาลมีสมการดังนี้ และมีชื่อเรียกว่าตัวแบบผลคูณของวินเตอร์

$$Y_t = (\mu_t + \beta_t t) I_t + \varepsilon_t$$

โดยที่ μ_t, β_t, I_t เป็นพารามิเตอร์แสดงระดับ ความชัน และฤดูกาล ของอนุกรมเวลาตามลำดับ และ ε_t คือ ความคลาดเคลื่อนสุ่ม โดยมีข้อสมมติพื้นฐาน คือ มีค่าเฉลี่ยศูนย์ ความแปรปรวนคงที่ และไม่มีสหสัมพันธ์

ตัวแบบข้างต้นเหมาะสมกับอนุกรมเวลา ที่มีการแกว่งหรือการผันแปรของฤดูกาล เป็นสัดส่วนโดยตรงกับระดับของอนุกรมเวลา (ค่าเฉลี่ยของอนุกรม) กล่าวคือ การแกว่งจะมากขึ้นขณะที่ระดับของอนุกรมเวลาเพิ่มขึ้น ส่วนการผันแปรของ ε_t ไม่ขึ้นอยู่กับระดับของอนุกรมเวลา

จากตัวแบบข้างต้น วิเคราะห์ได้ตัวแบบพยากรณ์ดังนี้

$$\hat{Y}_t(\ell) = (\hat{\mu}_t + \ell \hat{\beta}_t) \hat{I}_{t+1-m} \qquad \text{สำหรับ } t = m, m+1, \ldots$$

โดยที่ $\hat{\mu}_t = \alpha (Y_t / \hat{I}_{t-m}) + (1-\alpha)(\hat{\mu}_{t-1} + \hat{\beta}_{t-1})$

$\hat{\beta}_t = \gamma (\hat{\mu}_t - \hat{\mu}_{t-1}) + (1-\gamma)\hat{\beta}_{t-1}$

$\hat{I}_t = \delta (Y_t / \hat{\mu}_t) + (1-\delta)\hat{I}_{t-m}$

m = ความยาวของคาบฤดูกาล เช่น m = 12 (สำหรับอนุกรมเวลารายเดือน) หรือ m = 4 (สำหรับอนุกรมเวลารายไตรมาส)

การคำนวณค่าพยากรณ์ $\hat{Y}_t(\ell)$ ต้องกำหนดค่าเริ่มต้นของ $\hat{\mu}_t$, $\hat{\beta}_t$ และ $\hat{I}_t$ นอกเหนือจากการกำหนดค่าคงที่ปรับให้เรียบ α, γ, δ และทางหนึ่งในการกำหนดค่าเริ่มต้นคือให้

$$\hat{\mu}_m = (Y_1 + Y_2 + \ldots + Y_m)/m$$

$$\hat{I}_t = Y_t / \hat{\mu}_m \qquad \text{สำหรับ } t = 1, 2, \ldots, m$$

$$\hat{\beta}_m = 0$$

2.1.3 วิธีการแยกองค์ประกอบ (Decomposition Method)

วิธีการแยกองค์ประกอบนี้ เป็นวิธีการพยากรณ์ข้อมูลอนุกรมเวลาในระยะปานกลาง โดยวิเคราะห์เพื่อแยกองค์ประกอบของข้อมูลอนุกรมเวลาออกได้เป็น 3 องค์ประกอบใหญ่ ๆ คือ แนวโน้ม (T_t) ฤดูกาล (S_t) และเหตุการณ์ที่ผิดปกติ (ε_t) แล้วทำการพยากรณ์องค์ประกอบแต่ละส่วนเพื่อนำมาพยากรณ์ค่าของตัวแปรที่สนใจในอนาคต โดยควรจะทราบถึงรูปแบบของข้อมูลในอดีตจนถึงปัจจุบัน และถือว่าปัจจัยที่เกี่ยวข้องในอนาคตเหมือนกับในอดีต สำหรับรูปแบบของข้อมูลอนุกรมเวลาที่จะกล่าวถึงมี 2 รูปแบบ คือ

1. รูปแบบผลบวก (Additive Model) รูปแบบนี้เหมาะสำหรับข้อมูลอนุกรมเวลา ที่มีความแปรปรวนหรือการผันแปรของฤดูกาล ไม่แปรผันตามระดับค่าเฉลี่ยหรือแนวโน้ม มีรูปแบบดังนี้

$$Y_t = T_t + S_t + \varepsilon_t$$

2. รูปแบบผลคูณ (Multiplicative Model) รูปแบบนี้เหมาะสำหรับข้อมูลอนุกรมเวลา ที่มีความแปรปรวนหรือการผันแปรของฤดูกาล แปรผันตามระดับค่าเฉลี่ยหรือแนวโน้ม มีรูปแบบดังนี้

$$Y_t = T_t \cdot S_t \cdot \varepsilon_t$$

จากรูปแบบผลคูณข้างต้น สามารถแปลงให้อยู่ในรูปแบบผลบวกได้ ซึ่งใช้วิธีการแปลงค่าของตัวแปรตาม โดยการใส่ *ln* ในอนุกรม Y_t จะได้รูปแบบดังนี้

$$\ln Y_t = \ln T_t + \ln S_t + \ln \varepsilon_t$$

การพยากรณ์โดยวิธีการแยกองค์ประกอบนี้ จะพิจารณาองค์ประกอบที่มีอิทธิพล ต่อการเปลี่ยนแปลงของข้อมูลแต่ละส่วน ดังนี้

แนวโน้ม (Trend : T) การคำนวณหาแนวโน้มจะใช้หลักการของการวิเคราะห์การถดถอย โดยให้ตัวแปรอิสระคือเวลา (t) และตัวแปรตามคือแนวโน้ม ซึ่งแสดงถึงแนวโน้มของตัวแปร หรือข้อมูลอนุกรมเวลาที่สนใจในอนาคต ซึ่งมีรูปแบบทั่วไปเป็นพหุนาม (polynomial) ที่ระดับองศาต่ำ ๆ ดังนี้

$$T_t = \beta_0 + \sum_{i=1}^{k} \beta_i \frac{t^i}{i!}$$

โดยที่ k = ระดับองศาของพหุนาม (polynomial of degree)

ถ้า k = 1 : $T_t = \beta_0 + \beta_1 t$ เมื่อแนวโน้มมีลักษณะเป็นเส้นตรง (linear trend)

ถ้า k = 2 : $T_t = \beta_0 + \beta_1 t + \beta_2 (t^2/2)$ เมื่อแนวโน้มมีลักษณะเป็นเส้นโค้ง (quadratic trend)

ฤดูกาล (Seasonal : S) การคำนวณหาอิทธิพลของฤดูกาล จะใช้ตัวแปรบ่งชี้ที่สร้างขึ้นเพื่อวัดอิทธิพลของฤดูกาล ที่จะระบุว่าข้อมูลหนึ่ง ๆ เป็นข้อมูลที่เกิดขึ้นในช่วงฤดูกาลใด ซึ่งมีรูปแบบดังนี้

$$S_t = \sum_{i=1}^{m-1} \delta_i I_{it}$$

โดยที่ m = ความยาวของคาบฤดูกาล

เช่น m = 12 (สำหรับอนุกรมเวลารายเดือน)

หรือ m = 4 (สำหรับอนุกรมเวลารายไตรมาส)

และ I_{it} = ตัวแปรบ่งชี้ (indicator variable)

ซึ่งค่า $I_{it} = 1$ เมื่อข้อมูล Y_t อยู่ในช่วงฤดูกาลที่ i

และ $I_{it} = 0$ เมื่อข้อมูล Y_t ไม่อยู่ในช่วงฤดูกาลที่ i

เหตุการณ์ที่ผิดปกติ (Error : ε) หรือความผันแปรที่ไม่แน่นอน เนื่องจากเป็นความผันแปรที่อาจเกิดขึ้นหรือไม่เกิดขึ้นก็ได้ ดังนั้นจึงไม่นำค่าความผันแปรที่ไม่แน่นอน มาพยากรณ์ค่าของข้อมูลในอนาคต

2.2 การวิเคราะห์การถดถอย (Regression Analysis)

การวิเคราะห์การถดถอย เป็นเทคนิคเชิงสถิติเทคนิคหนึ่งสำหรับการศึกษาวิเคราะห์ และจำลองรูปแบบความสัมพันธ์พึ่งพิงเชิงคณิตศาสตร์ระหว่างตัวแปรสองกลุ่ม ตัวแปรกลุ่มหนึ่งเรียกว่า "ตัวแปรตาม" (dependent variable) หรือ "ตัวแปรผล" (response variable) มีหนึ่งตัวแปร เป็นตัวแปรที่นักสถิติหรือนักพยากรณ์ สนใจที่จะศึกษาลักษณะการเปลี่ยนแปลง หรือพยากรณ์ค่า หรือควบคุม โดยศึกษาวิเคราะห์หารูปแบบความสัมพันธ์กับตัวแปรอีกกลุ่มหนึ่ง เรียกตัวแปรในกลุ่มนี้ว่า "ตัวแปรอิสระ" (independent variables) หรือ "ตัวแปรให้ค่าพยากรณ์หรือค่าทำนาย" (predictor variables) ตัวแปรในกลุ่มนี้อาจมีหนึ่ง หรือมากกว่าหนึ่งตัวแปร และรูปแบบความสัมพันธ์เชิงคณิตศาสตร์หรือเชิงสถิติที่ได้ เรียกว่า "ตัวแบบการถดถอย" หรือ "สมการการถดถอย" จากสมการการถดถอย สามารถอธิบายลักษณะการเปลี่ยนแปลงของตัวแปรตาม หรือพยากรณ์ค่าของตัวแปรตาม หรือใช้ในการควบคุมตัวแปรตาม โดยใช้รูปแบบสมการและค่าของตัวแปรอิสระ

การหาสมการการถดถอย ในกรณีวิเคราะห์รูปแบบความสัมพันธ์พึ่งพิงเชิงเส้น ของตัวแปรตาม Y บนตัวแปรอิสระกลุ่มหนึ่ง $X_1, X_2, \ldots, X_k\ (k \geq 1)$ เราเรียกรูปแบบความสัมพันธ์นั้นว่า "ตัวแบบการถดถอยเชิงเส้น" (linear regression model) และมีรูปแบบทั่วไปดังนี้

$$\begin{aligned} Y &= E(Y) + \varepsilon \\ &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon \end{aligned}$$

โดยที่ ค่าเฉลี่ย $E(Y)$ หรือ $E(Y|X_1, X_2, \ldots, X_k)$ ของ Y เท่ากับ $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$ เมื่อกำหนดตัวแปรอิสระ $X_1, X_2, \ldots, X_k$

สำหรับตัวอย่างสุ่ม $(X_{1t}, X_{2t}, \ldots, X_{kt}, Y_t)$, $t = 1, 2, \ldots, n$ นั่นคือ $X_1 = X_{1t}$, $X_2 = X_{2t}, \ldots, X_k = X_{kt}$, $Y = Y_t$ และให้ $\varepsilon = \varepsilon_t$ สำหรับ $t = 1, 2, \ldots, n$ ซึ่งเขียนตัวแบบการถดถอยเชิงเส้นได้ใหม่ดังนี้

$$\begin{aligned} Y_t &= E(Y_t) + \varepsilon_t \\ &= \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \ldots + \beta_k X_{kt} + \varepsilon_t, \quad \text{สำหรับ } t = 1, 2, \ldots, n \end{aligned}$$

พารามิเตอร์ หรือค่าคงที่ไม่ทราบค่า $\beta_0, \beta_1, \ldots, \beta_k$ มีชื่อเรียกว่า "สัมประสิทธิ์การถดถอย" (regression coefficients) สำหรับ β_j ($j \neq 0$) คือ อัตราเปลี่ยนแปลงโดยเฉลี่ยใน Y เมื่อค่าของตัวแปร X_j เปลี่ยนแปลงไปหนึ่งหน่วย ขณะที่ค่าของตัวแปร X_i ($i \neq j$) ที่เหลือทั้งหมดไม่เปลี่ยนแปลง และสำหรับ β_0 คือ ค่าเฉลี่ยของ Y เมื่อตัวแปรอิสระทุกตัวมีค่าเท่ากับศูนย์ แต่ทั้งนี้ β_0 จะไม่ให้ความหมาย ถ้าหากตัวแบบไม่ครอบคลุมกรณีตัวแปรอิสระทุกตัวเท่ากับศูนย์

สำหรับ ε_t เป็นตัวแปรสุ่ม มีชื่อเรียกว่า "ค่าผิดพลาดสุ่ม" (random error) เป็นค่าคลาดเคลื่อนหรือค่าผิดพลาดต่าง ๆ ที่เกิดขึ้นจากการสร้างตัวแบบ ซึ่งจะรวมผลกระทบต่อค่า Y จากตัวแปรอิสระอื่น ๆ ที่ไม่ใช่ $X_1, X_2, \ldots, X_k$ และอาจมีค่าผิดพลาดซึ่งเกิดขึ้นได้อย่างสุ่มในการวัดค่าหรือบันทึกค่าของ Y_t โดยคุณสมบัติพื้นฐานของ ε_t คือมีค่าเฉลี่ยเท่ากับศูนย์ ความแปรปรวนคงที่เท่ากับ σ^2 สำหรับทุกค่า t และความแปรปรวนร่วมของ $(\varepsilon_i, \varepsilon_j)$, $i \neq j$ ทุกคู่เท่ากับศูนย์หรือไม่มีสหสัมพันธ์กัน และ ε_t มีการแจกแจงความน่าจะเป็นแบบปกติ นั่นคือ ε_t มีข้อสมมติดังนี้

$$E(\varepsilon_t) = 0, \; t = 1, 2, \ldots, n$$

$$V(\varepsilon_t) = E(\varepsilon_t^2) = \sigma^2, \; t = 1, 2, \ldots, n$$

$$Cov(\varepsilon_i, \varepsilon_j) = E(\varepsilon_i, \varepsilon_j) = 0, \; i \neq j \; ; \; i, j = 1, 2, \ldots, n$$

$$\varepsilon_t \sim N(0, \sigma^2), \; t = 1, 2, \ldots, n$$

ด้วยวิธีกำลังสองน้อยที่สุดแบบสามัญ (Ordinary Least Squares Method : OLS) ตัวประมาณแบบกำลังสองน้อยที่สุด $B_0, B_1, \ldots, B_k$ ของ $\beta_0, \beta_1, \ldots, \beta_k$ ตามลำดับ คำนวณหาได้โดยการแก้สมการ

$$\left.\frac{\partial Q}{\partial \beta_i}\right|_{\beta_0 = B_0, \ldots, \beta_k = B_k} = 0, \quad i = 0, 1, 2, \ldots, k$$

โดยที่ $Q = \sum_{t=1}^{n} \varepsilon_t^2 = \sum_{t=1}^{n} (Y_t - \beta_0 - \beta_1 x_{1t} - \ldots - \beta_k x_{kt})^2$

หรือนั่นคือ โดยการแก้ระบบสมการปกติดังนี้

$$nB_0 + B_1\sum_{t=1}^{n} X_{1t} + B_2\sum_{t=1}^{n} X_{2t} + \ldots + B_k\sum_{t=1}^{n} X_{kt} = \sum_{t=1}^{n} Y_t$$

$$B_0\sum_{t=1}^{n} X_{1t} + B_1\sum_{t=1}^{n} X_{1t}^2 + B_2\sum_{t=1}^{n} X_{1t}X_{2t} + \ldots + B_k\sum_{t=1}^{n} X_{1t}X_{kt} = \sum_{t=1}^{n} X_{1t}Y_t$$

$$B_0\sum_{t=1}^{n} X_{2t} + B_1\sum_{t=1}^{n} X_{2t}X_{1t} + B_2\sum_{t=1}^{n} X_{2t}^2 + \ldots + B_k\sum_{t=1}^{n} X_{2t}X_{kt} = \sum_{t=1}^{n} X_{2t}Y_t$$

$$\vdots$$

$$B_0\sum_{t=1}^{n} X_{kt} + B_1\sum_{t=1}^{n} X_{kt}X_{1t} + B_2\sum_{t=1}^{n} X_{kt}X_{2t} + \ldots + B_k\sum_{t=1}^{n} X_{kt}^2 = \sum_{t=1}^{n} X_{kt}Y_t$$

$$\underset{\sim}{Y} \sim N_n(\underset{\sim}{X}\underset{\sim}{\beta}, \underset{\sim}{I}\sigma^2)$$

ในกรณีการวิเคราะห์การถดถอยเชิงพหุคูณ การใช้เมตริกซ์จะช่วยทำให้การวิเคราะห์ ง่ายและสะดวกยิ่งขึ้น ดังนั้นจากระบบสมการปกติอาจเขียนในเทอมของเมตริกซ์ได้สั้น ๆ ดังนี้

$$(\underset{\sim}{X}'\underset{\sim}{X})\underset{\sim}{B} = \underset{\sim}{X}'\underset{\sim}{Y}$$

ซึ่ง $\underset{\sim}{X}'$ หมายถึง เมตริกซ์สับเปลี่ยน (transposed matrix) ของเมตริกซ์ $\underset{\sim}{X}$

$$\underset{\sim}{X} = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 1 & X_{12} & X_{22} & & X_{k2} \\ \vdots & & & & \\ 1 & X_{1n} & X_{2n} & & X_{kn} \end{bmatrix}$$

การแก้สมการปกติ หา $\underset{\sim}{B}$ จะสมมติว่าหาเมตริกซ์ผกผัน $(\underset{\sim}{X}'\underset{\sim}{X})^{-1}$ ของเมตริกซ์ $\underset{\sim}{X}'\underset{\sim}{X}$ ได้ ซึ่งเป็นจริงโดยทั่วไปในทางปฏิบัติ เพราะฉะนั้นตัวประมาณแบบกำลังสองน้อยที่สุดสามัญคือ

$$\underset{\sim}{B} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{Y}$$

และตัวแบบพยากรณ์ค่า Y หรือตัวประมาณของค่าเฉลี่ย E(Y) ของ Y เมื่อกำหนดให้ $X_1 = X_{1o}, X_2 = X_{2o}, \ldots, X_k = X_{ko}$ คือ

$$\hat{Y} = \underset{\sim}{X}'_o \underset{\sim}{B}$$
$$= B_o + B_1 X_{1o} + B_2 X_{2o} + \ldots + B_k X_{ko}$$

ซึ่ง $\underset{\sim}{X}'_o = (X_{1o}, X_{2o}, \ldots, X_{ko})$

เนื่องจากในการอนุมานเชิงสถิติเกี่ยวกับพารามิเตอร์ ซึ่งเกี่ยวกับการทดสอบข้อสมมติฐานต่าง ๆ และการประมาณค่าแบบช่วงของค่าพารามิเตอร์ ตลอดจนการทดสอบและการประมาณค่าแบบช่วงของค่าเฉลี่ย E(Y) ของ Y และของค่าพยากรณ์ของค่าจริง Y โดยทั่วไปการทดสอบและการประมาณค่าดังกล่าว จะกระทำภายใต้คุณสมบัติพื้นฐานของค่าผิดพลาดสุ่ม ε_t ดังนั้นจึงจำเป็นต้องตรวจสอบความเหมาะสม หรือความเพียงพอของตัวแบบ ด้วยการตรวจสอบคุณสมบัติของค่าผิดพลาด ε_t แต่เนื่องจากไม่ทราบค่าจริง ε_t ฉะนั้นจะตรวจสอบคุณสมบัติของค่าเศษเหลือตกค้าง e_t ซึ่งกำหนด $e_t = y_t - \hat{y}_t$ เป็นค่าประมาณของ ε_t เมื่อตัวแบบเหมาะสมเพียงพอ และตรวจสอบว่าค่า $e_t (t = 1, 2, \ldots, n)$ มีคุณสมบัติสอดคล้องภายใต้คุณสมบัติพื้นฐานของค่าผิดพลาดสุ่ม ε_t หรือไม่ นั่นคือ มีค่าเฉลี่ยเท่ากับศูนย์ มีความแปรปรวนคงที่ ไม่มีอัตสหสัมพันธ์ และมีการแจกแจงแบบปกติ หรือไม่ นอกจากนี้ อาจพบว่าตัวแบบไม่เหมาะสม เนื่องมาจากตัวแบบมีรูปแบบยังไม่ถูกต้องเหมาะสม วิธีการตรวจสอบนักพยากรณ์อาจเลือกใช้วิธีกราฟ หรือวิธีการทดสอบในเชิงสถิติ ซึ่งวิธีเชิงสถิติเป็นวิธีที่มีระเบียบ (formal) หรือมีทฤษฎี อย่างไรก็ตาม ถึงแม้ว่าวิธีกราฟอาจจะไม่เป็นวิธีเชิงระเบียบทฤษฎี แต่โดยทั่วไปก็เป็นวิธีการที่เพียงพอที่จะใช้วินิจฉัยตัวแบบ และเป็นวิธีการง่ายที่ใช้โดยทั่วไป

ถ้ากราฟระหว่างค่าเศษเหลือตกค้าง e_t หรือค่าเศษเหลือตกค้างมาตรฐาน (standardized residuals) $e_t / \sqrt{MSE}$ กับตัวแปร $\hat{y}_t$, ตัวแปรอิสระ X_{it} แต่ละตัว และกับเวลา (ถ้าเป็นอนุกรมเวลา) ทั้งหมดมีรูปแบบการกระจายของจุดเป็นแนวขนาน แสดงว่าตัวแบบการถดถอยเชิงเส้นมีรูปแบบเหมาะสม ค่าผิดพลาดมีค่าเฉลี่ยศูนย์ และความแปรปรวนคงที่ นอกจากนี้ในกรณีที่เป็นกราฟ e_t กับเวลา แสดงด้วยว่า ความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระ ไม่แปรเปลี่ยนตาม

เวลา หรือไม่มีอิทธิพลของเวลา และค่าผิดพลาดสุ่มไม่มีอัตสหสัมพันธ์กัน แต่ถ้ากราฟมีรูปแบบไม่เป็นแนวขนาน แสดงว่าค่าผิดพลาดไม่สอดคล้องคุณสมบัติบางข้อหรือทุกข้อของค่าผิดพลาดสุ่ม ε_t หรือตัวแบบยังมีรูปแบบไม่ถูกต้อง ถ้ากรณีที่ความแปรปรวนของค่าผิดพลาดไม่คงที่ การแก้ปัญหาอาจใช้วิธีแปลงค่าของตัวแปรตาม Y หรือใช้วิธีแก้ปัญหาเฉพาะในเรื่องนี้ เช่น ใช้วิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก (Weighted Least Squares Method) ส่วนกรณีที่ตัวแบบการถดถอยเชิงเส้นยังมีรูปแบบไม่เหมาะสม ควรมีเทอมค่าคงที่ หรือองค์ประกอบเชิงเส้นในตัวแบบ หรือควรมีเทอมที่มีกำลังสูงขึ้นในตัวแบบ และถ้าข้อมูลเป็นอนุกรมเวลา ควรมีเทอมเชิงเส้น หรือเทอมอันดับหนึ่งหรืออันดับสองของเวลาในตัวแบบด้วย

การตรวจสอบว่าค่าผิดพลาดมีสหสัมพันธ์กันหรือไม่ โดยปกติจะตรวจสอบเมื่อข้อมูลเป็นอนุกรมเวลา วิธีการมีหลายวิธี เช่น ใช้วิธีพิจารณาค่าของฟังก์ชันอัตสหสัมพันธ์ตัวอย่าง [Sample Autocorrelation Function (SACF)] $|r_k|$ ของ e_t ที่คาบเวลาห่างกัน k $(k = 1, 2, \ldots)$ เปรียบเทียบกับค่าตัดสินเชิงสถิติหรือค่าวิกฤต (Critical value) ซึ่งมีค่าเท่ากับ $2/\sqrt{n}$ โดยประมาณ ที่ระดับนัยสำคัญ 0.05 เมื่อพบว่าค่าผิดพลาดมีสหสัมพันธ์กัน ในการวิเคราะห์การถดถอยเชิงเส้น มีวิธีแก้ปัญหานี้โดยเฉพาะ เช่น วิธีการแปลงตัวแปร และวิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก เป็นต้น

การตรวจสอบว่าค่าผิดพลาดมีการแจกแจงแบบปกติหรือไม่ อาจใช้วิธีเขียนแผนภาพฮิสโตแกรม และ/หรือกราฟความน่าจะเป็นแบบปกติ ถ้าแผนภาพฮิสโตแกรมมีลักษณะโค้งสมมาตร หรือเส้นกราฟความน่าจะเป็นเป็นแนวเส้นตรง ก็มีเหตุผลได้ว่าค่าผิดพลาดมีการแจกแจงเข้ารูปแบบปกติ แต่ถ้าฮิสโตแกรมเบ้มาก หรือเส้นกราฟความน่าจะเป็นไม่เป็นแนวเส้นตรง ค่าผิดพลาดจะไม่สอดคล้องกับคุณสมบัติการแจกแจงแบบปกติ ควรแก้ไขตัวแบบ เช่น ใช้วิธีการแปลงค่าตัวแปรตาม Y

นอกจากการตรวจสอบคุณสมบัติต่าง ๆ ของค่าผิดพลาดสุ่ม และความถูกต้องของตัวแบบการถดถอยเชิงเส้นแล้ว ในการวิเคราะห์การถดถอย จะมีการคำนวณค่าของตัวสถิติต่าง ๆ สำหรับทดสอบผลกระทบของแต่ละตัวแปรอิสระต่อค่า Y ว่ามีมากพอที่จะยอมรับได้หรือไม่ในเชิงสถิติ รวมทั้งการตรวจสอบตัวแปรอิสระพร้อมกันทุกตัว (ถ้ามีมากกว่าหนึ่งตัว) ว่าตัวแปรอิสระทั้งกลุ่มมีผลต่อค่า Y มากพอหรือไม่ หรือช่วยให้ค่าพยากรณ์ถูกต้องมากขึ้นหรือไม่ (จากเดิมที่ใช้ค่าเฉลี่ย Y เท่านั้นโดยไม่อาศัยตัวแปรอิสระใด ๆ) ถ้าพบว่าตัวแปรอิสระตัวใดมีผลกระทบน้อย จนไม่อาจยอมรับได้ในเชิงสถิติ ก็ควรตัดตัวแปรนั้นออกจากตัวแบบ การทดสอบเกี่ยวกับผลกระทบของตัวแปรอิสระทีละตัว เราใช้ตัวสถิติทดสอบที และการทดสอบผลกระทบรวมกันเป็นกลุ่ม เราใช้ตัวสถิติทดสอบเอฟ

2.3 การพยากรณ์ของการไฟฟ้าส่วนภูมิภาค

การพยากรณ์ความต้องการไฟฟ้าของ กฟภ. มีวิธีการพยากรณ์ปริมาณความต้องการไฟฟ้าในส่วนภูมิภาคตามรายภาคต่าง ๆ ได้แก่ ภาคเหนือ ภาคตะวันออกเฉียงเหนือ ภาคกลาง และภาคใต้ โดยแยกตามสาขาต่าง ๆ ดังนี้

2.3.1 การพยากรณ์ความต้องการไฟฟ้าสาขาบ้านอยู่อาศัย

หน่วยจำหน่ายสาขาบ้านอยู่อาศัย ประกอบด้วย ประเภทบ้านอยู่อาศัยขนาดเล็ก และบ้านอยู่อาศัยขนาดใหญ่ ซึ่งทำการพยากรณ์โดยวิธี End – Use Approach สร้างแบบจำลอง Home Model ซึ่งคำนึงถึงพฤติกรรมของครัวเรือนในการถือครองเครื่องใช้ไฟฟ้า และลักษณะการใช้เครื่องใช้ไฟฟ้า นำไปสู่การบริโภคการใช้ไฟฟ้า เพื่อกำหนดพฤติกรรมและลักษณะการใช้ไฟฟ้าให้มีค่า และความสัมพันธ์ตามกลุ่ม จึงมีการแบ่งกลุ่มครัวเรือนที่ทำการศึกษา ตามระดับรายได้ ชนิดที่อยู่อาศัย และภูมิศาสตร์ โดยข้อมูลดังกล่าวได้มาจากการสำรวจ การประมาณการอัตราการถือครอง ชั่วโมงการใช้ ขนาดของเครื่องใช้ไฟฟ้า และประสิทธิภาพของเครื่องใช้ไฟฟ้า เป็นตัวกำหนดการใช้ไฟฟ้าในสาขาบ้านอยู่อาศัยในอนาคต ซึ่งมีรูปแบบจำลองดังนี้คือ

$$\sum E_{ijkt} = \sum (S_{ijkt} * ESR_{ijkt} * U_{ijkt})$$

โดยที่ E คือ ปริมาณการใช้ไฟฟ้า
S คือ อัตราการถือครองของเครื่องใช้ไฟฟ้า
ESR คือ ขนาดของเครื่องใช้ไฟฟ้า
U คือ ชั่วโมงการใช้ของเครื่องใช้ไฟฟ้า
i คือ ชนิดของเครื่องใช้ไฟฟ้า
j คือ ระดับรายได้
k คือ ชนิดของที่อยู่อาศัย
t คือ ปีที่ทำการพยากรณ์

แบบจำลองดังกล่าว แบ่งตามลักษณะทางภูมิศาสตร์ คือ ภาคเหนือ ภาคตะวันออกเฉียงเหนือ ภาคกลาง และภาคใต้ โดยมีรายละเอียดของการจัดทำการพยากรณ์ตามแผนภูมิ แผนภาพที่ 2.1 เริ่มจากการพยากรณ์จำนวนที่อยู่อาศัย และแบ่งกลุ่มตามระดับรายได้ โดยในแต่ละระดับ

รายได้แบ่งตามชนิดที่อยู่อาศัย นำผลจากการสำรวจเครื่องใช้ไฟฟ้า อัตราการถือครองเครื่องใช้ไฟฟ้า ขนาดของเครื่องใช้ไฟฟ้า ชั่วโมงการใช้ แบ่งตามกลุ่มรายได้และชนิดที่อยู่อาศัย และปัจจัยด้านประสิทธิภาพของเครื่องใช้ไฟฟ้า มาคำนวณการใช้ไฟฟ้าในสาขาบ้านอยู่อาศัยตามแบบจำลองข้างต้น

2.3.2 การพยากรณ์ความต้องการไฟฟ้าสาขาธุรกิจและอุตสาหกรรม

หน่วยจำหน่ายสาขาธุรกิจและอุตสาหกรรม ประกอบด้วย ประเภทกิจการขนาดเล็ก กิจการขนาดกลาง และกิจการขนาดใหญ่ จัดกลุ่มข้อมูลผู้ใช้ไฟฟ้าตามสาขาเศรษฐกิจ ทำการพยากรณ์หน่วยจำหน่ายตามสาขาเศรษฐกิจ ดังแผนภูมิแสดงวิธีการพยากรณ์ สาขาธุรกิจและอุตสาหกรรม แผนภาพที่ 2.2

โดยจะใช้วิธีการพยากรณ์หน่วยจำหน่ายต่อผลิตภัณฑ์มวลรวมประชาชาติ ของแต่ละสาขาเศรษฐกิจ หรือกล่าวโดยย่อ ๆ ว่าพยากรณ์อัตราส่วนการใช้ไฟฟ้าต่อมูลค่าเพิ่ม (Energy Intensity Ratio : EIR) ในแต่ละสาขาเศรษฐกิจ โดยคำนวณค่า EIR ในแต่ละสาขาเศรษฐกิจ ซึ่งจะพิจารณาข้อมูล EIR รวมของกิจการขนาดเล็ก กิจการขนาดกลาง และกิจการขนาดใหญ่แยกเป็นรายภาค

2.3.3 การพยากรณ์ความต้องการไฟฟ้าประเภทอื่น ๆ

หน่วยจำหน่ายประเภทอื่น ๆ ประกอบด้วย กิจการเฉพาะอย่าง ราชการและองค์กรที่ไม่แสวงหากำไร สูบน้ำเพื่อการเกษตร และไฟชั่วคราว ทำการพยากรณ์โดยกำหนดให้การใช้ไฟฟ้ามีความสัมพันธ์กับผลิตภัณฑ์มวลรวมประชาชาติ

1. กิจการเฉพาะอย่าง

ใช้วิธีการพยากรณ์อัตราส่วนการใช้ไฟฟ้าต่อมูลค่าเพิ่ม (EIR) ในสาขาของการโรงแรมและภัตตาคาร

2. ราชการและองค์กรที่ไม่แสวงหากำไร

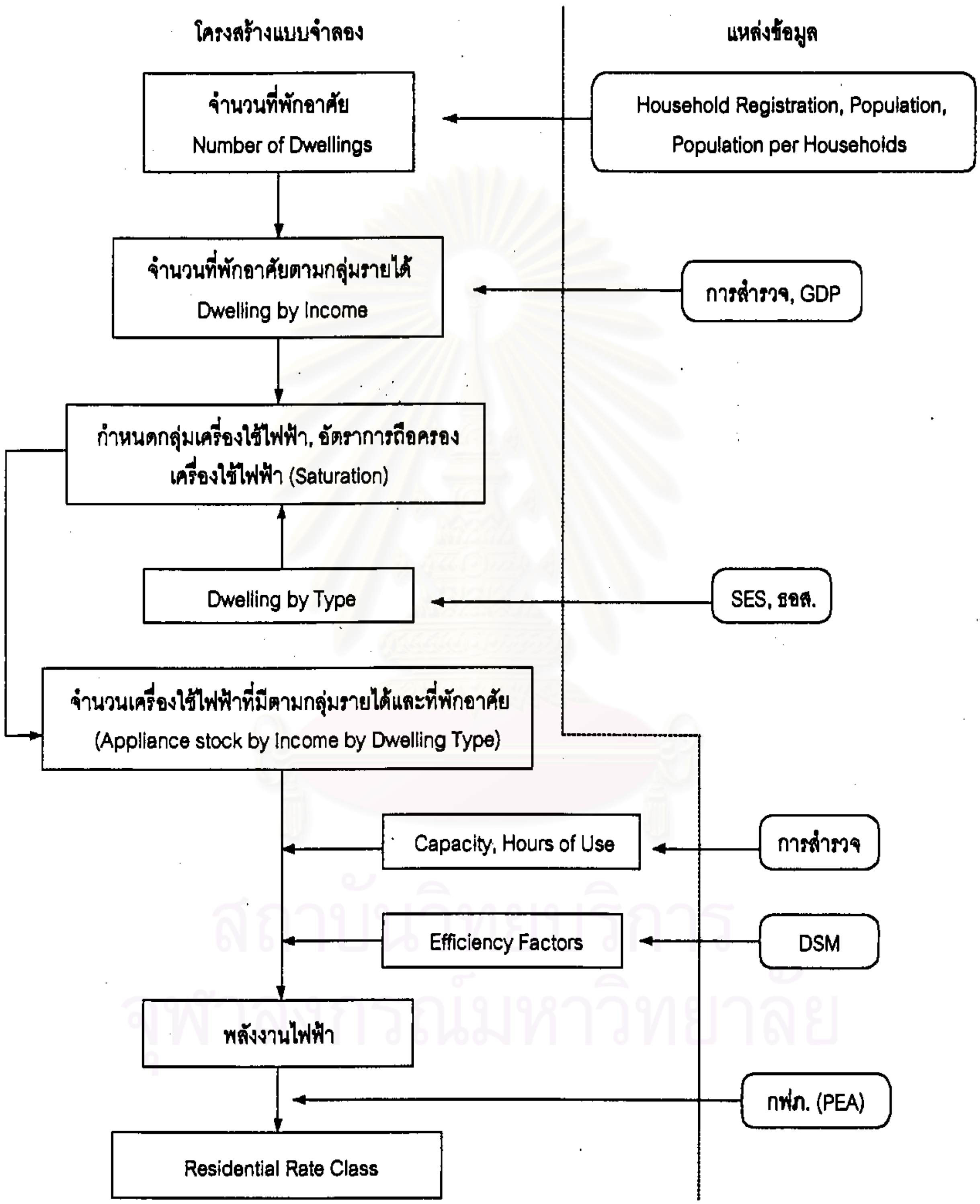
ใช้วิธีการพยากรณ์อัตราส่วนการใช้ไฟฟ้าต่อมูลค่าเพิ่ม (EIR) ในสาขาของการบริหารและการป้องกันประเทศ

3. สูบน้ำเพื่อการเกษตร และไฟชั่วคราว

ใช้วิธีการพยากรณ์อัตราส่วนการใช้ไฟฟ้าต่อมูลค่าเพิ่ม (EIR) ในสาขาของเกษตรกรรม

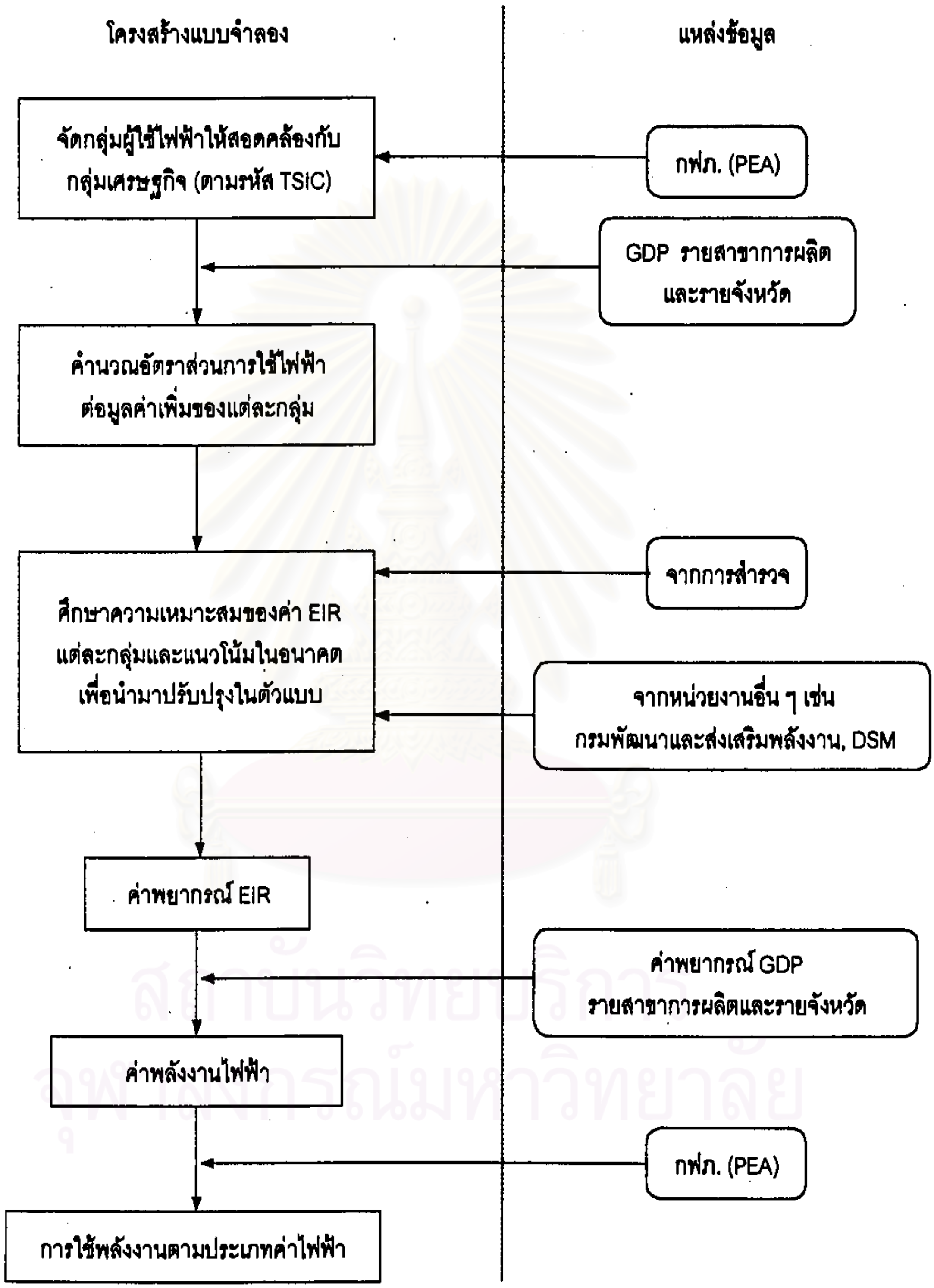
ที่มา : การพยากรณ์ความต้องการไฟฟ้า ปี 2536, แผนกสถิติการใช้ไฟฟ้า กองเศรษฐกิจพลังไฟฟ้า การไฟฟ้าส่วนภูมิภาค

แผนภาพที่ 2.1 แผนภูมิแสดงการจัดทำการพยากรณ์สาขาบ้านอยู่อาศัย



ที่มา : การพยากรณ์ความต้องการไฟฟ้า ปี 2536, แผนกสถิติการใช้ไฟฟ้า กองเศรษฐกิจพลังไฟฟ้า
การไฟฟ้าส่วนภูมิภาค

แผนภาพที่ 2.2 แผนภูมิแสดงการจัดทำการพยากรณ์สาขาธุรกิจและอุตสาหกรรม



ที่มา : การพยากรณ์ความต้องการไฟฟ้า ปี 2536, แผนกสถิติการใช้ไฟฟ้า กองเศรษฐกิจพลังไฟฟ้า การไฟฟ้าส่วนภูมิภาค