บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในการตัดสินใจลงทุนในตลาดหลักทรัพย์ ผู้ลงทุนจะรับและประมวลผลข้อมูลจากหลายแหล่ง เพื่อสร้างตัวแบบค่าคาดหวังของผลตอบแทนในอนาคตของหุ้นแต่ละตัว ซึ่งค่าคาดหวังของผลตอบแทนนี้จะถูกใช้เป็นปัจจัยในการตัดสินใจลงทุน ในกรณีที่ข้อมูลทางเศรษฐศาสตร์เพียงชี้ชัดว่าหุ้นในกลุ่มอุตสาหกรรมหนึ่งดีกว่าหุ้นในอีกกลุ่มอุตสาหกรรมหนึ่ง เราอาจใช้ตัวแบบอสมการเชิงเส้นเชิงเบส์ (Bayesian Linear Inequalities Model) เสนอโดย Chiarawongse และคณะ (2009) ในการแก้ปัญหานี้ ตัวแบบดังกล่าวเป็นดังนี้

ถ้าให้ $\underset{\sim}{\alpha}$ เป็นค่าคาดหวังของผลตอบแทน ซึ่งเป็นเวกเตอร์สุ่มของหุ้น k ตัว ให้การแจกแจงก่อน (prior distribution) ของ $\underset{\sim}{\alpha}$ เป็นการแจกแจงแบบปกติหลายตัวแปร ซึ่งมีค่าเฉลี่ย $\underset{\sim}{\Pi}$ และเมทริกซ์ความแปรปรวนร่วม Σ โดยที่ Σ เป็นเมทริกซ์ที่เป็นบวกแน่นอน (positive definite matrix)

$$\underset{\sim}{\alpha} \sim N_k\left(\underset{\sim}{\Pi}, \Sigma\right)$$

และให้ทัศนะของค่าคาดหวังของผลตอบแทน แสดงอยู่ในรูปของระบบอสมการเชิงเส้น

$$A\underset{\sim}{\alpha} \le \underset{\sim}{b}$$

โดยที่ A เป็นเมทริกซ์ขนาด $M \times k$ และ $\underset{\sim}{b}$ เป็นเวกเตอร์ขนาด M

พิจารณาค่าคาดหวังของผลตอบแทนของหุ้น k ตัว $\underset{\sim}{\alpha} = \left(\alpha^{(1)}, \alpha^{(2)}, \ldots, \alpha^{(k)}\right)$ การเรียงอันดับอย่างสมบูรณ์สามารถแสดงได้ด้วยอสมการเชิงเส้น k-1 อสมการ

$$\alpha^{(i_j)} \ge \alpha^{(i_{j+1})}$$

เมื่อ $j = 1,2,\ldots,k-1$ โดยที่ $(i_1, i_2, \ldots, i_k)$ เป็นการเรียงอันดับของ $(1,2,\ldots,k)$

ให้ $S = \left\{\underset{\sim}{\alpha} \in \Re^k \mid A\underset{\sim}{\alpha} \le \underset{\sim}{b}\right\}$ และ I_S เป็นตัวแปรบ่งชี้ (indicator variable) บน S ตัวแบบอสมการเชิงเส้นเชิงเบส์ นิยามโดย การแจกแจงภายหลัง (posterior distribution) ของค่าคาดหวังของผลตอบแทน ดังนี้

$$f\left(\underset{\sim}{\alpha} \mid S\right) \propto I_S \exp\left(-\frac{1}{2}\left[\left(\underset{\sim}{\alpha} - \underset{\sim}{\Pi}\right)^T \Sigma^{-1} \left(\underset{\sim}{\alpha} - \underset{\sim}{\Pi}\right)\right]\right)$$

ซึ่งการแจกแจงภายหลังนี้ เป็นการแจกแจงแบบปกติที่ถูกตัดหาง (truncated normal distribution)

การประมาณค่าคาดหวังของผลตอบแทนในตัวแบบนี้ เป็นดังนี้

$$\hat{\underset{\sim}{\alpha}} = E\left[\underset{\sim}{\alpha} \mid S\right] = \frac{1}{K}\int_{\underset{\sim}{x}\in S} \underset{\sim}{x} \exp\left(-\frac{1}{2}\left(\underset{\sim}{x}-\underset{\sim}{\Pi}\right)^{T}\Sigma^{-1}\left(\underset{\sim}{x}-\underset{\sim}{\Pi}\right)\right) d\underset{\sim}{x}$$

เมื่อ $K = \int_{\underset{\sim}{x}\in S} \exp\left(-\frac{1}{2}\left(\underset{\sim}{x}-\underset{\sim}{\Pi}\right)^{T}\Sigma^{-1}\left(\underset{\sim}{x}-\underset{\sim}{\Pi}\right)\right) d\underset{\sim}{x}$

การอนุมานสถิติเชิงเบส์เพื่อประมาณค่าคาดหวังของผลตอบแทนในตัวแบบดังกล่าวซึ่งเป็นค่าคาดหวังภายใต้การแจกแจงภายหลัง สามารถใช้เทคนิคในกลุ่มลูกโซ่มาร์คอฟมอนติคาร์โล (Markov Chain Monte Carlo หรือ MCMC) ซึ่ง Chiarawongse และคณะ (2009) เลือกใช้วิธี การสุ่มตัวอย่างแบบฮิตแอนด์รัน (Hit-and-Run sampler) ซึ่งเป็นวิธีหนึ่งในกลุ่มเทคนิคดังกล่าว ในการประมาณค่าคาดหวังของผลตอบแทน อย่างไรก็ตามเทคนิคในกลุ่มลูกโซ่มาร์คอฟมอนติคาร์โล มีหลายวิธี และ วิธีที่เป็นที่นิยมอย่างแพร่หลายวิธีหนึ่งได้แก่ การสุ่มตัวอย่างแบบกิบส์ (Gibbs sampler) ซึ่งสามารถใช้แก้ปัญหาดังกล่าวได้เช่นเดียวกัน ดังนั้นเพื่อให้การแก้ปัญหาดังกล่าวเป็นไปอย่างมีประสิทธิภาพ งานวิจัยชิ้นนี้จึงมีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพในการอนุมานสถิติของตัวแบบอสมการเชิงเส้นเชิงเบส์ ระหว่างวิธีการสุ่มตัวอย่างแบบฮิตแอนด์รันและการสุ่มตัวอย่างแบบกิบส์

1.2 วัตถุประสงค์ของการวิจัย

เพื่อทำการเปรียบเทียบความเร็วในการลู่เข้าของลูกโซ่มาร์คอฟมอนติคาร์โลสำหรับการอนุมานเชิงเบส์ เมื่อการแจกแจงภายหลังเป็นแบบปกติที่ถูกตัดหาง ระหว่างการสุ่มตัวอย่างแบบฮิตแอนด์รันกับการสุ่มตัวอย่างแบบกิบส์

1.3 ขอบเขตของการวิจัย

1. เปรียบเทียบเฉพาะการสุ่มตัวอย่างแบบฮิตแอนด์รันกับการสุ่มตัวอย่างแบบกิบส์
2. เกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพ ใช้ 2 เกณฑ์ คือ เปรียบเทียบด้วยค่าครึ่งช่วงความเชื่อมั่นที่คำนวณจากวิธีค่าเฉลี่ยกลุ่ม (batch means) และเปรียบเทียบด้วยค่า MPSRF ของ บรูกซ์-เกลแมน (Brooks–Gelman MPSRF)
3. จำนวนตัวแปร(มิติ)ของ $\underset{\sim}{\alpha}$ ที่ศึกษาเท่ากับ 2 ,3 ,4 ,5 ,7 และ 10 ตัวแปร

4. สำหรับการแจกแจงก่อน (prior distribution) ของ $\underset{\sim}{\alpha}$ ในตัวแบบอสมการเชิงเส้นเชิงเบส์ ซึ่งมีการแจกแจงแบบปกติหลายตัวแปร กำหนดให้ค่าเฉลี่ย $\underset{\sim}{\Pi}$ เท่ากับ $\underset{\sim}{0}$ และ เมทริกซ์ความแปรปรวนร่วม

$$\Sigma = \begin{bmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ \rho & \cdots & \rho & 1 & \rho \\ \rho & \cdots & \rho & \rho & 1 \end{bmatrix}$$ โดยที่ ρ มีค่า 0 ,0.5 และ 0.9

1.4 คำจำกัดความที่ใช้ในการวิจัย

การลู่เข้า (Convergence) หมายถึง ค่าเฉลี่ยของตัวอย่างมีแนวโน้มที่จะลู่เข้าสู่ค่าเฉลี่ยของประชากร เมื่อขนาดตัวอย่างเพิ่มขึ้น นั่นคือ

ให้ $\{X_n; n = 0,1,2,...\}$ เป็นกระบวนการเฟ้นสุ่มที่เป็นลูกโซ่มาร์คอฟเมื่อ X_n เป็นเวกเตอร์ของตัวแปรสุ่ม ซึ่ง $X_n \in S$ โดยที่ S เป็นปริภูมิสถานะ (state space) โดยเงื่อนไขทั่วไปบางเงื่อนไข(some regularity conditions) การแจกแจงของ X_n จะลู่เข้าสู่การแจกแจงคงตัว (stationary distribution) π และเป็นอิสระกับจุดเริ่มต้น x_0 ดังนี้

$$P(X_n \in A \mid X_0 = x_0) \to \pi(A) \text{ สำหรับทุก ๆ } A \in S$$

โดย Markov Chain Strong Law of Large Numbers (MCSLLN) จะได้ว่า

$$\frac{1}{N}\sum_{i=0}^{N-1} g(X_i) \to \pi g \text{ เมื่อ } \pi g = \int_S g(x)\pi(dx) \text{ สำหรับ } N \text{ ที่มีค่ามาก ๆ}$$

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อเป็นแนวทางในการตัดสินใจว่าควรใช้การสุ่มตัวอย่างแบบฮิตแอนด์รันหรือแบบกิบส์ในการจำลองข้อมูล ของการอนุมานเชิงเบส์ เมื่อการแจกแจงภายหลังเป็นแบบปกติที่ถูกตัดหาง

2. เพื่อทราบความเร็วในการลู่เข้า เมื่อหยุดการจำลองด้วยวิธีต่าง ๆ

1.6 วิธีดำเนินการวิจัย

1. ศึกษาค้นคว้าเอกสารและข้อมูลที่เกี่ยวข้องกับการวิจัย
2. จำลองข้อมูลด้วยการสุ่มตัวอย่างแบบฮิตแอนด์รันให้เป็นไปตามขอบเขตที่กำหนด
3. จำลองข้อมูลด้วยการสุ่มตัวอย่างแบบกิบส์ให้เป็นไปตามขอบเขตที่กำหนด
4. คำนวณค่าครึ่งช่วงความเชื่อมั่นจากวิธีค่าเฉลี่ยกลุ่ม
5. คำนวณค่า MPSRF ของ บรูกซ์-เกลแมน
6. วิเคราะห์ และสรุปผลการวิจัย