บทที่ 5

การสร้างและวิเคราะห์รูปแบบทางคณิตศาสตร์

5.1 รูปแบบทางคณิตศาสตร์ที่จะสร้าง

เนื่องจากการใช้ไฟฟ้าของผู้ใช้ไฟฟ้าแต่ละประเภทมีลักษณะแตกต่างกัน ปัจจัยที่มีผลต่อการใช้ไฟฟ้าของแต่ละประเภทก็แตกต่างกัน ดังนั้น ในการหารูปแบบทางคณิตศาสตร์ที่จะนำมาอธิบายความสัมพันธ์ระหว่างความต้องการใช้พลังงานไฟฟ้ากับปัจจัยต่าง ๆ ที่เกี่ยวข้องนั้นก็ควรจะแยกหาเป็นของแต่ละประเภท แล้วจึงรวมของทุกประเภทเข้าด้วยกัน ทั้งนี้เพื่อทำให้สามารถหารูปแบบที่แสดงความสัมพันธ์ใกล้เคียงกับสภาพความเป็นจริงมากขึ้น และเป็นประโยชน์ต่อการวางแผนขยายและปรับปรุงระบบการจ่ายพลังไฟฟ้า

ในการหารูปแบบทางคณิตศาสตร์สำหรับความต้องการใช้พลังงานไฟฟ้าในเขตนครหลวงของบทวิจัยนี้ จะแยกวิเคราะห์เป็น 4 ประเภทรวมของผู้ใช้ไฟฟ้าดังที่ได้กล่าวมาแล้ว จากนั้นจึงรวมผลที่ได้จากแต่ละประเภทเข้าด้วยกันเป็นความต้องการใช้พลังงานไฟฟ้าทั้งหมดของเขตนครหลวง กล่าวคือ

$$E_i = L_i + C_i + I_i + S_i$$

เมื่อ E_i = จำนวนพลังงานไฟฟ้าที่ใช้ทั้งหมดของเขตนครหลวงในปีที่ i

L_i = จำนวนพลังงานไฟฟ้าที่ใช้ในประเภทบ้านอยู่อาศัยของเขตนครหลวงในปีที่ i

C_i = จำนวนพลังงานไฟฟ้าที่ใช้ในประเภทธุรกิจการค้าของเขตนครหลวงในปีที่ i

I_i = จำนวนพลังงานไฟฟ้าที่ใช้ในประเภทอุตสาหกรรมของเขตนครหลวงในปีที่ i

S_i = จำนวนพลังงานไฟฟ้าที่ใช้ในประเภทไฟถนนสาธารณะของเขตนครหลวงในปีที่ i

(E = Electricity, L = Residential, C = Commercial, I = Industrial
S = Street lighting)

รูปแบบทางคณิตศาสตร์ที่จะนำมาใช้นั้นมีได้หลายรูป แต่สำหรับการวิจัยนี้จะเลือกใช้รูปแบบที่ไม่ยุ่งยากเกินไปและสามารถอธิบายความสัมพันธ์ระหว่างความต้องการใช้พลังงานไฟฟ้าและปัจจัยที่เกี่ยวข้องได้ดี 2 แบบคือ[1/]

ก. Standard Linear Model ซึ่งมีรูปเป็น

$$Y = B_0 + B_1X_1 + B_2X_2 + \cdots\cdots + B_pX_p + \epsilon$$

โดยที่ Y = ตัวแปรตาม (dependent variable)

X_i = ตัวแปรอิสระ (independent variables), $i = 1, 2, \ldots\ldots, p$

B_0 = พารามิเตอร์ ที่จะต้องประมาณค่าจากข้อมูล เป็นค่าคงที่ (Constant)

B_i = พารามิเตอร์ ซึ่งเป็น Regression Coefficient ของตัวแปรอิสระ X_i และเป็นค่าคงที่

ϵ = ค่าความคลาดเคลื่อน (error term)

เหตุที่เลือกใช้รูปสมการนี้เพราะปัจจัยที่มีผลต่อการใช้ไฟฟ้าแต่ละประเภทมีหลายอย่าง (มีตัวแปรอิสระหลายตัวให้เลือก) และ scatter diagram ที่แสดงความสัมพันธ์ระหว่างพลังงานไฟฟ้าที่ใช้และปัจจัยแต่ละอย่างก็มีแนวโน้มในลักษณะใกล้เคียงกับเส้นตรง

ข. Multiplicative Model มีรูปเป็น (เพื่อให้ดูง่าย สมมติว่ามีตัวแปรอิสระ 3 ตัว)

$$Y = A X_1^{B} X_2^{C} X_3^{D} \cdot \epsilon$$

โดยที่ A, B, C, D เป็น พารามิเตอร์

เป็น Nonlinear Model ชนิด Intrinsically Linear หรือแปลงรูปให้เป็น Linear Model ได้ โดยใช้ logarithm เข้าช่วย จะได้สมการ

$$\ln Y = \ln A + B \ln X_1 + C \ln X_2 + D \ln X_3 + \ln \epsilon$$

ซึ่งอยู่ในรูปลักษณะเดียวกับ Standard Linear Model และเรียกว่า Double log form

---

1/ Norman Draper and Harry Smith, Applied Regression Analysis (New York : John Wiley & Son, Inc.,1966), p. 128.

เหตุที่เลือกรูปแบบนี้เพราะว่าปัจจัยทางเศรษฐกิจส่วนใหญ่ซึ่งนำมาเป็นตัวแปรอิสระในการสร้างรูปแบบนี้จะมีการเปลี่ยนแปลงตามเวลาแบบ Exponential ดังได้กล่าวมาแล้วในบทที่ 4 และความต้องการทางด้านไฟฟ้าก็มีอัตราการเพิ่มที่เร็วเช่นกัน

5.2 วิธีการทางสถิติที่ใช้เลือกตัวแปรอิสระใน Regression

หมายถึง วิธีที่ใช้เลือกตัวแปรอิสระที่มีนัยสำคัญทางสถิติ (Statistical Significance) เพื่อสร้าง Regression กับตัวแปรตามได้ดีที่สุด หรือคือการตัดสินว่าจะมีตัวแปรอิสระใดที่สำคัญพอที่จะอยู่ในรูปแบบบ้างเพื่อให้ได้รูปแบบที่ดีที่สุด

วิธีที่ใช้กันมีอยู่หลายวิธี[2/] และผลที่ได้ก็อาจจะไม่ตรงกัน เช่น

วิธี All Possible Regressions คือการสร้างสมการ regression ทุกสมการที่จะมีได้สำหรับตัวแปรอิสระทั้งหมดที่มีให้เลือก ตั้งแต่มีตัวแปรอิสระ 1 ตัว จนถึงมีครบทุกตัวแล้วเลือกสมการที่ให้ค่าสัมประสิทธิ์สหสัมพันธ์เชิงซ้อน (Multiple Correlation Coefficient) หรือ R มากที่สุด (ปกติดูจากค่าร้อยละของ R^2)

$$R^2 = \frac{\text{explanined variation}}{\text{Total variation}} = \frac{\sum_i (\hat{Y}_i - \bar{Y})^2}{\sum_i (Y_i - \bar{Y})^2}$$

เมื่อ Y เป็นตัวแปรตาม

วิธี Backward Elimination เป็นวิธีที่สร้างสมการโดยให้มีตัวแปรอิสระทุกตัวที่จะเลือกอยู่ในสมการทั้งหมดก่อน แล้วจึงตัดตัวที่ไม่สำคัญออกทีละตัว

วิธี Forward Selection เป็นวิธีสร้างสมการโดยการเลือกตัวแปรอิสระเข้าไปในสมการทีละตัว จนกว่าตัวที่จะเข้านั้นไม่มีนัยสำคัญทางสถิติ ก็จะหยุด

วิธี Stepwise Regression วิธีนี้เป็นวิธีที่ปรับปรุงจากวิธี Forward selection คือมีการตัดตัวแปรอิสระที่เข้าอยู่ในสมการแล้วออก ถ้าหากการเพิ่มตัวแปรอิสระอีกตัวหนึ่งเข้าไปในสมการ ทำให้ตัวแปรอิสระที่เข้าก่อนนั้นไม่สำคัญ ทั้งนี้เนื่องจาก

2/ Norman Draper and Harry Smith, *ibid.*, p. 163.

ความสัมพันธ์ของตัวแปรอิสระทั้งสองมีมาก (high intercorrelation หรือ collinearity among independent variables)

รายละเอียดของวิธี Stepwise Regression นี้คือ :-

สมมติให้ $X_1, X_2 \ldots\ldots, X_k$ เป็นตัวแปรอิสระทั้งหมดที่มีให้เลือก เพื่อสร้างสมการ regression กับ Y ซึ่งเป็นตัวแปรตาม

วิธีการเลือกตัวแปรอิสระจะดำเนินการเป็นขั้นตอนดังนี้คือ

<u>ขั้นที่ 1</u> คำนวณหา simple correlation matrix ซึ่งเป็น matrix ที่แสดงค่าสัมประสิทธิ์สหสัมพันธ์อย่างง่าย หรือ r ระหว่างตัวแปรทีละคู่ทุกคู่ (ทั้งตัวแปรอิสระและตัวแปรตาม) ตัวแปรอิสระใดมีค่า r กับตัวแปรตามสูงสุดจะได้รับเลือกเข้าสมการ สมมติว่าในที่นี้คือ X_1 ดังนั้น จะได้สมการ $\hat{Y} = f\ (X_1)$ ทดสอบดูว่าตัวแปรอิสระที่เพิ่งจะเข้าสมการมีส่วนในการอธิบายค่า Y อย่างมีนัยสำคัญทางสถิติหรือไม่ โดยดูที่ค่า sequential F เทียบกับค่า F จากตารางในระดับนัยสำคัญที่เลือก [ในที่นี้คือเทียบกับ $F\ (1, n-2, 1-\alpha)$ ในตาราง โดยที่ n เป็นจำนวนข้อมูล (observations), α = ระดับนัยสำคัญที่เลือก] ถ้า Sequential F > F จากตารางแสดงว่า X_1 ยังคงอยู่ในสมการต่อไปได้ ถ้าน้อยกว่าก็แสดงว่า Y ไม่สัมพันธ์ตัวแปรอิสระตัวใดอย่างมีนัยสำคัญเลย

Sequential F เป็นอีกชื่อหนึ่งของ partial F ที่วัด variation ของตัวแปรตามที่อธิบายโดยตัวแปรอิสระตัวสุดท้ายในขั้นที่กำลังเลือก (ในที่นี้คือ X_1) โดยถือว่าตัวแปรตัวอื่นอยู่ในสมการแล้ว (ในขั้นนี้ยังไม่มี) ส่วนความหมาย partial F จะได้กล่าวต่อไป

ตามหลัก Regression ที่ถูกต้องนั้น การทดสอบ sequential F นี้จะหมายถึงการทดสอบ sequential F ของ Regression coefficient ของตัวแปรอิสระที่กำลังพิจารณา สมมุติว่าตัวแปรอิสระนั้นคือ X_i จะหมายถึงการทดสอบสมมติฐาน

$$H_0 : \quad B_i = 0$$

แต่ในที่นี้เพื่อให้ง่ายแก่การเข้าใจจะเรียกว่า "ทดสอบ Sequential F ของ X_i " ในกรณีของ partial F ก็เช่นกัน จะเรียกว่า "ทดสอบ partial F ของ X_i "

ขั้นที่ 2 หาค่า partial correlation coefficient ระหว่างค่า Y กับตัวแปรอิสระอื่น ๆ ที่ยังไม่ได้นำเข้าสมการ คือ X_j ($j \neq 1$)โดยถือว่า X_1 อยู่ในสมการเรียบร้อยแล้ว หรือกล่าวอีกนัยหนึ่งก็คือการหาสัมประสิทธิ์สหสัมพันธ์ระหว่าง

residual จากสมการ regression $\hat{Y} = f(X_1)$

กับ residual จากสมการ regression $\hat{X}_j = f_j(X_1)$

(residual คือผลต่างระหว่างค่าจริงและค่าประมาณจากสมการ regression เช่น $Y - \hat{Y}$) ตัวแปรอิสระที่ให้ค่า partial correlation กับ Y มากที่สุดจะได้รับเลือกเข้าสมการต่อไป สมมติในที่นี้คือ X_2 ดังนั้นสมการที่ได้จะเป็น $\hat{Y} = f(X_1, X_2)$

จากนั้น ทดสอบ sequential F ของ X_2 ดู ถ้า X_2 ไม่สำคัญพอก็จะตัด X_2 ออกเหลือเพียง $\hat{Y} = f(X_1)$ และหยุดเพียงเท่านี้ ถ้า X_2 ยังคงสำคัญพอก็จะมีตัวแปรอิสระ 2 ตัวในสมการตามเดิม และดำเนินการตามขั้นต่อไป

ขั้นที่ 3 ทดสอบ Partial F ของตัวแปรอิสระที่เลือกเข้าสมการก่อนหน้านี้ ในที่นี้คือทดสอบดูว่าหลังจากที่ X_2 เข้าไปอยู่ในสมการแล้ว X_1 ซึ่งได้รับเลือกให้เข้าสมการก่อนยังมีความสำคัญต่อไปอีกหรือไม่ โดยหาค่า partial F ของ X_1 เทียบกับค่า F จากตาราง [ในขั้นนี้คือ $F(1, n-3, 1-\alpha)$] ถ้า partial F ที่ได้ < F จากตาราง ก็จะตัด X_1 ออกไปจากสมการ ซึ่งจะได้ $\hat{Y} = f(X_2)$แต่ถ้า partial F ที่ได้ > F จากตารางก็จะทำขั้นต่อไป

การทดสอบค่า partial F ในที่นี้ก็คือการทดสอบดูว่าตัวแปรอิสระที่เข้าไปอยู่ในสมการแล้ว ถ้าเราให้เข้าเป็นตัวสุดท้าย โดยให้ตัวแปรอิสระตัวอื่นที่เลือกแล้วอยู่ในสมการก่อน จะสามารถอธิบายค่า Y ได้อีกมากน้อยเพียงใด กล่าวคือ หาค่า

F (1,n-p-1)

$$= \frac{\text{additional variance explained by the independent variables}}{\text{Unexplained variance}}$$ 3/

n = จำนวนข้อมูลที่หามาได้

p = จำนวนตัวแปรอิสระที่อยู่ในสมการขณะนี้

ขั้นที่ 4 ต่อไปก็คือการหา partial correlation coefficient ของตัวแปรอิสระที่เหลือเช่นเดียวกับวิธีการตั้งแต่ขั้นที่ 2 เรื่อยไป จนกระทั่งไม่มีตัวแปรอิสระใดที่เป็นที่ยอมรับให้เข้าสมการได้ (ไม่ผ่านการทดสอบ Sequential F test) ก็จะได้สมการที่ต้องการ

วิธี Stepwise Regression นี้ Draper และ Smith แนะนำว่า "เป็นวิธีเลือกตัวแปรเข้าสมการ Regression ที่ดีที่สุด อย่างไรก็ตาม การพิจารณาเลือกตัวแปรอิสระในขั้นต้นก่อนที่จะนำมาใช้กับวิธีนี้ก็เป็นสิ่งสำคัญ ไม่ควรจะไว้วางใจวิธีนี้ ซึ่งปกติจะทำโดยเครื่องคอมพิวเตอร์ให้มากเกินไป"

3/ Thomas H. Wonnacott, Ronald J. Wonnacott, Introductory Statistias for Businness and Economics (New York : John Wiley & Son, Inc., 1972), p. 351.

(การสร้างสมการ Regression ของวิธีทั้ง 4 ข้างต้น อาศัยการประมาณค่าของพารามิเตอร์ ด้วยวิธี กำลังสองน้อยที่สุด)

นอกจากวิธีทั้ง 4 นี้แล้ว ยังมีวิธีอื่นที่นิยมใช้กัน เช่น วิธี Stagewise Regression และการนำวิธีในสี่วิธีนี้มาใช้ผสมกัน เป็นต้น

ในการวิจัยนี้ได้เขียนโปรแกรมคอมพิวเตอร์ (Computer Program) ของวิธี Stepwise Regression นี้ตามวิธีการของ "Efroymson" ซึ่ง Draper และ Smith ได้นำรายละเอียดการคำนวณมาแสดงในหนังสือ "Applied Regression Analysis" โดยได้ดัดแปลงเพื่อให้ใช้ในการหารูปแบบของความต้องการใช้พลังงานไฟฟ้าทั้งในรูป Standard Linear และ Double log ซึ่งรายละเอียดของโปรแกรมและการทดสอบโปรแกรมได้แสดงไว้ในภาคผนวก จ.

มีข้อที่น่าสังเกตคือ การคำนวณค่าต่าง ๆ รวมทั้งค่า Regression coefficient ของสมการของวิธีที่ใช้ในโปรแกรมนี้ได้อาศัย correlation coefficient matrix ของตัวแปรอิสระที่เข้าสมการแล้ว กล่าวคือ เป็น matrix ที่แสดงค่า r ของตัวแปรอิสระ X_i กับ X_j ดังนี้

$$\begin{matrix} 1 & r_{12} & r_{13} & \cdots & r_{1p} \\ r_{21} & 1 & r_{23} & \cdots & r_{2p} \\ r_{31} & r_{32} & 1 & \cdots & r_{3p} \\ \vdots & & & & \vdots \\ r_{p1} & r_{p2} & r_{p3} & \cdots & 1 \end{matrix}$$

(เมื่อ p = จำนวนตัวแปรอิสระที่เลือกเข้าสมการ)

โดยเฉพาะการหาค่า Regression coefficient แต่ละค่าจะต้องหารด้วยค่า determinant ของ matrix ดังกล่าว ดังนั้นถ้าค่า determinant น้อยมาก (มีค่าใกล้ 0) ค่าของ Regression coefficient จะมากอย่างหาค่าไม่ได้ (Solutions approaches indeterminancy) การที่ค่า determinant มีค่าน้อยเนื่องจากค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรอิสระมีค่ามาก หรือหมายความว่า ตัวแปรอิสระมี

ความสัมพันธ์หรือขึ้นแก่กันมาก[4/] ดังนั้นในกรณีที่รูปแบบที่หาได้มีค่า determinant ของ correlation coefficient matrix น้อยก็ไม่ควรจะนำรูปแบบดังกล่าวมาใช้

5.3 หลักเกณฑ์ในการเลือกรูปแบบ

การหารูปแบบทางคณิตศาสตร์สำหรับความต้องการใช้พลังงานไฟฟ้าของแต่ละประเภทในการวิจัยนี้จะใช้วิธีเลือกตัวแปรอิสระต่าง ๆ จากปัจจัยที่มีผลต่อการใช้ไฟฟ้าของแต่ละประเภท ดังที่กล่าวมาแล้วในบทที่ 4 ให้เข้าสมการด้วยวิธี Stepwise Regression โดยที่จะทำแบบ Standard linear และ Double log แล้วจึงเปรียบเทียบผลที่ได้ของรูปแบบทั้งสองด้วยค่า R^2 และค่า residual mean squares หรือกำลังสองของค่าประมาณของ Standard error of estimate (s^2) โดยจะถือค่า s^2 เป็นสำคัญ (ในการเปรียบเทียบจะต้องหาค่า s^2 ของรูปแบบชนิด double log เป็นค่าในสเกลเลขคณิตธรรมดา) ถ้ารูปแบบใดให้ค่า R^2 สูงกว่า และค่า s^2 น้อยกว่า ก็จะเลือกรูปแบบนั้น ในขณะเดียวกันจะดูค่า determinant ของ correlation coefficient matrix ถ้ามีค่าน้อยมากก็ไม่เลือก

ถึงแม้ว่าหลักเกณฑ์การเลือกรูปแบบข้างต้นจะเป็นการเพียงพอในแง่ของทฤษฎีและวิธีการของ Stepwise Regression คือทำการเลือกจากตัวแปรทั้งหมดที่พิจารณาโดยวิธี Stepwise ในรูปของ Standard linear สมการหนึ่งและในรูปของ double log อีกสมการหนึ่ง แล้วนำมาเปรียบเทียบกันก็จะเพียงพอ แต่ในทางปฏิบัติในระยะแรกที่นำวิธีการดังกล่าวมาใช้พบว่า ปัจจัยต่าง ๆ ที่มีผลต่อการใช้ไฟฟ้าหรือเป็นตัวแปรอิสระที่จะให้เลือกนั้นแต่ละตัวมีความสัมพันธ์กับการใช้ไฟฟ้ามาก (สัมประสิทธิ์สหสัมพันธ์ของ Y กับ X มีค่ามาก) และในขณะเดียวกัน ความสัมพันธ์ระหว่างตัวแปรอิสระด้วยกันก็มีมากเช่นกัน (สัมประสิทธิ์สหสัมพันธ์ระหว่าง X_i กับ X_j , $i \neq j$ มีค่ามาก) ถึงแม้จะใช้วิธีของ Stepwise ก็ตามก็ยังไม่สามารถจะตัดตัวแปรอิสระที่มีความสัมพันธ์กันมากออกไป และจะเข้าไปอยู่ในสมการพร้อม ๆ กัน และตัวแปรอิสระที่เลือกเข้าสมการซึ่งมีความสัมพันธ์กันมากนี้มักจะมีเครื่องหมายของ regression coefficient ตรงกันข้ามกับที่ควรจะเป็น เช่น ตามหลักเศรษฐศาสตร์ ถ้ารูปแบบของจำนวนพลังงานไฟฟ้าที่ใช้นี้เป็นฟังชั่นของอุปสงค์ (Demand function) จริง เมื่อจำนวนประชากรเพิ่มขึ้น การใช้ไฟฟ้าควรจะเพิ่มขึ้น คือ regression coefficient ของตัวแปรอิสระ "ประชากร" ควรจะมีเครื่องหมายเป็น " + "หรือ

4/ Norman Draper and Harry Smith, ibid., p. 148.

อย่างน้อยก็มีค่าเท่ากับ 0 แต่พบว่าในบางกรณีเมื่อนำตัวแปรอิสระที่มีความหมายใกล้เคียงกับจำนวนประชากร เช่น จำนวนผู้ใช้ไฟ มาพิจารณา และถูกเลือกเข้าไปในสมการโดยมีจำนวนประชากรได้รับเลือกเข้าไปด้วย ปรากฏว่าเครื่องหมายของ regression coefficient ของจำนวนประชากรจะเป็น "-" ซึ่งผิดกับหลักความจริง ไม่ควรจะนำรูปแบบที่ได้มาใช้

ดังนั้น ในการวิจัยนี้จึงต้องใช้วิธีเลือกรูปแบบโดยวิธี Stepwise regression กับวิธี all possible regression ผสมกัน คือก่อนที่จะนำตัวแปรอิสระต่าง ๆ ที่มีอยู่ไปเลือกเข้าสมการด้วยวิธี Stepwise นั้น จะพิจารณาเลือกตัวแปรอิสระที่มีความหมายใกล้เคียงกันมากหรือมีค่าสัมประสิทธิ์สหสัมพันธ์สูงมากให้เข้าโปรแกรม Stepwise พร้อมกับตัวแปรอิสระอื่น ๆ ชุดละตัวเป็นชุด ๆ ไป (เพื่อไม่ให้ตัวแปรอิสระที่มีความสัมพันธ์กันมากเข้าไปอยู่ในสมการพร้อมกันและทำให้เครื่องหมายของสัมประสิทธิ์ที่ได้ผิดความจริงไป) ซึ่งผลที่ออกมาจะได้สมการในรูปของ Linear และ log หลาย ๆ ชุด

จากนั้น จะใช้หลักเกณฑ์ในการพิจารณาดังนี้คือ เลือกสมการที่ดีและเหมาะสมที่สุดจากชุดของสมการทางด้าน linear และ log แล้วนำมาเปรียบเทียบหาสมการที่ดีที่สุด กล่าวคือ ในรูปแบบแต่ละชนิดคือ linear หรือ log นั้น จะมีสมการอยู่หลายสมการ ซึ่งได้จากการทำ Stepwise regression หลายครั้ง สมการที่ให้ค่า R^2 สูงที่สุด และ s^2 ต่ำสุด ก็จะรับเลือกให้เป็นสมการที่ดีที่สุดของแต่ละชนิดของรูปแบบ แต่ในขณะที่เลือกนั้นจะพิจารณาเครื่องหมายของ regression coefficient และค่าของ determinant ด้วย ถึงแม้สมการใดจะมีค่า R^2 สูง และ s^2 ต่ำ แต่ว่าเครื่องหมายที่ได้ขัดกับความจริง หรือค่า determinant น้อยก็จะไม่เลือกเช่นกัน จะพิจารณาเลือกสมการอื่นที่ให้ค่า R^2 และ s^2 ที่ดีรองลงมา เมื่อได้สมการที่เหมาะสมที่สุดจากด้าน linear และ log แล้วก็จะนำมาเปรียบเทียบค่า s^2 สมการที่ให้ค่า s^2 ที่น้อยกว่า จะเป็นสมการที่เลือกใช้สำหรับเป็นรูปแบบทางคณิตศาสตร์ของประเภทนั้น

ในการตัดสินว่า สมการใดดีที่สุดจะถือเอา s^2 หรือ Residual MS. เป็นสำคัญ เพราะว่าการเพิ่มตัวแปรอิสระเข้าไปในสมการมากขึ้นจะทำให้ค่า R^2 เพิ่มขึ้น แต่ในขณะเดียวกันก็ทำให้ degree of freedom ของ Residual น้อยลง ซึ่งในบางกรณี อาจจะทำให้ s^2 มากขึ้นก็ได้ 5/

5/ Norman Draper and Harry Smith, ibid., p.118.

สำหรับรายละเอียดของการหารูปแบบสำหรับความต้องการใช้พลังงานไฟฟ้าในเขตนครหลวงของผู้ใช้ไฟฟ้าแต่ละประเภทจะได้นำมากล่าวต่อไป

5.4 ข้อสมมุติในการหารูปแบบ

ในการสร้างรูปแบบทางคณิตศาสตร์ในรูปของ Standard Linear

$$Y_i = B_0 + B_1 X_{1i} + B_2 X_{2i} + \ldots\ldots + B_p X_{pi} + \epsilon_i$$

$$i = 1, 2, \ldots\ldots\ldots\ldots\ldots, n$$

ตามที่กล่าวข้างต้น ซึ่งประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุด และทดสอบสมมุติฐานด้วยค่าสถิติ (Statistic) ต่าง ๆ เช่น Partial F เป็นต้น หรือการหาขอบเขตความเชื่อมั่น (Confidence Limits) ก็ตาม ได้อาศัยข้อสมมุติดังต่อไปนี้

1. ϵ_i เป็นตัวแปรเชิงสุ่มที่เป็นอิสระแก่กัน (Independent random variable) ซึ่งจะได้ $cov\ (\epsilon_i \epsilon_j) = 0$, $i \neq j$
2. $E\ (\epsilon_i) = 0$ หรือมีค่ามัชฌิมเลขคณิต $= 0$
3. $V\ (\epsilon_i) = \sigma^2$ หรือมีค่าแปรปรวน (variance) $= \sigma^2$ คงที่
4. ϵ_i มีการแจกแจงแบบปกติ (Normally Distributed)

ด้วยข้อสมมุติดังกล่าว ทำให้ค่าที่ประมาณด้วยวิธีกำลังสองน้อยที่สุด เป็นค่าประมาณที่ดีที่สุด คือเป็น Best Linear Unbiased Estimators[6]

6/ Alexander M. Mood and Franklin A. Graybill. Introduction to the theory of Statistics (international student ed., 2 nd. ed; Tokyo : Kōgakusha co., Ltd., 1963) , p. 340.

ส่วนในการสร้างรูปแบบของ Double log ก็จะต้องมีข้อสมมุติสำหรับ ln ϵ_i ในลักษณะเดียวกับข้อสมมุติของรูปแบบชนิด Standard Linear เช่นกัน เพียงแต่แทนที่ ϵ_i ด้วย ln ϵ_i เท่านั้น

5.5 รูปแบบทางคณิตศาสตร์สำหรับประเภทบ้านอยู่อาศัย

5.5.1 การหารูปแบบ

ปัจจัยที่คาดว่าจะมีผลต่อการใช้ไฟฟ้าหรือตัวแปรอิสระที่จะนำมาพิจารณาในประเภทนี้ ตามการวิจัยในบทที่ 4 คือ รายได้สุทธิส่วนบุคคล มูลค่าผลิตภัณฑ์ของจังหวัด และราคาไฟฟ้า (ตัวแปรทั้งสามอยู่ในราคาคงที่ปี 2505) นอกจากตัวแปรทั้งสามนี้แล้วจะพิจารณาตัวแปรอีกตัวหนึ่ง คือ "เวลา" ซึ่งมีระยะเป็นปีให้เป็นตัวแทนของตัวแปรอื่น ๆ ที่ไม่ได้นำมาพิจารณา เช่น เมื่อรายได้ของประชากรคงที่ ราคาไฟฟ้าคงที่ แต่จำนวนพลังงานไฟฟ้าที่ใช้อาจจะเปลี่ยนไปได้เมื่อเวลาเปลี่ยนไปเป็นต้น

สมมุติให้ L = จำนวนพลังงานไฟฟ้าที่ใช้ในประเภทบ้านอยู่อาศัย

หน่วย : ล้านกิโลวัตต์-ชั่วโมง

DI = รายได้สุทธิส่วนบุคคล (Disposable Income)

หน่วย : พันล้านบาท

GP = มูลค่าผลิตภัณฑ์ของจังหวัด (GPP)

หน่วย : พันล้านบาท

P_1 = ราคาไฟฟ้าของประเภทบ้านอยู่อาศัย (Price)

หน่วย : สตางค์/หน่วย

T = เวลาเป็นปี (Time) โดยให้ปี 2506 = 1, 2507 = 2,.........

และเนื่องจาก DI และ GP มีความหมายใกล้เคียงกันมาก ดังนั้นชุดของตัวแปรที่ให้เลือกด้วยวิธี Stepwise regression จึงแยกออกเป็น 2 ชุด คือ

(1) P_1, DI, T

(2) P_1, GP, T

เมื่อนำตัวแปรทั้ง 2 ชุดมาเลือกด้วยวิธี Stepwise regression แล้วจะได้ผลทั้งในรูป Linear และ log ดังแสดงในตาราง 5.5.1

เมื่อพิจารณาสมการที่ได้จากทางด้านรูป linear จะเห็นว่าสมการที่ดีที่สุดคือสมการที่ 3 ในตาราง เพราะให้ค่า R^2 มากที่สุด และ s^2 น้อยที่สุด เป็นสมการที่เลือกจากตัวแปรชุดที่ (2)

ทางด้านรูป log สมการที่ดีที่สุดคือสมการที่ 4 ในตาราง

แต่เมื่อเปรียบเทียบค่า s^2 ซึ่งคำนวณในสเกลเลขคณิต ของสมการที่ดีที่สุดจากทั้ง 2 รูป ปรากฏว่า สมการที่ 3 ให้ค่า s^2 = 258.375 ซึ่งน้อยกว่าค่า s^2 = 698.617 ของรูป log ดังนั้นสมการที่ 3 ซึ่งมีรูปเป็น

$$\hat{L} = -250.2879 + 17.9092\ GP \qquad (5.1)$$

จึงเป็นรูปแบบทางคณิตศาสตร์ที่ดีที่สุดที่เลือกสำหรับความต้องการใช้พลังงานไฟฟ้าของผู้ใช้ไฟฟ้าประเภทบ้านอยู่อาศัยในเขตนครหลวง (ผลจากโปรแกรมคอมพิวเตอร์ของการเลือกสมการนี้โดยวิธี **Stepwise Regression** ได้แสดงไว้ในภาคผนวก ง - 1)

จากสมการจะสังเกตเห็นว่า ถึงแม้ค่า s^2 ของสมการที่ 1 จะน้อยกว่าสมการที่ 4 ซึ่งเป็นสมการที่ดีที่สุดทางรูป log ก็ตาม แต่ก็ไม่ควรจะนำสมการที่ 1 มาใช้ เพราะว่าเครื่องหมายของค่าสัมประสิทธิ์ (Regression coefficient) ของตัวแปร "ราคาไฟฟ้า" แสดงค่าเป็น " + " ซึ่งขัดกับหลักทฤษฎีทางเศรษฐศาสตร์ที่ว่า เมื่อราคาของสินค้าหรือบริการลดลง อุปสงค์ต่อสินค้านั้นจะมากขึ้น โดยเฉพาะกับพลังงานไฟฟ้าซึ่งเป็นสิ่งจำเป็นต่อการดำรงชีวิตในปัจจุบัน

<u>หมายเหตุ</u> ในการเลือกตัวแปรอิสระเข้าสมการ regression โดยวิธี Stepwise

ตารางที่ 5.5.1 ผลของ Stepwise Regression ประเภทบ้านอยู่อาศัย

ตัวแปรที่พิจารณา		สมการ Regression ที่ได้	R^2 (%)	s^2	det
(1) P_1, DI, T	1	$\hat{L} = -502.995 + 82.97\ T + 6.737\ P_1$ (7.6982) (2.3662)	99.28	502.950	0.0466
	2	$\widehat{\ln L} = 16.844 - 2.720 \ln P$ (0.1199)	97.91	0.009 (3268.370)	1
(2) P_1, GP, T	3	$\hat{L} = -250.288 + 17.909\ GP$ ** (0.3450)	99.59	258.375	1
	4	$\widehat{\ln L} = 0.971 + 1.276 \ln GP + 0.217 \ln T$ (0.1473) (0.0677)	99.31	0.003 (698.617)	0.096

หมายเหตุ ค่าที่อยู่ในวงเล็บ ( ) ใต้ค่า Regression coefficient คือค่าประมาณของความคลาดเคลื่อนของค่าประมาณข
Regression coefficient

ค่าที่อยู่ในวงเล็บ ( ) ใต้ค่า s^2 คือค่า s^2 ที่หาในสเกลเลขคณิต

** คือสมการที่เลือก

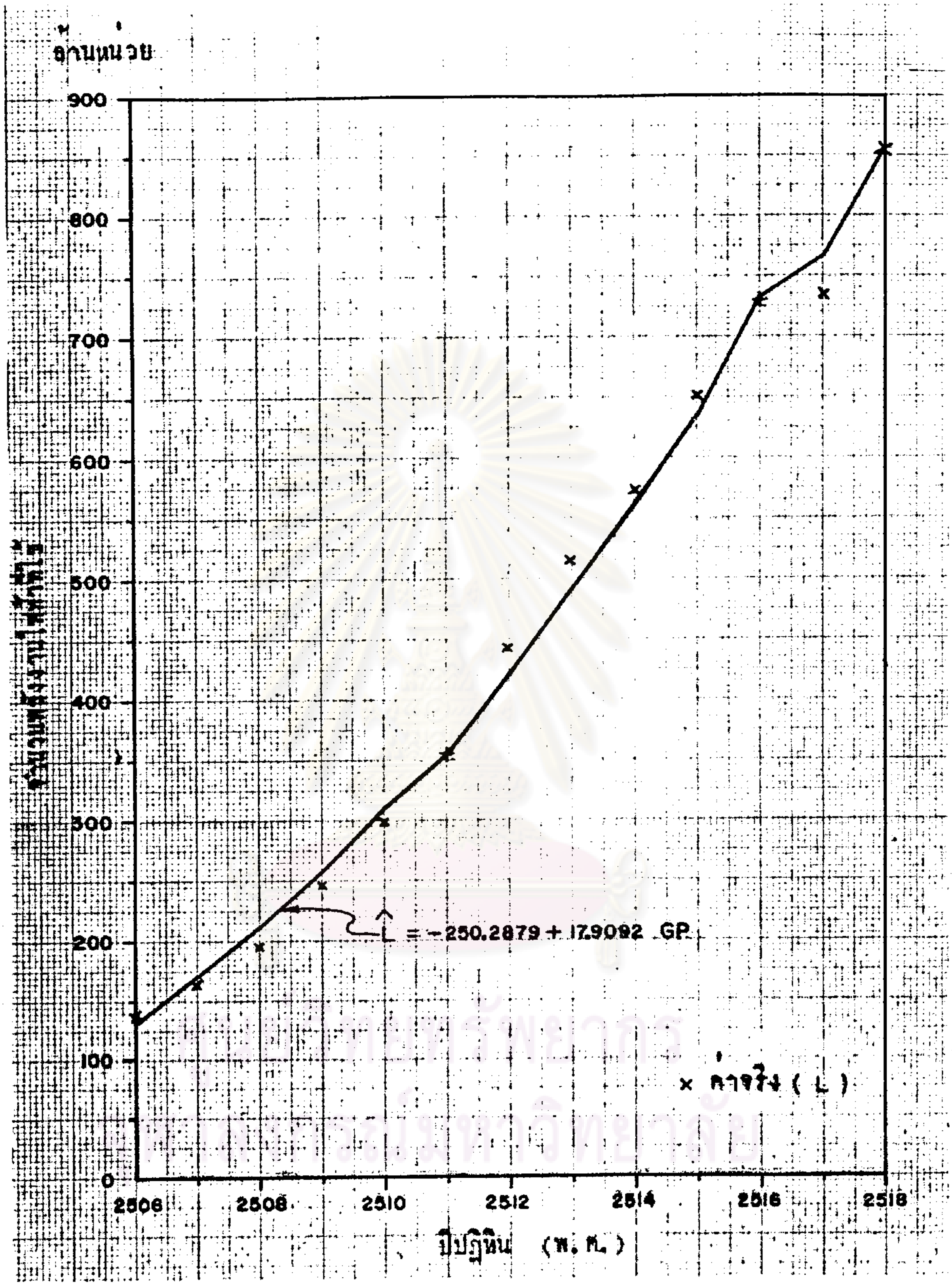
(ตารางที่ 5.6.1 - 5.8.1 ใช้หมายเหตุเดียวกันนี้)

Regression ของการวิจัยนี้ใช้ค่า F (1, 13-4, 0.95) = 5.12 จากตารางของค่า F ให้เป็นค่าคงที่สำหรับทดสอบค่า F ของตัวแปรอิสระที่จะนำเข้าหรือตัดออกจากสมการ (Sequential F และ Partial F test) โดยตั้งเกณฑ์ไว้ว่าจะให้ตัวแปรอิสระเข้าสมการไม่เกิน 3 ตัว ณ. ระดับนัยสำคัญ = 0.05 และถึงแม้จะใช้ระดับนัยสำคัญ = 0.10 หรือ F = 3.36 จากตารางก็ตาม ผลการเลือกส่วนใหญ่ก็ยังเหมือนเดิม หรือบางครั้งทำให้ค่าของสัมประสิทธิ์ผิดข้อเท็จจริง ดังนั้น จึงใช้ F = 5.12 เป็นเกณฑ์ในการทดสอบเท่านั้น

อนึ่ง จากผลการวิจัยปรากฏว่า รูปแบบที่ได้รับเลือกส่วนมากจะมีตัวแปรอิสระเข้าสมการไม่เกิน 3 ตัว ดังนั้นระดับนัยสำคัญในการทดสอบจริงจึงน้อยกว่า 0.05 เช่น ในกรณีของสมการที่ 4 จากตารางที่ 5.5.1 มีตัวแปรอิสระเข้าสมการ 2 ตัว ค่า partial F ของ GP และ T เท่ากับ 75.0753 และ 10.2407 ตามลำดับ แต่ค่าจากตารางของ F คือ F (1, 13 - 3, 0.95) = 4.96, F (1, 13 - 3, 0.975) = 6.94, F (1, 13-3, 0.99) = 10.04 แสดงว่า partial F ของตัวแปรทั้งสองผ่านการทดสอบในระดับนัยสำคัญ 0.025 และ 0.01 ด้วย

5.5.2 การวิเคราะห์ผลของรูปแบบ

จากรูปแบบที่เลือกได้ ของประเภทบ้านอยู่อาศัยนี้คือสมการ (5-1) แสดงให้เห็นว่าการใช้ไฟฟ้าของผู้ใช้ไฟฟ้าประเภทนี้ในเขตนครหลวงมีความสัมพันธ์กับมูลค่าผลิตภัณฑ์ทั้งหมดของเขตนครหลวงในราคาคงที่ปี 2505 มากที่สุด โดยมีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรทั้ง 2 สูงถึง 0.9979 และเป็นความสัมพันธ์ในทิศทางเดียวกัน ในลักษณะของสมการเส้นตรง ซึ่งมี intercept = -250.2879 และ slope = 17.9092 หรือหมายความว่า ถ้า GPP ของเขตนี้เพิ่มขึ้น 1 พันล้านบาท ความต้องการใช้พลังงานไฟฟ้าของประเภทนี้ก็จะเพิ่มขึ้นประมาณ 17.9092



ภาพที่ 5.1 เปรียบเทียบค่าจริงและค่าประมาณของจำนวนพลังงานไฟฟ้าที่ใช้ประเภทบ้านอยู่อาศัย

ล้านกิโลวัตต์ - ชั่วโมง (ล้านหน่วย) แต่ถ้า GPP ลดลง 1 พันล้านบาท การใช้ไฟฟ้าก็จะลดลงประมาณ 17.9092 ล้านหน่วย เช่นกัน และถึงแม้จะไม่มี GPP ผู้ใช้ไฟฟ้าประเภทนี้ก็ยังคงมีความต้องการที่จะใช้ไฟฟ้าอีกประมาณ 250.2879 ล้านหน่วย (ในสภาพความเป็นจริง GPP > 0 เสมอ)

การพิจารณาเลือกรูปแบบข้างต้นตามที่กล่าวมาแล้วเป็นผลจากการเลือกโดยใช้เทคนิคทางสถิติ แต่ถ้าพิจารณาจากความหมายของตัวแปรที่ได้รับเลือกแล้ว จะเห็นว่าสมการที่ 4 เป็นสมการที่น่าสนใจสำหรับการพยากรณ์แนวโน้มการใช้พลังงานไฟฟ้าในระยะยาวต่อไป เพราะได้รวมตัวแปรไว้ในสมการถึง 2 ตัว คือ GPP และเวลา โดยเฉพาะตัวแปร "เวลา" จะเป็นตัวแปรที่แทนตัวแปรอื่น ๆ ที่ไม่ได้นำมาพิจารณาซึ่งเปลี่ยนแปลงตามเวลาและมีผลต่อการใช้ไฟฟ้า

และการที่ตัวแปร "รายได้สุทธิส่วนบุคคล" หรือ "DI" ไม่ได้รับเลือกให้เข้าสมการดังเช่นผลการวิจัยของ EEI ที่ว่า ความต้องการใช้พลังงานไฟฟ้าของประเภทบ้านอยู่อาศัยนี้มีความสัมพันธ์กับรายได้สุทธิส่วนบุคคล (กล่าวมาแล้วในบทที่ 4) ทั้งนี้อาจจะเป็นเพราะว่าข้อมูลของตัวแปรดังกล่าวที่นำมาใช้ในการสร้างสมการนี้ต้องทำการประมาณมาหลายขั้นตอน ทำให้ค่าที่ได้มีความผิดพลาด หากในอนาคตมีการรวบรวมข้อมูลนี้อย่างสมบูรณ์แล้ว ผลการเลือกตัวแปรดังกล่าวอาจจะเป็นไปตามที่ EEI ศึกษาก็ได้

อย่างไรก็ตามรูปแบบที่เลือกได้ตามวิธีการทางสถิตินี้ก็มีเหตุผลดี กล่าวคือความต้องการใช้พลังงานไฟฟ้าของผู้ใช้ไฟฟ้าประเภทนี้ซึ่งเป็นประชาชนส่วนใหญ่ ขึ้นอยู่กับมูลค่ารวมของผลิตภัณฑ์ในเขตนี้ หรือผลิตผลซึ่งเกิดจากทุกสาขาเศรษฐกิจภายใน 4 จังหวัดนี้ และรูปแบบดังกล่าวมีรูปง่าย ๆ สะดวกต่อการคำนวณและหาข้อมูล

5.6 รูปแบบทางคณิตศาสตร์สำหรับประเภทธุรกิจการค้า

5.6.1 การหารูปแบบ

จากผลการศึกษาในบทที่ 4 ตัวแปรที่ควรจะนำมาพิจารณาสำหรับการสร้างรูปแบบทางคณิตศาสตร์ของประเภทนี้ได้แก่ GPP สาขาบริการ GPP สาขาค้าส่งและค้าปลีก และราคาไฟฟ้า (ในราคาคงที่ปี 2505 ของเขตนครหลวง) ตัวแปรอีกตัวหนึ่งนอกเหนือจาก

ที่กล่าวมานี้และจะนำมาพิจารณาเช่นเดียวกับประเภทบ้านอยู่อาศัย คือ "เวลา"

สมมุติให้ C = จำนวนพลังงานไฟฟ้าที่ใช้ในประเภทธุรกิจการค้า หน่วย : ล้านกิโลวัตต์-ชั่วโมง

SV = มูลค่ารวมของ GPP สาขาบริการ (Services) หน่วย : พันล้านบาท

WR = มูลค่ารวมของ GPP สาขาค้าส่งและค้าปลีก (Wholesale and Retail trade) หน่วย : พันล้านบาท

P_2 = ราคาไฟฟ้าประเภทธุรกิจการค้า หน่วย : สตางค์/หน่วย

ดังนั้น ชุดของตัวแปรที่จะนำมาพิจารณาด้วยวิธี Stepwise Regression ชุดหนึ่งคือ

(1) P_2, SV, WR, T

เนื่องจาก SV และ WR ต่างก็เป็นสาขาหนึ่งของมูลค่าผลิตภัณฑ์รวมของจังหวัด (GP) และจากผลของการหารูปแบบสำหรับประเภทบ้านอยู่อาศัยซึ่ง GP เป็นตัวแปรสำคัญที่ได้รับเลือกเข้าสมการ ทำให้สนใจที่จะพิจารณาตัวแปรอีกชุดหนึ่งคือ

(2) P_2, GP, T

และตัวแปรอีกชุดหนึ่งที่จะนำมาพิจารณาในประเภทนี้คือ

(3) P_2, T, L

เพราะจากการศึกษาแนวโน้มและโครงสร้างการใช้ไฟฟ้าในบทที่ 3 ปรากฏว่า ลักษณะการใช้ไฟฟ้าของประเภทธุรกิจการค้าคล้ายคลึงกับประเภทบ้านอยู่อาศัย เช่นเดียวกับผลการวิจัยของการไฟฟ้าแห่งประเทศฟิลิปปินส์[7] ซึ่งพบว่า จำนวนพลังงานไฟฟ้าที่จำหน่ายในฟิลิปปินส์ประเภทธุรกิจการค้านี้มีความสัมพันธ์กับจำนวนพลังงานไฟฟ้าที่จำหน่ายแก่ประเภทบ้านอยู่อาศัย ซึ่งแสดงในสมการ Regression ดังนี้

$$\hat{C} = -232.1805 + 1.2683\ L$$

โดยที่ $R^2 = 99.4\%$

7/ จากเอกสาร "Medium Term Sales Forecast Methodology" ที่ส่งมายัง กฟภ.

เมื่อนำตัวแปรทั้ง 3 ชุดไปเลือกตัวแปรอิสระด้วยวิธี Stepwise Regression โดยใช้ค่า F = 5.12 ในการทดสอบเช่นกัน จะได้ผลตามที่แสดงในตาราง 5.6.1

จะเห็นว่าสมการที่ให้ค่า R^2 มากที่สุด และ s^2 น้อยที่สุดเมื่อเปรียบเทียบทั้งรูป linear และ log แล้วคือสมการที่ 6 ในตาราง เป็นสมการรูป log ที่ได้จากตัวแปรชุดที่ (3) รองลงมาคือสมการที่ 5 เป็นสมการที่ได้จากตัวแปรชุดเดียวกัน แสดงว่าความต้องการใช้พลังงานไฟฟ้าประเภทธุรกิจการค้า (C) มีความสัมพันธ์กับความต้องการใช้พลังงานไฟฟ้าประเภทบ้านอยู่อาศัย (L) มาก และสัมพันธ์กับราคาไฟฟ้าของประเภทธุรกิจเอง (P_2) ด้วยค่าสัมประสิทธิ์สหสัมพันธ์เชิงซ้อน (R) ที่สูงมาก คือ R^2= 99.86% ในรูปของ log และ R^2= 99.83% ในรูป linear แต่สมการทั้งสองนี้เหมาะที่จะนำมาแสดงความสัมพันธ์ระหว่างตัวแปรดังกล่าวเมื่อทราบค่าของ L เท่านั้น ไม่ควรจะนำมาเป็นรูปแบบในการพยากรณ์ความต้องการในอนาคต เพราะจะขัดกับหลักของ Regression ที่ว่า "ค่าของตัวแปรอิสระที่จะนำมาสร้างสมการ regression จะต้องวัดมาอย่างไม่มีความผิดพลาด" 8/ เนื่องจากค่า L เป็นค่าที่เราสนใจที่จะหาแนวโน้มเพื่อการพยากรณ์เช่นกัน ถ้าผลจากการสร้างรูปแบบของ L นี้ผิดพลาด เมื่อนำไปแทนค่าเป็นตัวแปรอิสระในการพยากรณ์ C จะทำให้ค่าที่ได้ผิดพลาดมากขึ้น ดังนั้นจึงควรพิจารณาเลือกสมการอื่นมาใช้ต่อไป

สมการที่ดีที่สุดจากสมการที่เหลือคือ สมการที่ 1 ซึ่งเป็นสมการในรูป linear จากสมการชุดที่ (1) ที่ให้เลือก โดยมีค่า R^2 = 99.66% , s^2 = 141.188

สมการที่ดีที่สุดในรูป log คือสมการที่ 4 แต่ก็ยังด้อยกว่าสมการที่ 3 ซึ่งมีรูปเป็น linear เมื่อเปรียบเทียบค่า R^2 และ s^2

ถ้าเปรียบเทียบระหว่างสมการที่ 1, 3, 4 จะเห็นว่าคุณภาพของสมการทั้งสามมีความทัดเทียมกันมาก กล่าวคือมีค่า R^2 และ s^2 ใกล้เคียงกันมาก และถ้าพิจารณาค่า determinant ของ Correlation matrix ของตัวแปรอิสระที่เข้าสมการ จะเห็นว่าสมการที่ 4 ซึ่งเป็นรูป log ก็ยังคงเป็นสมการที่น่าสนใจคือ det = 0.0955 สูงกว่าอีกสองสมการ

8/ Robert G.D. Steel, James H. Torrei, Principles and Procedures of Statistics (New York : McGraw - Hill Book Co., 1960), p. 165

ตารางที่ 5.6.1 ผลของ Stepwise Regression ประเภทธุรกิจการค้า

ตัวแปรที่พิจารณา		สมการ Regression ที่ได้	R^2 %	s^2	det
(1) P_2, SV, WR, T	1	$\hat{C} = -253.302 + 56.722\ SV + 45.134\ WR$ ** (11.6496) (9.6753)	99.66	141.188	0.0309
	2	$\widehat{\ln C} = 1.850 + 1.946 \ln WR$ (0.0629)	98.86	0.003 (641.067)	1
(2) P_2, GP, T	3	$\hat{C} = 1.081 + 11.689\ GP - 2.336\ P_2$ (0.9376) (1.0132)	99.63	152.069	0.0797
	4	$\widehat{\ln C} = 1.829 + 1.080 \ln GP + 0.130 \ln T$ (0.1195) (0.0549)	99.26	0.002 (293.541)	0.0955
(3) P_2, T, L	5	$\hat{C} = 230.123 + 0.669\ L - 1.998\ P_2$ (0.0361) (0.7011)	99.83	71.369	0.0718
	6	$\widehat{\ln C} = 3.816 + 0.629 \ln L - 0.406 \ln P_2$ (0.0576) (0.1528)	99.86	0.000 (65.131)	0.0265

อย่างไรก็ตามสมการที่จะเลือกสำหรับการใช้ไฟฟ้าของประเภทธุรกิจการค้านี้คือสมการที่ 1 ซึ่งมีรูปเป็น

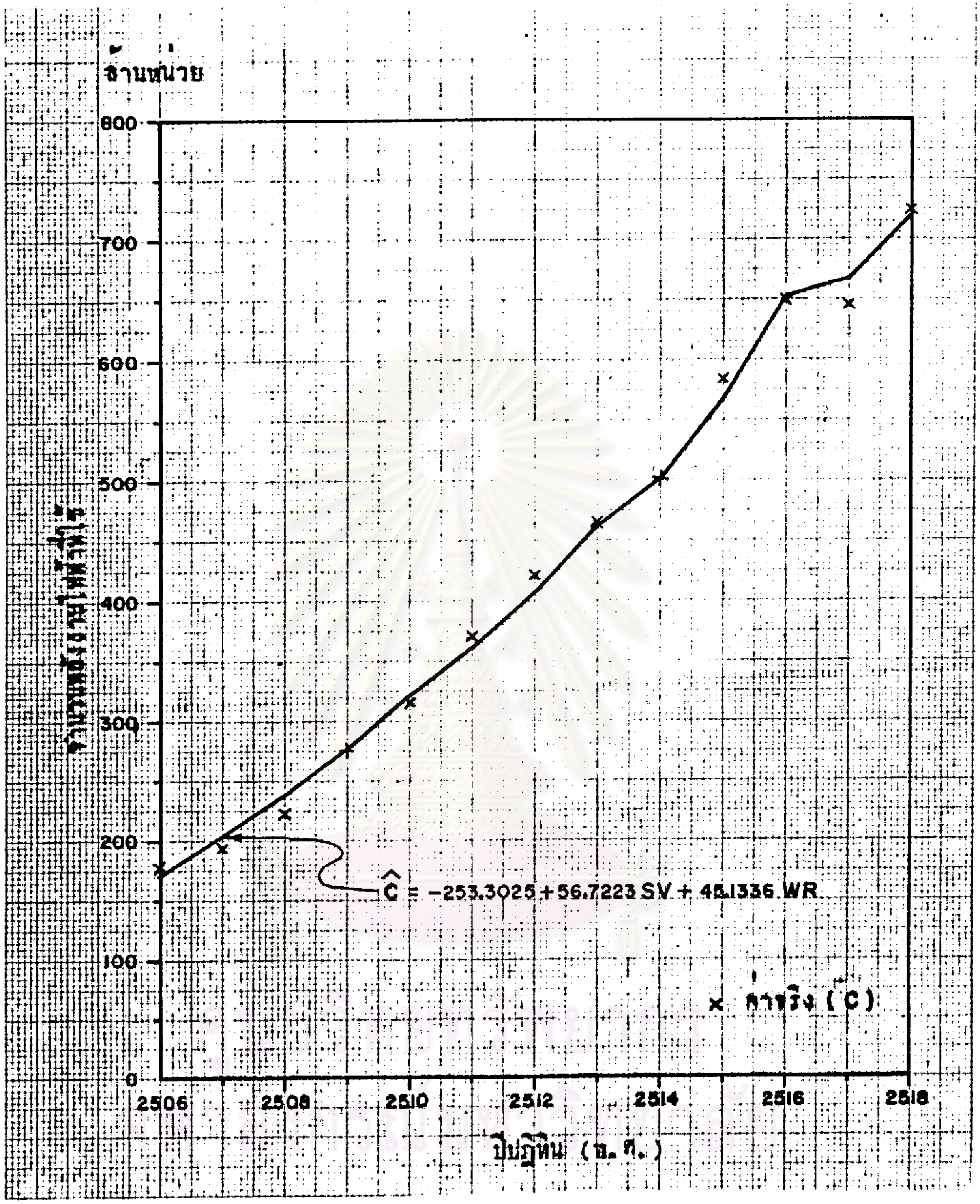
$$\hat{C} = -253.3025 + 56.7223\ SV + 45.1336\ WR \qquad (5\text{-}2)$$

เพราะให้ค่า s^2 น้อยที่สุด

(รายละเอียดของการหารูปแบบนี้แสดงในภาคผนวก ง-)

5.6.2 การวิเคราะห์ผลของรูปแบบ

จากรูปแบบที่ได้คือสมการ (5-2) แสดงให้เห็นว่า ความต้องการใช้พลังงานไฟฟ้าของประเภทธุรกิจการค้านี้มีแนวโน้มขึ้นอยู่กับ GPP สาขาบริการ (SV) และ GPP สาขาค้าส่งและค้าปลีก (WR) ดังที่คาดหมายไว้ในบทที่ 4 แสดงว่าตัวแปรทั้งสองนี้มีความหมายใกล้เคียงกับ "รายจ่ายภาคเอกชนเพื่อซื้อบริการ" ซึ่ง EEI ได้วิเคราะห์ไว้ว่ามีความสัมพันธ์กับการใช้พลังงานไฟฟ้าของประเภทนี้ โดยที่ SV และ WR ต่างก็สัมพันธ์กับความต้องการใช้พลังงานไฟฟ้าของประเภทนี้ในทิศทางเดียวกัน ด้วยค่าประมาณของ **Partial regression coefficient** เท่ากับ 56.7223 และ 45.1336 ตามลำดับ กล่าวคือ เมื่อ SV เพิ่มขึ้น 1 พันล้านบาท และ WR คงที่ จำนวนพลังงานไฟฟ้าที่ต้องการของประเภทธุรกิจการค้านี้จะเพิ่มขึ้นประมาณ 56.7223 ล้านหน่วย และถ้า SV ลดลง 1 พันล้านบาท จำนวนพลังงานไฟฟ้าที่ใช้ก็จะลดลงประมาณ 56.7223 ล้านหน่วยเช่นกัน ในทำนองเดียวกันเมื่อ WR เปลี่ยนแปลง 1 พันล้านบาท โดยที่ SV คงที่ จำนวนพลังงานที่ต้องการใช้ก็จะเปลี่ยนแปลงเท่ากับประมาณ 45.1336 ล้านหน่วย ตามกัน และถึงแม้ SV และ WR จะมีค่าเท่ากับ 0 ก็ยังจะมีความต้องการจำนวนพลังงานไฟฟ้าอีกประมาณ 253.3025 (**intercept** ของสมการ = -203.3525) แสดงว่าการใช้



ภาพที่ 5.2 เปรียบเทียบค่าจริงและค่าประมาณของจำนวนพลังงานไฟฟ้าที่ใช้ประเภทธุรกิจการค้า

ไฟฟ้าของประเภทนี้ขึ้นอยู่กับกิจกรรมทางธุรกิจ คือ ขายสินค้าและบริการ

จากผลการเลือกสมการของประเภทนี้ แสดงให้เห็นว่า GP หรือผลิตภัณฑ์ของจังหวัดยังคงเป็นตัวแปรที่น่าสนใจ โดยได้รับเลือกเข้าเป็นตัวแปรอิสระในสมการทั้งรูป linear และ log และสมการที่ได้ก็มีคุณภาพในการประมาณดี (ค่า R^2 สูง, s^2 น้อย, det มาก) โดยเฉพาะสมการที่ 4 ในตาราง 5.6.1 นี้ ตัวแปรที่ได้เป็นชุดเดียวกับสมการที่ 4 ในตาราง 5.5.1 ของประเภทบ้านอยู่อาศัย แต่เนื่องจาก GPP สาขาบริการและสาขาค้าส่งและค้าปลีกแสดงผลต่อการใช้ไฟฟ้าเด่นชัดกว่า GPP รวมทุกสาขา สมการที่ 1 จึงดีกว่า อย่างไรก็ตามสำหรับการพยากรณ์ในระยะยาวนั้นคาดว่าสมการที่ 4 ก็ใช้ได้ดีเช่นกัน เพราะมีตัวแปร "เวลา" อยู่ด้วย

5.7 รูปแบบทางคณิตศาสตร์สำหรับประเภทอุตสาหกรรม

5.7.1 การหารูปแบบ

จากผลการวิเคราะห์ในบทที่ 3 และจากความรู้ที่ได้จากการหารูปแบบสำหรับ 2 ประเภทแรกที่ผ่านมา ทำให้สามารถหารูปแบบที่เหมาะสมสำหรับประเภทอุตสาหกรรมนี้ได้ง่ายขึ้นดังนี้

ถ้าให้ I = จำนวนพลังงานไฟฟ้าที่ใช้ในประเภทอุตสาหกรรม
หน่วย : ล้านกิโลวัตต์-ชั่วโมง

MF = มูลค่าของ GPP สาขาอุตสาหกรรม (Manufacturing)
หน่วย : พันล้านบาท

PA = มูลค่าของ GPP สาขาบริหารราชการและป้องกันประเทศ (Public Administration and defence)
หน่วย : พันล้านบาท

P_3 = ราคาไฟฟ้าประเภทอุตสาหกรรม

(ตัวแปรทั้งสามนี้เป็นค่าในราคาคงที่ปี 2505 และอยู่ในขอบเขตพื้นที่นครหลวง)

ชุดของตัวแปรที่จัดให้เลือก คือ

(1) MF, SV, PA, P_3, T

(2) GP, P_3, T

และผลที่ได้จาก Stepwise Regression แสดงในตาราง 5.7.1 (ใช้ค่า F = 5.12 ในการทดสอบ sequential F และ partial F ของ regression coefficient เช่นกัน)

รูปแบบที่ดีที่สุดที่เลือกสำหรับประเภทอุตสาหกรรมนี้คือสมการที่ 1 ที่มีรูปเป็น

$$\hat{I} = -1198.9253 + 135.6404\ MF + 604.7317\ PA \quad (5\text{-}3)$$

เพราะมีค่า R^2 สูงที่สุด และ s^2 น้อยที่สุด รองลงไปคือสมการที่ 3 ในตาราง ซึ่งเป็นรูป linear เช่นกัน

ส่วนสมการทางด้านรูป log นั้น ถึงแม้จะมีค่า R^2 ที่สูงใกล้เคียงกับของรูป linear ก็ตาม แต่ถ้านำค่า s^2 มาเปรียบเทียบกันจะเห็นว่าค่า s^2 ของสมการทางด้านรูป log สูงกว่ามาก แสดงว่าไม่เหมาะสมที่จะนำมาใช้ โดยเฉพาะสมการที่ 2 ค่า det ของ correlation matrix มีค่าน้อยมาก คือ 0.0083 แสดงว่าตัวแปรอิสระที่เลือกเข้าสมการมีความสัมพันธ์กันมาก

5.7.2 การวิเคราะห์ผลของรูปแบบ

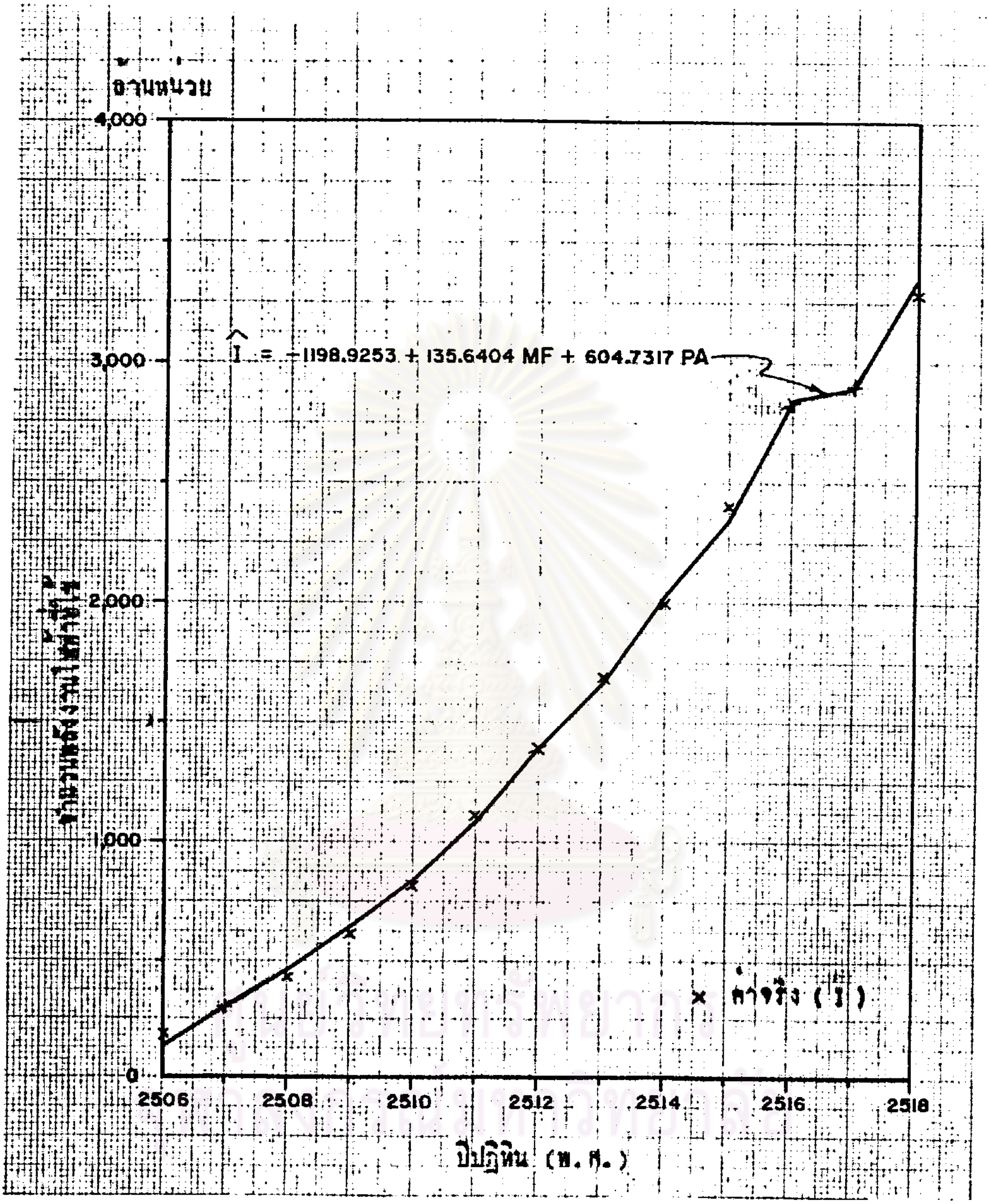
ตารางที่ 5.7.1 ผลของ Stepwise Regression ประเภทอุตสาหกรรม

ตัวแปรที่พิจารณา		สมการ Regression ที่ได้	R^2 (%)	s^2	det
(1) MF, SV, PA, P_3, T	1	$\hat{I} = -1198.925 + 135.640\ MF + 604.732\ PA$ ** (10.3351) (117.0650)	99.90	1424.000	0.0341
	2	$\widehat{\ln I} = 6.968 + 0.626 \ln T + 1.248 \ln PA - 0.509 \ln P_3$ (0.0621) (0.1581) (0.1417)	99.85	0.002 (6841.351)	0.0083
(2) GP, P_3, T	3	$\hat{I} = -1608.314 + 79.899\ GP$ (1.0016)	99.83	2177.727	1
	4	$\widehat{\ln I} = 1.798 + 0.754 \ln T + 1.079 \ln GP$ (0.0796) (0.1731)	99.61	0.004 (13949.021)	0.0955

จากรูปแบบที่เลือกได้นี้แสดงว่า การใช้ไฟฟ้าในประเภทอุตสาหกรรมนี้มีแนวโน้มขึ้นอยู่กับมูลค่าผลิตภัณฑ์ของจังหวัดสาขาอุตสาหกรรม (MF) และสาขาบริหารราชการและป้องกันประเทศ (PA) โดยมีความสัมพันธ์ในทิศทางเดียวกัน ด้วยค่าประมาณของ partial Regression coefficient เท่ากับ 135.6404 และ 604.7317 ตามลำดับ และค่า intercept = -1198.9253 แสดงว่าผลที่ได้เป็นไปตามผลการเลือกปัจจัยที่มีผลต่อการใช้ไฟฟ้าประเภทอุตสาหกรรมนี้โดยดูจากจำนวนพลังงานไฟฟ้าที่ใช้ในประเภทนี้แยกตามสาขาเศรษฐกิจ (ในบทที่ 4) ซึ่งสาขาที่ใช้พลังงานไฟฟ้าในประเภทนี้มากที่สุดเรียงตามลำดับจากมากไปน้อยคือ MF, SV และ PA แต่ด้วยเหตุที่ MF และ SV มีความสัมพันธ์กันมาก (r = 0.9964) ด้วยวิธีของ Stepwise Regression จึงทำให้ตัวแปรทั้งสองเข้าสมการพร้อมกันไม่ได้ และเลือก MF เข้าสมการเท่านั้น เพราะ MF มีความสัมพันธ์กับการใช้ไฟฟ้าประเภทนี้มากกว่า

เมื่อเปรียบเทียบค่าประมาณของ parameter ทั้งหมดที่ได้ กล่าวคือค่าประมาณของ partial regression coefficient และค่า intercept ของรูปแบบที่เลือกในประเภทอุตสาหกรรมนี้ กับค่าประมาณที่ได้จากรูปแบบที่เลือกได้ในอีก 2 ประเภทที่ผ่านมาแล้ว ในสมการที่ (5-1) และ (5-2) ซึ่งต่างก็เป็นรูป linear ด้วยกัน ตัวแปรที่เข้าสมการคือ GP, SV, WR, MF, PA ต่างก็มีหน่วยเป็นพันล้านบาท จะพบว่า ค่าประมาณเหล่านี้ในประเภทอุตสาหกรรมสูงกว่าอีก 2 ประเภท เช่น เมื่อ MF เพิ่มขึ้น 1 พันล้านบาท จะทำให้การใช้ไฟฟ้าประเภทอุตสาหกรรมเพิ่มขึ้น 135.6404 ล้านหน่วย (ค่าสัมประสิทธิ์ของ MF) แต่เมื่อ SV เพิ่มขึ้น 1 พันล้านบาท การใช้ไฟฟ้าประเภทธุรกิจการค้าจะเพิ่มขึ้นเพียง 56.7223 พันล้านบาท และเมื่อ GP เพิ่มขึ้น 1 พันล้านบาท การใช้ไฟฟ้าประเภทบ้านอยู่อาศัยก็จะเพิ่มขึ้นเพียง 17.9092 บาทเท่านั้น แสดงให้เห็นว่าประเภทอุตสาหกรรมนี้มีส่วนในการใช้ไฟฟ้ามาก

มีข้อที่น่าสังเกตสำหรับการหารูปแบบประเภทอุตสาหกรรมนี้เช่นเดียวกับที่ผ่านมาคือ เมื่อพิจารณาจากตัวแปรชุดที่ (2) ซึ่งมี GP อยู่ด้วย ปรากฏว่า GP จะได้รับเลือกเข้าทั้งสมการในรูป linear และรูป log โดยเฉพาะรูป log ตัวแปรที่เลือกได้คือ GP และ T ซึ่งเป็นตัวแปรคู่เดียวกันที่เลือกได้จากตัวแปรชุดเดียวกันใน 2 ประเภทที่ผ่านมา แสดงว่า GP มีความสัมพันธ์กับการใช้ไฟฟ้าทั้ง 3 ประเภทนี้ และคาดว่า T จะเป็นตัวแปรตัวหนึ่งที่จะอธิบายผลจากตัวแปรอื่นที่ไม่ได้อยู่ในสมการได้ดี



ภาพที่ 5.3 เปรียบเทียบค่าจริงและค่าประมาณของจำนวนพลังงานไฟฟ้าที่ใช้

ประเภทอุตสาหกรรม

5.8 รูปแบบทางคณิตศาสตร์สำหรับประเภทไฟถนนสาธารณะ

5.8.1 การหารูปแบบ

จากการศึกษาในบทที่ 4 ปัจจัยที่จะนำมาพิจารณาในที่นี้คือ จำนวนประชากร จำนวนบ้าน GPP สาขาบริการ และเวลา ดังนั้นวิธีการเลือกสมการจะเป็นดังนี้

กำหนดให้ S = จำนวนพลังงานไฟฟ้าที่ใช้ในประเภทไฟถนนสาธารณะ หน่วย : ล้านกิโลวัตต์ - ชั่วโมง

PP = จำนวนประชากร (Population) หน่วย : ล้านคน

H = จำนวนบ้าน (Houses) หน่วย : พันหลัง

สมการชุดที่จะให้พิจารณาเลือกด้วยวิธี Stepwise Regression มีดังนี้

(1) PP, T

(2) H, T

(3) SV, T

และจากผลที่ผ่านมา จะนำ GP มาพิจารณาอีกชุดหนึ่งคือ

(4) GP, T เช่นเคย

ผลที่ได้จาก Stepwise Regression ด้วยค่า F = 5.12 ได้แสดงไว้ในตาราง 5.8.1 และมีวิธีการตัดสินดังนี้

ทางด้านรูป linear สมการที่ให้ผลดีที่สุดคือ สมการที่ 7 เพราะให้ค่า R^2 ที่มากกว่า และ s^2 ที่น้อยกว่า (R^2 = 95.08%, s^2 = 5.562)

ทางด้านรูป log นั้น สมการที่ให้ผลดีที่สุดคือ สมการที่ 8 โดยมีค่า R^2 = 95.39% s^2 = 0.020 (แต่มีค่าเท่ากับ 10.605 เมื่อคำนวณในคาสเกลเลขคณิต)

เมื่อเปรียบเทียบรูปที่ได้จากด้าน linear กับ log จะเห็นว่า ถึงแม้ R^2 ในรูป log จะมีค่ามากกว่า แต่ค่า s^2 ของรูป linear น้อยกว่า ดังนั้นรูปแบบที่เลือกสำหรับประเภทไฟถนนสาธารณะนี้คือ สมการที่ 7 จากตาราง 5.8.1 หรือ

$$\hat{S} = -10.0721 + 0.7377\ GP \qquad (5\text{-}4)$$

ตารางที่ 5.8.1 ผลของ Stepwise Regression ประเภทไฟถนนสาธารณะ

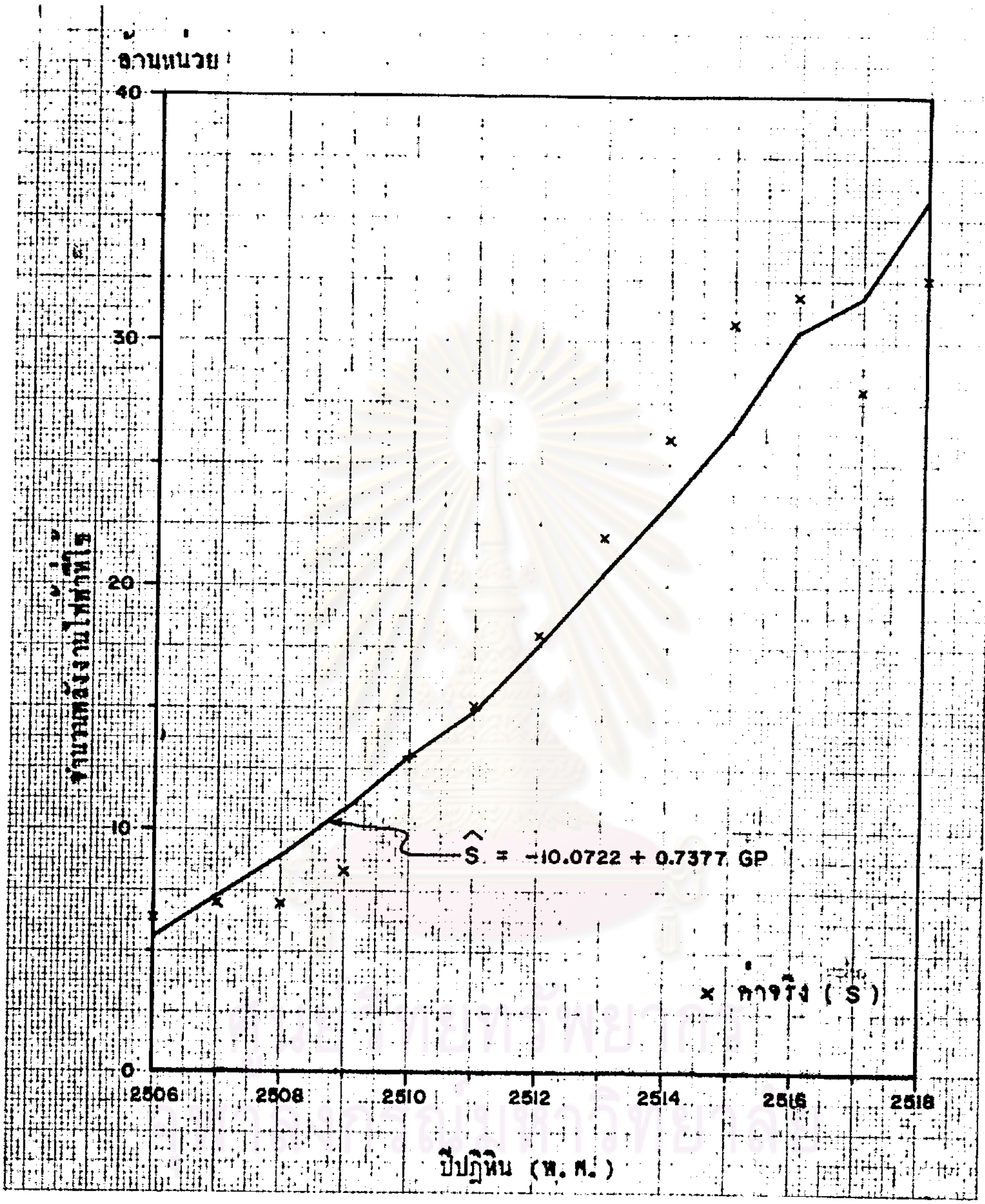
ตัวแปรที่พิจารณา	สมการ Regression ที่ได้		R^2 (%)	s^2	det
(1) PP, T	1	$\hat{S} = 1.157 + 2.543\ T$ (0.1816)	94.69	6.000	1
	2	$\widehat{\ln S} = -2.945 + 3.965 \ln PP$ (0.3474)	92.21	0.034 (21.664)	1
(2) H,T	3	(ได้ผลเช่นเดียวกับสมการที่ 1)			
	4	$\widehat{\ln S} = -17.597 + 3.199 \ln H$ (0.2460)	93.89	0.026 (15.666)	1
(3) SV, T	5	(ได้ผลเช่นเดียวกับสมการที่ 1)			
	6	$\widehat{\ln S} = -0.259 + 1.887 \ln SV$ (0.1445)	93.94	0.026 (15.800)	1
(4) GP, T	7	$\hat{S} = -10.072 + 0.738\ GP$ ** (0.0506)	95.08	5.562	1
	8	$\widehat{\ln S} = -3.538 + 1.746 \ln GP$ (0.1158)	95.39	0.020 (10.605)	1

5.8.2 การวิเคราะห์ผลของรูปแบบ

จากรูปแบบที่เลือกได้ คือสมการที่ (5-4) แสดงว่า ผลิตภัณฑ์ของจังหวัด (GP) ก็ยังเป็นปัจจัยที่สำคัญในการกำหนดอุปสงค์ของการใช้ไฟฟ้าประเภทนี้ หรือเป็นตัวแปรอิสระที่จะอธิบายผลที่มีต่อการใช้ไฟฟ้าประเภทนี้ ในรูปของสมการเส้นตรง ด้วยค่าคงที่ intercept = -10.072 และค่าคงที่ที่แสดงการเปลี่ยนแปลงของจำนวนพลังงานไฟฟ้าที่ใช้ต่อ 1 หน่วยของ GP ที่เปลี่ยนไป หรือค่าสัมประสิทธิ์ของ GP = 0.738

สมการที่ให้ผลคู่คี่กับสมการที่เลือกคือสมการที่ 1 ในตาราง ซึ่งตัวแปรที่เลือกคือ "เวลา" แต่ในการวิจัยนี้ไม่ได้นำค่ายกกำลังของตัวแปรอิสระที่มากกว่ากำลังหนึ่งมาพิจารณา (T^2, T^3 เป็นต้น) ผลที่ได้จึงเป็นเพียงสมการรูปกำลังหนึ่งของตัวแปร ซึ่งหากนำสมการนี้ไปใช้กับการพยากรณ์แนวโน้มในระยะยาว จะทำให้ขัดกับหลักความจริง เพราะการใช้ไฟฟ้าย่อมจำกัดอยู่ในขอบเขตของพื้นที่หนึ่ง ๆ (ในที่นี้คือเขตนครหลวง) และจะต้องมีจุดอิ่มตัว ไม่ใช่เมื่อเวลาเพิ่มขึ้น การใช้ไฟฟ้าก็เพิ่มขึ้นตามกันอย่างไม่มีที่สิ้นสุด โดยเฉพาะในเรื่องไฟถนนสาธารณะ ก็ยิ่งเป็นไปไม่ได้ ดังนั้นจึงไม่ควรเลือกสมการดังกล่าว อย่างไรก็ตามเท่าที่เป็นอยู่ในขณะนี้ในเขตนครหลวงนี้ยังมีถนนหนทางต่าง ๆ ที่ยังไม่มีไฟฟ้าใช้อีกมาก ดังนั้นในระยะสั้นสมการนี้ก็ยังจะใช้ได้ผลดี และไม่ต้องหาข้อมูลของตัวแปรอื่นมาใช้

ถ้าหากนำรูปแบบที่ได้ของทุกประเภทมาเปรียบเทียบกันจะพบว่า ประสิทธิภาพของรูปแบบที่ได้ในประเภทไฟถนนสาธารณะนี้น้อยกว่าของประเภทอื่น ๆ ที่กล่าวมา คือในรูปแบบอื่น ๆ จะมีค่า R^2 ตั้งแต่ 99% ขึ้นไป และมีค่าสัมประสิทธิ์ของการกระจาย (Coefficient of variation หรือ C.V.) น้อยกว่า 4 แต่ในประเภทนี้มีค่า R^2 เพียง 95.07%และ



ภาพที่ ๕.๔ เปรียบเทียบค่าจริงและค่าประมาณของจำนวนพลังงานไฟฟ้าที่ใช้ประเภทไฟถนนสาธารณะ

C.V. สูงถึง 12.44%

$$\text{โดยที่ C.V.} = \frac{\text{ค่าประมาณของ Standard error of estimate (s)}}{\text{ค่ามัชฌิมเลขคณิตของตัวแปรตาม}}$$

และแสดงในชื่อ "STD. ERROR AS A % OF MEAN RESPONSE" ในผลที่ได้จากคอมพิวเตอร์ของรูปแบบที่เลือกซึ่งแสดงในภาคผนวก ง.

แสดงว่า ตัวแปรอิสระต่าง ๆ ที่พิจารณาไม่แสดงผลต่อการใช้ไฟฟ้าประเภทนี้มากเท่ากับตัวแปรอิสระในประเภทอื่น ซึ่งเป็นไปตามผลการวิเคราะห์ในบทที่ 3 ที่ว่า จำนวนการเปลี่ยนแปลงของไฟถนนสาธารณะที่ใช้นี้ขึ้นอยู่กับนโยบายการควบคุมมากกว่าผลทางเศรษฐกิจและสังคม แต่ที่รูปแบบที่เลือกสำหรับประเภทนี้ มีตัวแปรอิสระคือ GP นั้นอาจจะเป็นเพราะว่า ในระยะที่รัฐบาลดำเนินนโยบายประหยัดน้ำมันนั้น ได้ลดไฟฟ้าสาธารณะที่ใช้ลงเป็นจำนวนมาก และเป็นระยะเดียวกับที่ GP ต่ำ จึงทำให้เห็นผลของ GP ที่มีต่อการใช้ไฟฟ้าเด่นชัด

5.9 รูปแบบทางคณิตศาสตร์สำหรับรวมทุกประเภท

จากผลของรูปแบบแต่ละประเภทที่เลือกได้คือ สมการที่ (5-1) ถึง (5-4) ซึ่งได้แก่

$$\hat{L} = -250.2879 + 17.9092\ GP$$

$$\hat{C} = -253.3025 + 56.7223SV + 45.1336\ WR$$

$$\hat{I} = -1198.9253 + 135.6404\ MF + 604.7317\ PA$$

$$\hat{S} = -10.0721 + 0.7377\ GP$$

เมื่อนำมารวมกันเข้าทุกประเภทจะได้

$$\hat{E} = \hat{L} + \hat{C} + \hat{I} + \hat{S}$$

$$= -1712.5878 + 18.6469\ GP + 56.7223\ SV + 45.1336\ WR + 135.6404\ MF + 60.7317\ PA \quad \text{———— (5-5)}$$

เป็นรูปแบบที่สร้างขึ้นเพื่อแสดงความสัมพันธ์ระหว่างจำนวนพลังงานไฟฟ้าที่ใช้ทั้งหมดในเขตนครหลวงกับปัจจัยต่าง ๆ ที่เกี่ยวข้องคือ GPP และสาขาต่าง ๆ ของ GPP (สาขาบริการ, ค้าส่งและค้าปลีก, อุตสาหกรรม, บริหารราชการแผ่นดินและป้องกันประเทศ)

เนื่องจากรูปแบบดังกล่าวได้จากผลรวมของรูปแบบที่ดีที่สุดที่เลือกสำหรับความต้องการใช้พลังงานไฟฟ้าของแต่ละประเภท ดังนั้นรูปแบบที่ได้นี้หรือสมการ (5-5) จึงเป็นรูปแบบที่ดีที่สุดสำหรับความต้องการใช้พลังงานไฟฟ้ารวมทุกประเภทด้วย

และเนื่องจากรูปแบบดังกล่าวประกอบด้วยตัวแปรอิสระถึง 5 ตัว ทำให้สามารถอธิบายความกระจายของตัวแปรตามได้ดี และตัวแปรอิสระที่ได้ก็สามารถแสดงผลของการเปลี่ยนแปลงตามภาวะเศรษฐกิจ ดังนั้น รูปแบบดังกล่าวจึงใช้ได้ดีกับการพยากรณ์ ความต้องการใช้พลังงานไฟฟ้าในอนาคตทั้งในระยะสั้นและระยะยาว เช่น 10-20 ปี เพราะเราสามารถจะกำหนดค่าของตัวแปรอิสระตามความคาดหมายทางภาวะเศรษฐกิจได้โดยไม่ต้องแก้ไขรูปของสมการ หรือค่าคงที่ต่าง ๆ โดยเฉพาะในเวลาที่ภาวะเศรษฐกิจเปลี่ยนไปและได้พยากรณ์การใช้ไฟฟ้าล่วงหน้าไว้แล้ว จะทำให้สามารถแก้ไขผลการพยากรณ์ในอนาคตให้ใกล้เคียงกับสภาพความเป็นจริงได้ทันห่วงที

ในทางปฏิบัติ อาจจะมีข้อเสียสำหรับการใช้รูปแบบดังกล่าวอยู่บ้าง เพราะจะต้องหาข้อมูลสำหรับตัวแปรอิสระหลายตัว และจะต้องคำนวณมากกว่ารูปแบบที่มีจำนวนตัวแปรน้อย

อย่างไรก็ตามข้อเสียดังกล่าวเป็นเพียงส่วนน้อยเมื่อเทียบกับข้อดี ดังนั้นจึงยังคงเหมาะสมที่จะใช้รูปแบบนี้ต่อไป

ถึงแม้รูปแบบต่าง ๆ ที่ได้จะมีประสิทธิ์ภาพดีมากในแง่ของสถิติ คือมีค่า R^2 สูงมากก็ตาม แต่ก็เป็นเพียงการแสดงความเหมาะสมของสมการที่ได้กับข้อมูลที่นำมาวิเคราะห์เท่านั้น (fit กับข้อมูลได้ดี) และจะนำไปใช้ในการพยากรณ์ได้ดีก็ต่อเมื่อภาวะเศรษฐกิจและโครงสร้างของสังคมของเขตนครหลวงนี้ไม่เปลี่ยนแปลงไปจากที่เป็นอยู่ในช่วงของข้อมูลที่นำมาศึกษามากจนผิดปกติเท่านั้น

5.10 การพยากรณ์

ค่าประมาณจากรูปแบบทางคณิตศาสตร์ของแต่ละประเภทผู้ใช้ไฟฟ้าซึ่งอยู่ในลักษณะ

$$\hat{Y} = b_0 + b_1x_1 + b_2x_2 + \dots\dots\dots + b_px_p$$

จะเป็นค่าประมาณของ

$$E\ (Y) = B_0 + B_1X_1 + B_2X_2 + \cdots\cdots + B_pX_p$$ ซึ่งเป็นค่ามัชฌิมของ

Y ณ. ค่า x ชุดหนึ่ง ๆ ที่กำหนดให้

หรือถ้าเขียนในรูปของ Matrix ค่าประมาณที่ได้คือ

$$\hat{\underline{Y}} = \underline{X}\ \underline{b}$$

เมื่อ $$\hat{\underline{Y}} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix}, \quad \underline{X} = \begin{bmatrix} 1 & X_{11} & X_{12} \cdots\cdots X_{1p} \\ 1 & X_{21} & X_{22} \cdots\cdots X_{2p} \\ \vdots & & \vdots \\ 1 & \cdots\cdots & X_{np} \end{bmatrix}, \quad \underline{b} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_p \end{bmatrix}$$

n = จำนวนค่าของข้อมูลที่นำมาสร้างรูปแบบ (observations)

p = จำนวน parameters

สมมุติว่า $'\underline{X}_0$ เป็นค่า X ชุดหนึ่งที่กำหนดให้คือ

$$\underline{X}_0 = \begin{bmatrix} 1 \\ X_{01} \\ X_{02} \\ \vdots \\ X_{0p} \end{bmatrix}$$

จะได้ $$\hat{Y} = \underline{X}_0'\ b = \begin{bmatrix} 1 & X_{01} & X_{02} \cdots\cdots\cdots X_{0p} \end{bmatrix} \times \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_p \end{bmatrix}$$

ค่าประมาณของค่าแปรปรวนของ $\hat{Y}$ ที่ค่า $\underline{X}_0$ หรือ

ค่าประมาณของ $$V(\hat{Y}) = s^2\ (\underline{X}_0'\ \underline{C}\ \underline{X}_0)$$

โดยที่ $$\underline{C} = (\underline{X}'\ \underline{X})^{-1}$$

และสามารถหาขอบเขตความเชื่อมั่น 1- α % ของค่ามัชฌิมของ Y (1 - α % Confidence Limits on the true mean value) ณ. ค่า $\underline{X}_0$ ที่กำหนดให้ได้คือ 9/

$$\hat{Y} \pm t \left\{ (n - p - 1), 1 - \tfrac{1}{2}\alpha \right\} \cdot s \sqrt{\underline{X}_0' \underline{C} \underline{X}_0}$$

ดังนั้นถ้าเราหาค่า $\underline{X}_0$ ต่าง ๆ ซึ่งเป็นตัวแปรอิสระในอนาคตได้ จะสามารถพยากรณ์ค่ามัชฌิม และขอบเขตความเชื่อมั่นของค่ามัชฌิมของ Y ที่ค่า $\underline{X}_0$ เหล่านั้นได้

ในที่นี้ได้หาค่าพยากรณ์พร้อมทั้งขอบเขตความเชื่อมั่น 95% ของค่ามัชฌิมของจำนวนพลังงานไฟฟ้าที่ใช้ของผู้ใช้ไฟฟ้าแต่ละประเภทสำหรับปี 2519 และปีในระยะแผนพัฒนาที่ 4 คือ ปี 2520-2524 ด้วยรูปแบบที่หาได้ และค่าประมาณของตัวแปรอิสระคือ GPP และสาขาต่าง ๆ ของ GPP ในปี 2519-2524 โดยอาศัยโปรแกรมคอมพิวเตอร์ซึ่งเขียนขึ้นตามวิธีการคำนวณดังกล่าวข้างต้น (รายละเอียดต่าง ๆ ของผลการพยากรณ์ได้แสดงไว้ในภาคผนวก ฉ.)

สำหรับค่าพยากรณ์ของจำนวนพลังงานไฟฟ้าที่ใช้ทั้งหมดของเขตนครหลวง (รวมทุกประเภท) ได้จากการรวมค่าพยากรณ์ของแต่ละประเภทในปีเดียวกันเข้าด้วยกันคือ

$$\hat{E}_i = \hat{L}_i + \hat{C}_i + \hat{I}_i + \hat{S}_i$$

และการหาขอบเขตความเชื่อมั่นก็เช่นเดียวกัน คือ

ขีดจำกัดล่าง (Lower Limit) = ผลรวมของขีดจำกัดล่างของแต่ละประเภทในปีเดียวกัน

ขีดจำกัดบน (Upper Limit) = ผลรวมของขีดจำกัดบนของแต่ละประเภทในปีเดียวกัน

(รายละเอียดของการพยากรณ์ได้แสดงไว้ในผนวก ฉ. เช่นกัน)

ผลการพยากรณ์จำนวนพลังงานไฟฟ้าที่ใช้ทั้งหมดในเขตนครหลวงนี้ ได้สรุปไว้ในตารางที่ 5.10.1

9/ Norman Draper and Harry Smith, ibid., p. 121.

ตารางที่ 5.10.1 ผลการพยากรณ์จำนวนพลังงานไฟฟ้าที่ใช้ทั้งหมดในเขตนครหลวง สำหรับปี พ.ศ. 2519 - 2524

หน่วย : ล้านกิโลวัตต์ - ชั่วโมง

ปี พ.ศ.	บ้านอยู่อาศัย	ธุรกิจการค้า	อุตสาหกรรม	ไฟถนนสาธารณะ	รวม
2519	975.5	808.6	4,061.2	40.4	5,885.7
2520	1,088.2	875.1	4,609.7	45.1	6,618.1
2521	1,211.8	945.8	5,196.1	50.2	7,403.8
2522	1,346.4	1,020.8	5,822.2	55.7	8,245.1
2523	1,493.0	1,100.6	6,488.2	61.7	9,143.5
2524	1,652.5	1,185.1	7,194.5	68.3	10,100.4