บทที่ 3

วิธีดำเนินการวิจัย

ในการวิจัยครั้งนี้ต้องการศึกษาเปรียบเทียบ สถิติทดสอบที่ใช้ในการตรวจสอบความเอนเอียงของความคลาดเคลื่อนในการพยากรณ์ 4 วิธีคือ การตรวจสอบแบบผลรวมสะสมอย่างง่าย (Simple Cumulative Sum) การตรวจสอบแบบปรับความคลาดเคลื่อนให้เรียบ (Smoothed error) การตรวจสอบแบบผลรวมสะสมย้อนหลัง (Backward Cumulative Sum) และการตรวจสอบแบบสหสัมพันธ์ในตัวเอง (Autocorrelation) โดยจะศึกษาจำนวนตัวอย่างที่ต้องใช้ตรวจสอบโดยเฉลี่ย (ARL) ในระดับแนวโน้มที่เปลี่ยนแปลง จากน้อยไปมากเป็นเกณฑ์ในการประเมินประสิทธิภาพของตัวสถิติทดสอบต่าง ๆ การวิจัยนี้จะมีลักษณะเป็นการวิจัยเชิงทดลอง ซึ่งจำลองขึ้นด้วยการทำงานของเครื่องคอมพิวเตอร์ โดยเทคนิคมอนติคาร์โล เพื่อหาผลสรุปในการเปรียบเทียบจำนวนตัวอย่างที่ใช้ตรวจสอบความเอนเอียงของความคลาดเคลื่อนในการพยากรณ์ ของสถิติทดสอบ 4 วิธีที่กล่าวมาแล้ว ซึ่งวิธีการดำเนินการวิจัยมีรายละเอียดดังนี้

3.1 การวางแผนการทดลอง

วิธีที่ใช้ตรวจสอบความเอนเอียงของความคลาดเคลื่อนในการพยากรณ์ที่นำมาเปรียบ-เทียบกันในงานวิจัยนี้ จะทำการเปรียบเทียบภายใต้ตัวแบบแนวโน้มเชิงเส้นไม่คงที่ตลอดช่วง ซึ่งมีตัวแบบดังนี้

$$Z_t = \beta_o + \beta_1 t + \beta_2 (t - \ell)\, \delta_t + \varepsilon_t \quad t = 1,2,\ldots \qquad (3.1)$$

$$\delta_t = \begin{cases} 0 & t \leq \ell \\ 1 & t > \ell \end{cases}$$

ซึ่ง t คือ คาบเวลาซึ่งเป็นตัวแปรอิสระ

Z_t คือ ค่าของข้อมูล ณ เวลา t ซึ่งเป็นตัวแปรตาม

β_o , β_1, β_2 คือ พารามิเตอร์ในตัวแบบ

ε_t คือ ค่าผิดพลาดสุ่ม ณ เวลา t ซึ่งมีการแจกแจงแบบปกติมีค่าเฉลี่ยเท่ากับศูนย์ มีค่าความแปรปรวนคงที่

ℓ คือ จำนวนคาบเวลาหรือจำนวนค่าสังเกตเริ่มต้น (Run-in period) ก่อนที่จะมีการเปลี่ยนระดับแนวโน้ม (ความชัน)

การหาค่าพยากรณ์ของตัวแบบนี้ จะทำการหาค่าพยากรณ์ล่วงหน้า 1 คาบเวลา ด้วยวิธีการปรับให้เรียบ แบบเอ็กซ์โพเนนเชียลสองครั้ง (Double Exponential Smoothing) ซึ่งมีวิธีการพยากรณ์ (จากสมการ (2.9) บทที่ 2) ดังนี้

$$\hat{Z}_n(1) = \left(2 + \frac{\alpha}{1-\alpha}\right) S_n^{(1)} - \left(1 + \frac{\alpha}{1-\alpha}\right) S_n^{(2)} \qquad (3.2)$$

โดยที่

$$S_n^{(1)} = \alpha Z_n + (1 - \alpha) S_{n-1}^{(1)}$$

$$S_n^{(2)} = \alpha S_n^{(1)} + (1 - \alpha) S_{n-1}^{(2)}$$

$$S_o^{(1)} = \hat{\beta}_{o,o} - \frac{1-\alpha}{\alpha} \hat{\beta}_{1,0}$$

$$S_o^{(2)} = \hat{\beta}_{o,o} - 2 \frac{1-\alpha}{\alpha} \hat{\beta}_{1,0}$$

$$\hat{\beta}_{1,0} = \frac{12 \sum_{t=1}^{n} \left(t - \frac{n+1}{2}\right) Z_t}{n^3 - n}$$

$$\hat{\beta}_{o,o} = \bar{Z} - \hat{\beta}_{1,0} \left(\frac{n+1}{2}\right)$$

สำหรับการวิจัยครั้งนี้ได้กำหนดค่าพารามิเตอร์ในตัวแบบแนวโน้มเชิงเส้นไม่คงที่ตลอดช่วง $\beta_o = 0, \beta_1 = 1$ ซึ่งการเปลี่ยนแปลงของ β_o , β_1 ตามที่ได้ทดลองแล้วพบว่าไม่มีผลต่อดัชนีที่นักวิจัยใช้เป็นเกณฑ์ในการประเมินประสิทธิภาพของการทดสอบ นอกจากนี้ได้กำหนดจำนวนข้อมูลในคาบเวลาเริ่มต้น (ℓ) เท่ากับ 20 40 60 และ 100 ตามลำดับ

นอกจากจะทำการเปรียบเทียบตัวสถิติต่าง ๆ ภายใต้ตัวแบบแนวโน้มเชิงเส้นไม่คงที่ตลอดช่วงแล้ว การเปรียบเทียบจะกระทำภายใต้สถานะการณ์ต่อไปนี้ด้วย

(1) ค่าความแปรปรวน MAD และ MSE มีค่าคงที่โดยไม่ปรับค่าตามข้อมูลที่เพิ่มขึ้นในช่วงการทดสอบ ซึ่งในกรณีนี้จะศึกษาเปรียบเทียบการทดสอบทั้ง 4 วิธี โดยกำหนดค่าคงที่ที่ใช้ในการปรับให้เรียบของสูตรพยากรณ์ (α) มีค่าเท่ากับ 0.1 0.2 และ 0.3 และค่าคงที่ที่ใช้ในการปรับให้เรียบของตัวสถิติทดสอบ (γ) มีค่าเท่ากับค่าคงที่ α

(2) ค่าความแปรปรวน MAD และ MSE มีค่าไม่คงที่โดยปรับค่าตามข้อมูลที่เพิ่มขึ้นในช่วงการทดสอบ ซึ่งในกรณีนี้จะศึกษาเปรียบเทียบการตรวจสอบ 3 วิธีคือ การตรวจสอบแบบผลรวมสะสมอย่างง่าย การตรวจสอบแบบปรับความคลาดเคลื่อนให้เรียบ และการตรวจสอบแบบสหสัมพันธ์ในตัวเองโดยกำหนดให้ค่าคงที่ α มีค่าเท่ากับ 0.1 0.2 และ 0.3 และค่าคงที่ γ มีค่าเท่ากับค่าคงที่ α นอกจากนี้ยังศึกษาเพิ่มเติมในกรณีที่ค่าคงที่ γ มีค่าน้อยกว่าค่าคงที่ α โดยกำหนดให้ค่าคงที่ γ เท่ากับ 0.05 และ 0.01 สำหรับ α = 0.1, 0.2 ,0.3

3.2 ขั้นตอนการวิจัย

ขั้นตอนหลักในการวิจัยแบ่งเป็น 4 ขั้นตอนคือ

1. การสร้างโปรแกรมย่อย (Subroutines) สำหรับสร้างการแจกแจงของค่าผิดพลาดสุ่ม (ε_t) ตามที่กำหนด

2. สร้างข้อมูล (Z_t ,t) ที่มีความสัมพันธ์กันเป็นไปตามตัวแบบแนวโน้มเชิงเส้นไม่คงที่ตลอดช่วง ตามสมการ(3.1) และหาค่าประมาณของพารามิเตอร์ β_o , β_1 โดยใช้ข้อมูลในคาบเวลาเริ่มต้น (ℓ) ก่อนที่จะมีการเปลี่ยนแปลงระดับแนวโน้ม ซึ่งจะใช้ค่าพารามิเตอร์นี้ในการหาค่าพยากรณ์ตามสมการ (3.2) ต่อไป

3. หาขอบเขตควบคุมของแต่ละตัวสถิติทดสอบภายใต้ตัวแบบแนวโน้มเชิงเส้นไม่คงที่ตลอดช่วง ที่ยังไม่มีการเปลี่ยนแปลงระดับแนวโน้ม ภายใต้ค่า (ℓ, α, γ) ต่าง ๆ ตามที่กล่าวในหัวข้อ 3.1

4. หาจำนวนคาบเวลาที่ต้องใช้ในการตรวจสอบโดยเฉลี่ย (ARL) ของแต่ละตัวสถิติ เมื่อมีการเปลี่ยนแปลงระดับแนวโน้มของข้อมูล (Z_t, t) ที่มีความสัมพันธ์กันตามสมการ (3.1) โดยใช้ขอบเขตควบคุมที่ได้จากข้อ 3 มาเป็นขอบเขตควบคุมที่ใช้ในตัวสถิติ

รายละเอียดสำหรับแต่ละขั้นตอนเป็นดังนี้

3.2.1 <u>การสร้างโปรแกรมย่อย (Subroutines) สำหรับสร้างการแจกแจงของค่าผิดพลาดสุ่ม (ε_t) ตามที่กำหนด</u>

การสร้างลักษณะการแจกแจงของค่าผิดพลาดสุ่ม (ε_t) ตามที่กำหนดไว้ในข้อตกลงเบื้องต้น ให้มีการแจกแจงแบบปกติ จะต้องใช้ตัวเลขสุ่ม (Random number) ซึ่งมีการแจกแจงแบบยูนิฟอร์มในช่วง (0,1) เป็นพื้นฐานในการสร้าง โดยใช้โปรแกรมภาษาฟอร์แทรนโฟร์ (Fortran IV) ที่ใช้กับเครื่อง IBM 370/3031 การสร้างการแจกแจงต่าง ๆ ดังที่กล่าวมาแล้วมีวิธีการสร้างดังนี้

การแจกแจงแบบยูนิฟอร์มในช่วง (0,1)

การผลิตตัวเลขสุ่มที่มีการแจกแจงแบบยูนิฟอร์มในช่วง (0,1) ในการวิจัยครั้งนี้ ใช้วิธี 'Multiplicative Congruential Method' ซึ่งมีตัวผลิต* เป็นดังนี้

$$R_i = aR_{i-1} \pmod{m}$$

---

*เป็นตัวผลิตเดียวกันกับที่ใช้ในโปรแกรมสำเร็จรูป IMSL (International Mathematics and Statistical Libraries) ของบริษัทไอบีเอ็ม

โดยกำหนด

$$a = 7^5 = 16807$$

$$m = 2^{31}$$

R_o = ค่าเริ่มต้นมีค่าอยู่ระหว่าง 0 และ 1

โปรแกรมย่อยที่ใช้สร้างตัวเลขสุ่ม แสดงไว้ในภาคผนวก ก

การแจกแจงแบบปกติ $N(\mu, \sigma^2)$

การผลิตเลขสุ่มที่มีการแจกแจงแบบปกติ โดยวิธีของ Box และ Muller (1958) ซึ่งจะทำการสร้างเลขสุ่มที่มีการแจกแจงแบบปกติมาตรฐาน N (0,1) พร้อมกัน 2 ค่าเป็นอิสระกัน โดยใช้ตัวผลิต (generator) Z_1 และ Z_2

$$Z_1 = (-2 \ln R_1)^{\frac{1}{2}} \cos (2\pi R_2)$$

$$Z_2 = (-2 \ln R_1)^{\frac{1}{2}} \sin (2\pi R_2)$$

ซึ่ง R_1 และ R_2 เป็นตัวเลขสุ่มที่สร้างจากโปรแกรมย่อย FUNCTION RAND (OX) เมื่อได้ตัวเลขสุ่มที่มีการแจกแจงแบบปกติมาตรฐานแล้ว ทำการแปลงค่าเลขสุ่มดังกล่าวโดยอาศัยฟังก์ชัน

$$Z_1' = \mu + \sigma Z_1$$

และ $$Z_2' = \mu + \sigma Z_2$$

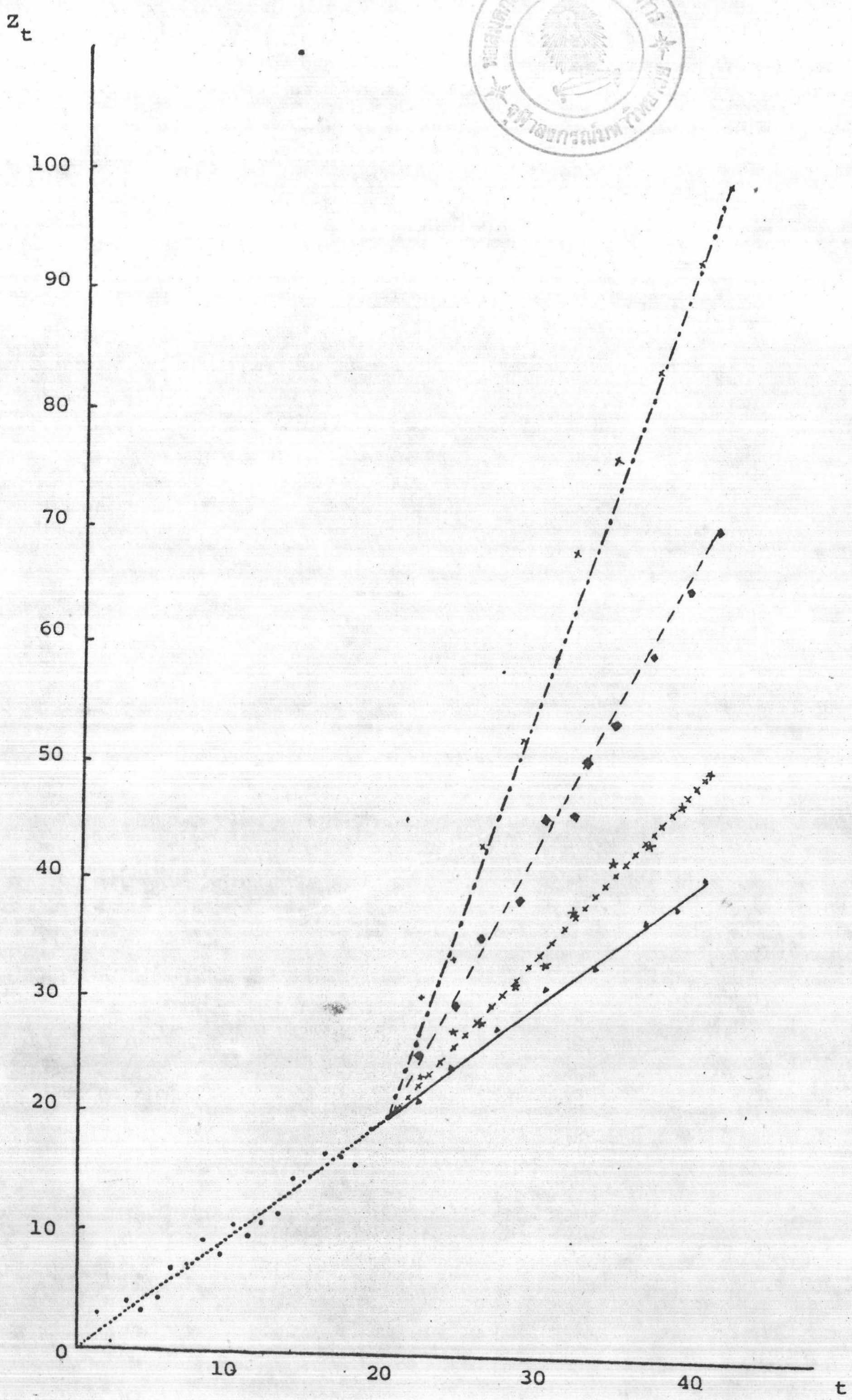
ซึ่งจะได้ว่า Z_1' และ Z_2' มีการแจกแจงเป็นแบบปกติที่มีค่าเฉลี่ย $E(X) = \mu$ และความแปรปรวน $V(X) = \sigma^2$ $(Z_i' \sim N(\mu, \sigma^2) \; ; i = 1,2)$

โปรแกรมย่อยที่ใช้สร้างตัวเลขสุ่มให้มีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับ μ ความแปรปรวนเท่ากับ σ^2 คือ FUNCTION NORMAL (MEAN,SD) ดังแสดงในภาคผนวก ก

3.2.2 สร้างข้อมูล (Z_t , t) ที่มีความสัมพันธ์กันเป็นไปตามตัวแบบแนวโน้มเชิงเส้นไม่คงที่ตลอดช่วง

การสร้างข้อมูล (Z_t , t) ที่มีความสัมพันธ์กันตามตัวแบบแนวโน้มเชิงเส้นไม่คงที่ตลอดช่วง กำหนดให้ $t = 1, 2, \ldots$ แล้วสร้างค่า Z_t ตามรูปแบบความสัมพันธ์ในสมการ (3.1) และ ε_t เป็นค่าผิดพลาดสุ่มที่มีการแจกแจงแบบปกติมีค่าเฉลี่ยเป็นศูนย์ความแปรปรวนเป็น 1 ซึ่งในการสร้างข้อมูลนั้นจะทำการกำหนดระดับแนวโน้มที่เปลี่ยนแปลงไปตามที่ระบุไว้ในข้อตกลงเบื้องต้นของการทำวิจัย และกำหนด $\ell = 20$

รูปที่ 3.1 แสดงให้เห็นตัวอย่างข้อมูลที่สร้างขึ้นตามความสัมพันธ์ในสมการ (3.1)



.... ข้อมูลในช่วง run-in period

——————— ข้อมูลเมื่อเปลี่ยนระดับแนวโน้มไป 0.0σ

+ + + + + + + ข้อมูลเมื่อเปลี่ยนระดับแนวโน้มไป 0.5σ

– – – – – – – ข้อมูลเมื่อเปลี่ยนระดับแนวโน้มไป 1.5σ

–·–·–·–·– ข้อมูลเมื่อเปลี่ยนระดับแนวโน้มไป 3.0σ

รูปที่ 3.1 แสดงรูปแบบความสัมพันธ์ระหว่าง Z_t และ t เมื่อระดับแนวโน้มของตัวแบบเปลี่ยนไป

ในช่วงคาบเวลาเริ่มต้น ℓ จะหาค่าประมาณของ β_o, β_1 ตามสมการ(2.12) และ (2.13) เพื่อใช้ค่า $\hat{\beta}_o$, $\hat{\beta}_1$ ที่ประมาณได้นี้ ในการหาค่าพยากรณ์ตามสมการ (3.2) ต่อไป

3.2.3 <u>การหาขอบเขตควบคุม</u>

การหาขอบเขตควบคุมจะใช้ข้อมูล $(Z_t\ , t)$ ที่สร้างได้ตามข้อ 3.2.2 โดยยังไม่มีการเปลี่ยนระดับแนวโน้ม และค่าพยากรณ์ $(Z_{t-1}(1), t)$ ตามสมการ (3.2) ในช่วงเวลาเดียวกันกับ (Z_t,t) มาหาขอบเขตควบคุมโดยใช้หลักการค้นหาที่เรียกว่า 'Binary Search'

ขั้นที่ 1 สร้างตัวเลขสุ่ม U(0,1)

ขั้นที่ 2 สร้างค่าของ $\varepsilon_t \sim N(0,1)$

ขั้นที่ 3 กำหนดค่าพารามิเตอร์ $\beta_o = 0$, $\beta_1 = 1$, $\beta_2 = 0$

ขั้นที่ 4 กำหนดจำนวนคาบเวลาเริ่มต้น $\ell = 20$

ขั้นที่ 5 กำหนดค่าคงที่ $\alpha = 0.1$, $\gamma = 0.05$

ขั้นที่ 6 สร้างข้อมูล $(Z_t\ , t)$ ให้มีความสัมพันธ์กันในรูป

$$Z_t = \beta_o + \beta_1 t + \beta_2 (t - \ell)\delta_t + \varepsilon_t$$

โดยที่ $t = 1,2,...$

$$\delta_t = \begin{cases} 0 & t \le \ell \\ 1 & t > \ell \end{cases}$$

ขั้นที่ 7 ในช่วงคาบเวลาเริ่มต้น (ℓ) คำนวณค่าประมาณ $\hat{\beta}_o$, $\hat{\beta}_1$ ของ β_o , β_1 ด้วยวิธีกำลังสองน้อยที่สุดแบบสามัญ

ขั้นที่ 8 ใช้ $\hat{\beta}_o$, $\hat{\beta}_1$ เป็นค่าเริ่มต้นในการหาค่าพยากรณ์ตามสมการ (3.2) ซึ่งจะเริ่มทำการหาค่าพยากรณ์ล่วงหน้า 1 คาบเวลาตั้งแต่คาบเวลาที่ ℓ

ขั้นที่ 9 กำหนดขอบเขตควบคุมขึ้นมา 2 ค่า เป็นค่ามากที่สุด H และค่าน้อยที่สุด L ที่คาดว่าจะเป็นไปได้ และใช้ขอบเขตควบคุม L เป็นขอบเขตควบคุมเริ่มต้น C (นั่นคือให้ C = L)

ขั้นที่ 10 เริ่มใช้ตัวสถิติทดสอบตั้งแต่คาบเวลาที่ $\ell + 1$ คำนวณค่าสถิติเปรียบเทียบกับขอบเขตควบคุม C

ขั้นที่ 11 นับจำนวนคาบเวลาที่ต้องใช้ในการตรวจสอบก่อนที่ค่าสถิติจะมีค่ามากกว่าค่า C

ขั้นที่ 12 ทำขั้นที่ 10 และ 11 ซ้ำ 1,000 ครั้งแล้วหาจำนวนคาบเวลาที่ต้องใช้ในการตรวจสอบโดยเฉลี่ย

ขั้นที่ 13 พิจารณาค่า ARL ว่าค่าที่ได้อยู่ในช่วง 99 - 101 หรือไม่ ถ้ามีค่าอยู่ในช่วงนี้จะใช้ค่า C นั้นเป็นขอบเขตควบคุมของตัวสถิติทดสอบที่กำลังพิจารณา ถ้าค่า ARL < 99 ทำในขั้นที่ 14 ต่อไป หรือถ้าค่า ARL > 101 ทำในขั้นที่ 15 ต่อไป

ขั้นที่ 14 ให้ $L = C$

$$C = (L + H)/2$$

และทำในขั้นที่ 16 ต่อไป

ขั้นที่ 15 ให้ $H = C$

$$C = (L + H)/2$$

และทำในขั้นที่ 16 ต่อไป

ขั้นที่ 16 เมื่อได้ค่าขอบเขตควบคุม C แล้ว ย้อนกลับไปทำขั้นตอนที่ 10 - 13 จนกว่าจะได้ขอบเขตควบคุมที่มีค่า ARL อยู่ในช่วงที่ต้องการ

การหาขอบเขตควบคุมของตัวสถิติแต่ละตัวจะกระทำภายใต้ค่าพารามิเตอร์ต่าง ๆ ดังนี้

1. $\ell = 20, 40, 60, 100$

2. $\alpha = 0.1, 0.2$ และ 0.3

3. กรณีที่ค่าความแปรปรวน MAD และ MSE มีค่าคงที่โดยปรับค่าตามข้อมูลที่เพิ่มขึ้นในช่วงการทดลอบ จะกำหนดค่า

   $\gamma = \alpha$

4. กรณีที่ค่าความแปรปรวน MAD และ MSE มีค่าไม่คงที่โดยปรับค่าตามข้อมูลที่เพิ่มขึ้นในช่วงการทดลอบ จะกำหนดค่า $\gamma = 0.05$ 0.1, α เมื่อ $\ell = 20$ และกำหนดค่า $\gamma = 0.05$, α เมื่อ $\ell = 40, 60, 100$

3.2.4 <u>การหาจำนวนคาบเวลาที่ต้องใช้ในการตรวจสอบโดยเฉลี่ย (ARL) เมื่อระดับแนวโน้มของตัวแบบเปลี่ยนแปลงไป</u>

ในขั้นตอนนี้จะนำขอบเขตควบคุมที่ได้จาก 3.2.3 มาใช้ในตัวสถิติทดสอบแต่ละตัว และในแต่ละกรณีที่ศึกษา เพื่อหาค่า ARL เมื่อระดับแนวโน้มของตัวแบบเปลี่ยนแปลงไป โดยมีขั้นตอนต่าง ๆ ดังนี้

ขั้นที่ 1 สร้างตัวเลขสุ่ม $U(0,1)$

ขั้นที่ 2 สร้างค่าของ $\varepsilon_t \sim N(0,1)$

ขั้นที่ 3 กำหนดพารามิเตอร์ $\beta_o = 0$, $\beta_1 = 1$, $\beta_2 = 0.5$

ขั้นที่ 4 กำหนดจำนวนคาบเวลาที่เริ่มต้น $\ell = 20$

ขั้นที่ 5 กำหนดค่าคงที่ $\alpha = 0.1$, $\gamma = 0.05$

ขั้นที่ 6 สร้างข้อมูล (Z_t , t) ให้มีความสัมพันธ์กันในรูป

$$Z_t = \beta_o + \beta_1 t + \beta_2 (t - \ell)\, \delta_t + \varepsilon_t$$

โดยที่ $t = 1,2,\ldots$

$$\delta_t = \begin{cases} 0 & t \leq \ell \\ 1 & t > \ell \end{cases}$$

ขั้นที่ 7 ในช่วงคาบเวลาเริ่มต้น (ℓ) คำนวณค่าประมาณ $\hat{\beta}_o$, $\hat{\beta}_1$ ของ β_o , β_1 ด้วยวิธีกำลังสองน้อยสุดแบบสัมบูรณ์

ขั้นที่ 8 ใช้ $\hat{\beta}_o$, $\hat{\beta}_1$ เป็นค่าเริ่มต้นในการหาค่าพยากรณ์ตามสมการ (3.2) ซึ่งจะเริ่มทำการหาค่าพยากรณ์ล่วงหน้า 1 คาบเวลาตั้งแต่คาบเวลาที่ ℓ

ขั้นที่ 9 เริ่มใช้ตัวสถิติทดสอบตั้งแต่คาบเวลาที่ $\ell + 1$ คำนวณค่าสถิติเปรียบเทียบกับขอบเขตควบคุมที่ได้มาจากขั้นตอนในการหาขอบเขตควบคุม

ขั้นที่ 10 นับจำนวนคาบเวลาที่ต้องใช้ในการตรวจสอบก่อนที่ค่าสถิติจะมีค่ามากกว่าขอบเขตควบคุม

ขั้นที่ 11 ทำขั้นที่ 9 และ 10 ซ้ำ 1,000 ครั้ง แล้วหาจำนวนคาบเวลาที่ต้องใช้ตรวจสอบโดยเฉลี่ย (ARL) สำหรับตัวสถิติที่กำลังพิจารณา

ขั้นที่ 12 เปลี่ยนค่า β_2 เป็น 1.0, 1.5, 2.0, 2.5 และ 3.0 แล้วทำขั้นที่ 6-11 จนกระทั่งครบทุกค่า β_2 ที่เปลี่ยนไป

การหาค่า ARL ของตัวสถิติแต่ละตัวกระทำภายใต้ค่าพารามิเตอร์ต่าง ๆ เช่นเดียวกับการหาขอบเขตควบคุม

เมื่อได้ค่า ARL ของแต่ละระดับแนวโน้มที่เปลี่ยนแปลงไปของแต่ละตัวสถิติทดสอบในแต่ละกรณีที่ศึกษามาแล้ว จะนำมาเปรียบเทียบกันในเชิงสถิติ โดยพิจารณาว่าตัวสถิติใดมีประสิทธิภาพมากที่สุด ตัวสถิตินั้นจะต้องมีค่า ARL ต่ำกว่าตัวสถิติตัวอื่น ๆ