บทที่ 3
วิธีดำเนินการวิจัย

ในการวิจัยครั้งนี้สนใจที่จะศึกษาเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ในตัวแบบอนุกรมเวลากรณีที่มีค่าผิดปกติเกิดขึ้น โดยวิธีที่ทำการศึกษาเปรียบเทียบนั้นมี 4 วิธี คือ

1.วิธีกำลังสองน้อยสุดแบบมีเงื่อนไข (Conditional Least Squares Method : CLS)

2.วิธีประมาณร่วมพารามิเตอร์ตัวแบบและผลกระทบของค่าผิดปกติ (Joint Estimation of Model Parameters and Outliers Effect : JEMPOE)

3.วิธีประมาณแบบเอ็ม (M-Method : M)

4.วิธีกำลังสองน้อยสุดแบบถ่วงน้ำหนักด้วยบูทสแทร็พ (Bootstrap Weighted Least Squares Method : BWLS)

ในการวิจัยครั้งนี้ใช้เทคนิคการจำลองแบบมอนติคาร์โล (Monte Carlo Simulation Method) โดยใช้โปรแกรมภาษา Delphi ในการจำลองซ้ำ 1,000 รอบในแต่ละสถานการณ์การทดลอง ในการเปรียบเทียบวิธีการประมาณจะพิจารณาจากค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ย (Average Mean Square Error : AMSE) ของวิธีประมาณทั้ง 4 วิธี โดยตัวประมาณที่มีค่า AMSE ต่ำที่สุดจะถือว่าวิธีประมาณนั้นมีประสิทธิภาพในการประมาณสูงที่สุดในกลุ่มวิธีประมาณที่ทำการเปรียบเทียบ

ในบทนี้จะกล่าวถึงรายละเอียดเกี่ยวกับวิธีการดำเนินการ ซึ่งมีรายละเอียดในแต่หัวข้อดังต่อไปนี้

3.1 การวางแผนการทดลอง

การวิจัยครั้งนี้มีแผนการทดลองดังนี้

1. ศึกษาในกรณีที่ข้อมูลอนุกรมเวลามีค่าผิดปกติเกิดขึ้น ในตัวแบบดังต่อไปนี้

   1.1 ตัวแบบ AR(1)

   1.2 ตัวแบบ MA(1)

   1.3 ตัวแบบ ARMA(1,1)

2. กำหนดค่าพารามิเตอร์ในแต่ละตัวแบบอนุกรมเวลาที่จะศึกษาดังนี้

   2.1 ตัวแบบ AR(1) กำหนดค่าพารามิเตอร์ ϕ 6 ค่า ดังนี้ -0.9,-0.5,-0.3 , 0.3 , 0.5 และ 0.9

2.2 ตัวแบบ MA(1) กำหนดค่าพารามิเตอร์ θ 6 ค่า ดังนี้ -0.9,-0.5,-0.3 , 0.3 , 0.5 และ 0.9

2.3 ตัวแบบ ARMA (1,1) กำหนดค่าพารามิเตอร์ (ϕ,θ) 5 ชุดดังนี้ (0.2,0.7) (0.8 ,0.3) ,(-0.5,0.4) ,(-0.4,-0.9) และ (0.7,-0.9)

3. ขนาดตัวอย่างมี 6 ระดับ คือ $40,50,60,80,100$ และ 120

4. ค่าเฉลี่ยอนุกรมเวลา $\mu = 100$

5. a_t มีการแจกแจงแบบปกติปลอมปน มีรูปแบบฟังก์ชันคือ

$$(1-P).N(0,10^2)+P.N(0,(C.10)^2)$$

กำหนดองค์ประกอบสเกล C เท่ากับ 3 และ 10 โดย

3 แทน ขนาดของค่าผิดปกติแบบไม่รุนแรง (Mild Outliers) และ

10 แทน ขนาดของค่าผิดปกติแบบรุนแรง (Extreme Outliers)

กำหนดสัดส่วนการปลอมปน P เท่ากับ $0.5, 0.10, 0.15$ และ 0.20

การเปรียบเทียบนั้นจะพิจารณาจากค่าเฉลี่ยของค่าความคลาดเคลื่อนกำลังสองของตัวประมาณพารามิเตอร์ เพื่อที่จะหาวิธีที่ดีที่สุดในแต่ละสถานการณ์ โดยจำนวนสถานการณ์ที่ใช้ในการวิจัยครั้งนี้รวมทั้งสิ้น $1,020$ สถานการณ์

3.2 ขั้นตอนการวิจัย

การวิจัยครั้งนี้มีขั้นตอนการวิจัย ดังต่อไปนี้

1. จำลองค่าความคลาดเคลื่อน a_t โดยจากตัวแบบกำหนดไว้ข้างต้น
2. จำลองข้อมูลอนุกรมเวลา z_t ตามตัวแบบที่กำหนด ซึ่งมีทั้งหมด 3 ตัวแบบ ดังในข้างต้น
3. ประมาณค่าพารามิเตอร์ของตัวแบบอนุกรมเวลา ทั้ง 4 วิธี
4. คำนวณค่าเฉลี่ยค่าความคลาดเคลื่อนกำลังสอง AMSE ของตัวประมาณพารามิเตอร์
5. ทำการเปรียบเทียบค่า AMSE ของพารามิเตอร์แต่ละตัว

3.3 วิธีการจำลองโดยใช้เทคนิคมอนติคาร์โล

เทคนิคที่ใช้แก้ปัญหาในการคำนวณทางสถิตินั้นมีอยู่หลายวิธี วิธีการจำลองโดยใช้เทคนิคมอนติคาร์โลก็เป็นอีกวิธีหนึ่ง ที่นิยมใช้แก้ปัญหากันอย่างแพร่หลาย ซึ่งหลักการจำลองของเทคนิค

ดังกล่าว คือ การใช้เลขสุ่ม (random numbers) มาช่วยในการหาคำตอบของปัญหาที่ต้องการ ขั้นตอนหลัก ๆ ของเทคนิคมอนติคาร์โลสามารถแบ่งได้ออกเป็น 3 ขั้นตอน คือ

1. การสร้างตัวเลขสุ่ม การสร้างตัวเลขสุ่มนั้นเป็นสิ่งที่สำคัญมาก โดยตัวเลขสุ่มที่นำมาใช้นั้นจะมีการแจกแจงแบบเอกรูป ที่มีค่าระหว่าง 0 ถึง 1 ในส่วนของการสร้างนั้นมีผู้คิดค้นหลายวิธี โดยลักษณะของเลขสุ่มที่ดีนั้นควรจะมีลักษณะ ดังนี้ คือ ตัวแปรสุ่มแต่ละตัวนั้นจะต้องเป็นอิสระกัน และมีวัฏจักร ที่ยาว

2. นำตัวเลขสุ่มที่สร้างขึ้นมา นำมาใช้ในการแก้ปัญหาที่กำลังศึกษา ซึ่งบางปัญหานั้นสามารถนำตัวเลขสุ่มมาใช้แก้ปัญหาได้โดยตรง แต่บางปัญหานั้นจะต้องมีขั้นตอนอีกหลายขั้นตอน ถึงจะสามารถแก้ปัญหานั้น ๆ ได้

3. ทดลองกระทำ โดยกระทำในลักษณะซ้ำ ๆ กันหลาย ๆ ครั้งเพื่อหาคำตอบ

3.4 วิธีการสร้างเลขสุ่มและการจำลองข้อมูลที่มีการแจกแจงแบบปกติ

3.4.1 การสร้างเลขสุ่มที่เป็นอิสระกันและมีการแจกแจงเอกรูปในช่วง $(0,1)$ [1]

วิธีการคณิตศาสตร์ในการจำลองเลขสุ่ม(เทียม) มีหลายวิธีการ สำหรับวิธีการที่ได้รับความนิยมใช้กันมากในปัจจุบัน คือ วิธีสมภาค (Congruential Method) ซึ่งมีสูตรหรือตัวแบบหนึ่งที่ใช้กันมาก คือ

$$X_i = (c + aX_{i-1}) \bmod m \ , \quad i = 1,2,3,...$$

โดยที่ค่า c,a และ m เป็นค่าคงที่จำนวนเต็มค่าไม่เป็นลบ และความหมายของตัวแบบคือ X_i เป็นเศษเหลือที่เป็นจำนวนเต็มที่ได้จากการหาร $(c + aX_{i-1})$ ด้วย m นั่นคือ $X_i = c + aX_{i-1} - mk_i$ ซึ่ง $k_i = \left\lfloor \frac{(c + aX_{i-1})}{m} \right\rfloor$ (หมายถึงจำนวนเต็มใหญ่ที่สุดที่น้อยกว่าหรือเท่ากับผลหาร $\frac{(c + aX_{i-1})}{m}$) ดังนั้นค่าเป็นไปได้ของ X_i คือ $0,1,...,m-1$ และก่อนที่จะได้ค่าของ $X_1, X_2,...$ ต้องกำหนดค่าของ c,a,m และ X_0 เราเรียก X_0 ว่า ซีด (seed) หรือ ค่าเริ่มต้น (starting value) จาก X_i ที่ได้จากการคำนวณนำมาหาค่า R_i ซึ่ง

$$R_i = \frac{X_i}{m} \quad \text{เมื่อ } i = 1,2,3,...$$

---

[1] ที่มา : มานพ วราภักดิ์, การจำลองเบื้องต้น (กรุงเทพฯ: ศูนย์ผลิตตำราเรียนสถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ, 2547), หน้า 43

จะได้ R_i มีค่าอยู่ในช่วง $[0,1)$ เรียก $R_1, R_2, \ldots$ ว่า เลขสุ่มเทียม หรือ เลขสุ่มคล้าย

ตัวแบบจำลองสมภาคแบบผลคูณที่ใช้กันมากตัวแบบหนึ่ง ซึ่งได้ผ่านการตรวจสอบคุณสมบัติแล้วหลายประการ คือ กำหนด $c = 0, m = 2^{31} - 1 = 2147483647, a = 7^5 = 16807$ และ X_0 เป็นจำนวนเต็มบวกที่ไม่เกิน m

3.4.2 การจำลองตัวแปรสุ่มปกติด้วยวิธีบอกซ์-มูลเลอร์[2]

George E.P.Box และ Mervin E.Muller ได้คิดค้นวิธีการจำลองตัวแปรสุ่มปกติมาตรฐาน Z_1 และ Z_2 ที่เป็นอิสระกัน ได้ตัวแบบการจำลอง

$$Z_1 = \sqrt{-2\ln R_1}\cos(2\pi R_2)$$
$$Z_2 = \sqrt{-2\ln R_1}\sin(2\pi R_2) \qquad (1)$$

โดยที่ $R_1, R_2 \sim U(0,1)$ และเป็นอิสระกัน (รายละเอียดของการจำลองตัวแปรสุ่มแบบปกติด้วยวิธีบอกซ์-มูลเลอร์แสดงไว้ใน ภาคผนวก-ก.) จากตัวแบบที่ (1) จะเห็นว่าต้องจำลองเลขสุ่มสองตัว ทำให้เราได้ค่าของตัวแปรสุ่ม $N(0,1)$ สองค่าที่เป็นอิสระกัน ซึ่งในทางปฏิบัติสามารถเลือกใช้เฉพาะสูตรใดสูตรหนึ่งก็ได้ เมื่อเราได้เลขสุ่มที่มีการแจกแจงแบบปกติมาตรฐานแล้ว จากนั้นจึงแปลงค่าเลขสุ่มดังกล่าวให้มีการแจกแจงแบบปกติที่มีค่าพารามิเตอร์ μ และ σ ที่ต้องการ โดยใช้สมการ $Normal_1 = \mu + \sigma Z_1$ หรือ $Normal_2 = \mu + \sigma Z_2$ ก็ได้

3.5 การจำลองข้อมูลอนุกรมเวลา

รายระเอียดของแต่ละขั้นตอนมีดังนี้

3.5.1 การจำลองค่าความคลาดเคลื่อน a_t

ในการวิจัยครั้งนี้จะใช้วิธีผลประกอบ (Composition Method) มีรายละเอียดคือ จำลอง u_i เมื่อ u_i มีการแจกแจงแบบ $U(0,1)$ ถ้า u_i มีค่ามากกว่า P ให้จำลอง a_t จาก $N(0,10^2)$ แต่ถ้าน้อยกว่าหรือเท่ากับ P ให้จำลอง a_t จาก $N(0,(C.10)^2)$

3.5.2 การจำลองข้อมูลอนุกรมเวลา

การจำลองข้อมูลอนุกรมเวลาทั้ง 3 ตัวแบบมีรายละเอียดดังต่อไปนี้

---

[2] ที่มา : มานพ วราภักดิ์, การจำลองเบื้องต้น (กรุงเทพฯ: ศูนย์ผลิตตำราเรียนสถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ, 2547), หน้า 142

1). การสร้างข้อมูลอนุกรมเวลา ตัวแบบ AR(1) มีวิธีการคือ

1.1 จำลองค่าเริ่มต้น $z_0 \sim N(\mu, \frac{\sigma_a^2}{1-\phi^2})$

1.2 จำลอง $a_t \sim (1-P).N(0,10^2) + P.N(0,(C.10)^2)$ โดยใช้วิธีการในหัวข้อ 3.5.1

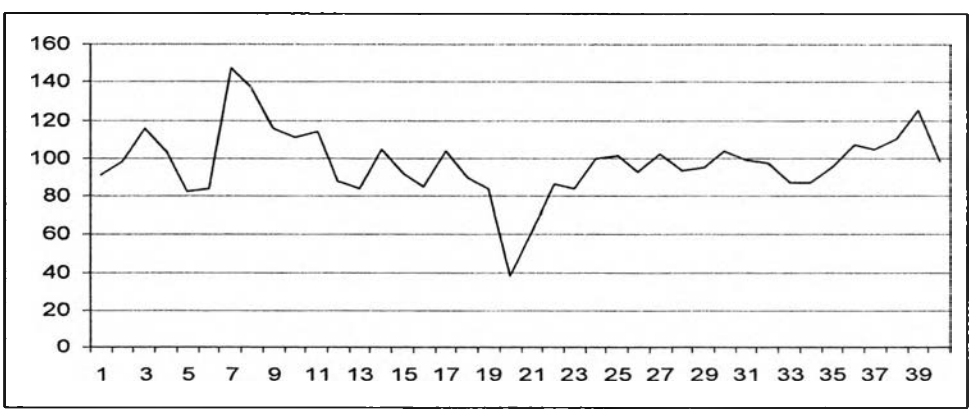
1.3 สร้าง z_t เมื่อ $t=1,\ldots,n$ จากตัวแบบ $z_t = \mu(1-\phi) + \phi z_{t-1} + a_t$

สำหรับการวิจัยครั้งนี้ ในส่วนการการสร้างข้อมูลอนุกรมเวลานั้นจะจำลองข้อมูลทิ้งก่อน 100 ตัวก่อนที่จำลองข้อมูลที่จะนำมาใช้จริง

ตารางที่ 3.1 ตัวอย่างข้อมูลอนุกรมเวลาที่ได้จากการจำลองจากตัวแบบ AR(1) เมื่อ $\phi = 0.5$ $n = 40$ และ $a_t \sim (0.95)N(0,10^2) + (0.05)N(0,(100)^2)$

t	z_t	t	z_t	t	z_t	t	z_t
1	90.9219	11	114.0232	21	61.5708	31	98.8252
2	98.5959	12	87.5358	22	86.0942	32	97.2280
3	115.5659	13	83.6972	23	84.2395	33	86.8198
4	103.0329	14	104.6776	24	99.8304	34	87.4675
5	82.4361	15	91.7280	25	101.6729	35	94.7807
6	83.6142	16	84.8560	26	92.7037	36	106.8277
7	147.1148	17	103.4631	27	102.2497	37	104.7367
8	136.7401	18	89.1346	28	93.3827	38	110.2600
9	115.3304	19	83.7018	29	95.2823	39	125.0164
10	111.1501	20	37.8497	30	103.8866	40	98.1526

จากตัวอย่างข้อมูลอนุกรมเวลาที่ได้จากการจำลองในตารางที่ 3.1 แสดงกราฟได้ดังรูปที่ 3.1



รูปที่ 3.1

2) การสร้างข้อมูลอนุกรมเวลา ตัวแบบ *MA(1)* มีวิธีการคือ

2.1 จำลอง $a_t \sim (1-P).N(0,10^2)+P.N(0,(C.10)^2)$ โดยใช้วิธีการในหัวข้อ 3.5.1

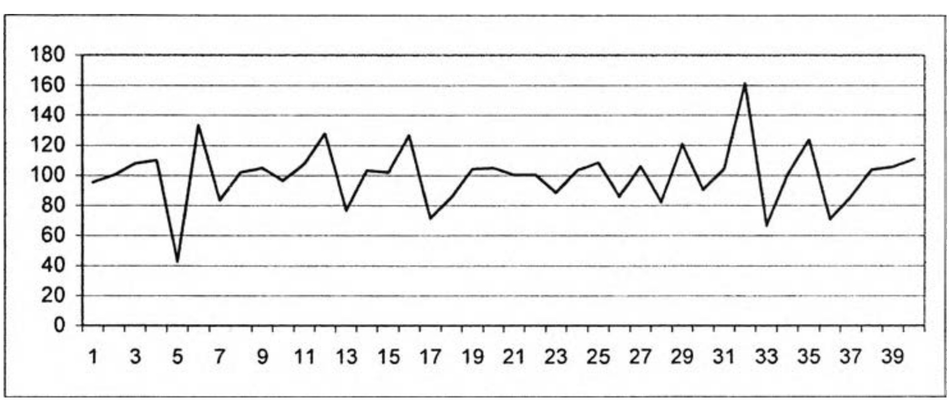
2.2 สร้าง z_t เมื่อ $t=1,\ldots,n$ จากตัวแบบ $z_t = \mu + a_t - \theta a_{t-1}$

สำหรับการวิจัยครั้งนี้ ในส่วนการการสร้างข้อมูลอนุกรมเวลานั้นจะจำลองข้อมูลทิ้งก่อน 100 ตัวก่อนที่จำลองข้อมูลที่จะนำมาใช้จริง

ตารางที่ 3.2 ตัวอย่างข้อมูลอนุกรมเวลาที่ได้จากการจำลองจากตัวแบบ MA(1) เมื่อ $\theta = 0.5$ $n = 40$ และ $a_t \sim (0.95)N(0,10^2)+(0.05)N(0,(100)^2)$

t	z_t	t	z_t	t	z_t	t	z_t
1	95.3055	11	107.5621	21	100.3373	31	104.4960
2	100.1828	12	127.8862	22	100.3361	32	161.5561
3	108.0086	13	76.3726	23	88.5518	33	66.6068
4	110.2785	14	103.2568	24	103.6155	34	100.7441
5	42.3708	15	102.1201	25	108.5816	35	123.4493
6	133.4154	16	126.7460	26	85.9382	36	71.0008
7	83.4166	17	71.4597	27	106.0231	37	85.8820
8	101.9379	18	85.6817	28	81.9856	38	103.7781
9	104.8311	19	104.0455	29	120.5782	39	105.7134
10	96.3178	20	105.0295	30	90.4048	40	110.8645

จากตัวอย่างข้อมูลอนุกรมเวลาที่ได้จากการจำลองในตารางที่ 3.2 แสดงกราฟได้ดังรูปที่ 3.2



รูปที่ 3.2

I22610650

3) การสร้างข้อมูลอนุกรมเวลา ตัวแบบ ARMA(1,1) มีวิธีการคือ

3.1 จำลองค่าเริ่มต้น $z_0 \sim N(\mu, \left[\frac{1-2\phi\theta+\theta^2}{1-\phi^2}\right]\sigma_a^2)$

3.2 จำลอง $a_t \sim (1-P).N(0,10^2) + P.N(0,(C.10)^2)$ โดยใช้วิธีการในหัวข้อ 3.5.1

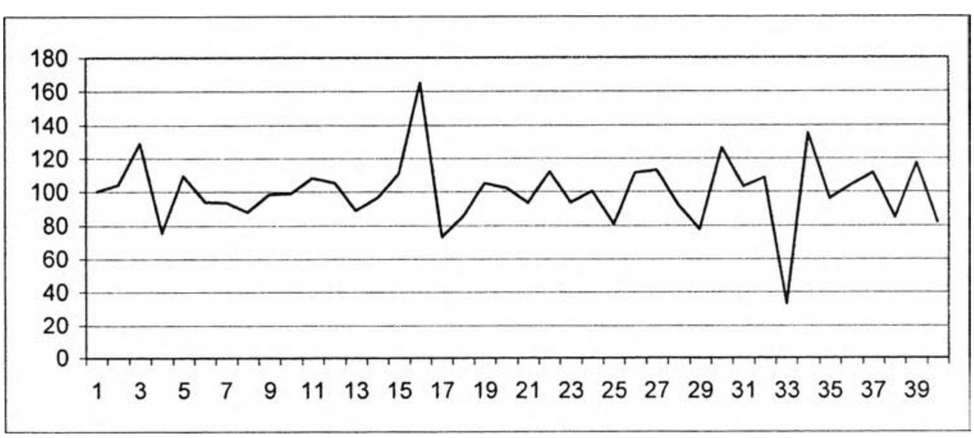
3.3 สร้าง z_t เมื่อ $t=1,\ldots,n$ จากตัวแบบ $z_t = \mu(1-\phi) + \phi z_{t-1} + a_t$

สำหรับการวิจัยครั้งนี้ ในส่วนการการสร้างข้อมูลอนุกรมเวลานั้นจะจำลองข้อมูลทิ้งก่อน 100 ตัวก่อนที่จำลองข้อมูลที่จะนำมาใช้จริง

ตารางที่ 3.3 ตัวอย่างข้อมูลอนุกรมเวลาที่ได้จากการจำลองจากตัวแบบ ARMA(1,1) เมื่อ $\phi = 0.2$ $\theta = 0.7$ และ $n = 40$ และ $a_t \sim (0.95)N(0,10^2) + (0.05)N(0,(100)^2)$

t	z_t	t	z_t	t	z_t	t	z_t
1	100.5757	11	108.2531	21	93.4725	31	103.2709
2	104.1332	12	105.5819	22	112.1878	32	108.7408
3	129.2719	13	89.0981	23	93.6950	33	33.1869
4	75.4671	14	96.6896	24	100.4078	34	135.3104
5	109.6695	15	110.9903	25	80.7318	35	96.0380
6	94.1029	16	165.2168	26	111.4232	36	104.3971
7	93.7364	17	73.3304	27	113.1279	37	111.5610
8	88.0402	18	85.7355	28	92.0181	38	84.4854
9	98.4800	19	105.3026	29	77.5937	39	117.1273
10	99.2944	20	102.1658	30	126.1856	40	81.8797

จากตัวอย่างข้อมูลอนุกรมเวลาที่ได้จากการจำลองในตารางที่ 3.3 แสดงกราฟได้ดังรูปที่ 3.3



รูปที่ 3.3

3.6 การคำนวณค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณพารามิเตอร์

การทดลองในสถานการณ์หนึ่ง ๆ เมื่อได้ค่าพารามิเตอร์ครบทั้ง 4 วิธีแล้ว จะนำค่าที่ประมาณได้มาเปรียบเทียบกับค่าจริง เพื่อคำนวณหาค่าเฉลี่ยของค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณพารามิเตอร์ ตามสูตรคำนวณดังนี้

$$MSE(\hat{\beta}_j) = \sum_{k=1}^{1000} \frac{(\hat{\beta}_{jk} - \beta_j)^2}{1000}$$

และ

$$AMSE = \frac{\sum_{j=1}^{p} MSE(\hat{\beta}_j)}{p}$$

โดยที่ β_j คือ ค่าจริงของพารามิเตอร์ เช่น μ , θ หรือ ϕ

$\hat{\beta}_{jk}$ คือ ค่าประมาณพารามิเตอร์ตัวที่ j ในการทำซ้ำรอบที่ k

k. คือ รอบที่ของการทำซ้ำ

$MSE(\hat{\beta}_i)$ คือ ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ $\hat{\beta}_i$

p. คือ จำนวนพารามิเตอร์ทั้งหมด

จากที่กล่าวมาทั้งหมดขั้นตอนการวิจัยสามารถสรุปเป็นผังขั้นตอนดังรูปที่ 3.4

รูปที่ 3.4 แสดงผังงานการหาค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของค่าประมาณพารามิเตอร์