บทที่ 3

วิธีดำเนินการวิจัย

การศึกษาวิจัยในครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของการสุ่มตัวอย่างด้วยวิธีกิบส์และการสุ่มตัวอย่างด้วยวิธีฮิตแอนด์รันสำหรับทำการวิเคราะห์ปัจจัยเชิงเบส์เมื่อมีปัจจัย 1 ปัจจัย โดยมีวัตถุประสงค์เพื่อการจัดอันดับ ทำการทดลองโดยใช้ข้อมูลความเสี่ยงทางการเมือง-เศรษฐกิจจาก MCMCpack R Package โดย Martin และ Quinn 2004 เพื่อจัดอันดับประเทศ 62 ประเทศตามความเสี่ยงทางการเมือง-เศรษฐกิจ ซึ่งมีดัชนี 5 ตัว คือ

1. ความเป็นอิสระของระบบตุลาการ (courts) เป็นตัวแปรเชิงอันดับที่มี 2 กลุ่ม
2. ส่วนต่างอัตราแลกเปลี่ยนเงินตราต่างประเทศในตลาดมืด (barb2) เป็นตัวแปรแบบต่อเนื่อง
3. ความเสี่ยงจากการยึดทรัพย์สินในกิจการมาเป็นของรัฐ (prsexp2) เป็นตัวแปรเชิงอันดับที่มี 6 กลุ่ม
4. การทุจริตในภาครัฐ (prscorr2) เป็นตัวแปรเชิงอันดับที่มี 6 กลุ่ม
5. ผลิตภัณฑ์มวลรวมภายในประเทศ (gdpw2) เป็นตัวแปรแบบต่อเนื่อง

ในการประมาณค่าพารามิเตอร์ ใช้ขั้นตอนวิธีของเทคนิคของลูกโซ่มาร์คอฟมอนติคาร์โล (Markov Chain Monte Carlo) หรือ MCMC ซึ่งคำนวณมาจากความน่าจะเป็นภายหลังซึ่งมีรูปแบบของความสัมพันธ์กับความน่าจะเป็นก่อนจากในบทที่ 2 ดังนี้

$$p(X^*, \gamma, \Lambda, \phi, \Psi \mid X) \propto p(X \mid X^*, \gamma) p(X^* \mid \Lambda, \phi, \Psi) p(\gamma) p(\Lambda), p(\Phi) p(\Psi)$$

โดย $p(\gamma)p(\Lambda), p(\Phi)p(\Psi)$ เป็นไปตามตัวแบบในบทที่ 2

เมื่อ X เป็นเมทริกซ์ของค่าสังเกต

X^* เป็นเมทริกซ์ของตัวแปรแฝง

γ เป็นเวกเตอร์ของจุดตัด

λ_1 เป็นเวกเตอร์หลักที่ 1 ของ Λ แสดงค่าเฉลี่ยของตัวแปรที่ j

λ_2 เป็นเวกเตอร์หลักที่ 2 ของ Λ แสดงน้ำหนักของปัจจัยของตัวแปรที่ j

ϕ_i เป็นเวกเตอร์หลักที่ 2 ของ Φ แสดงคะแนนของปัจจัยของหน่วยที่ i

ψ_{jj} เป็นความแปรปรวนของค่าคลาดเคลื่อนของตัวแปรที่ j

ในการสุ่มตัวอย่างแบบกิบส์จำเป็นจะต้องใช้ Full Conditional Distribution ซึ่งจากตัวแบบในบทที่ 2 Full Conditional Distribution ของพารามิเตอร์ที่เกี่ยวข้องเป็นดังนี้

สำหรับตัวแปรเชิงอันดับ Full conditional distribution ของ γ_{jc} มีการแจกแจงแบบสม่ำเสมอบนช่วงระหว่าง ค่าต่ำสุดเท่ากับ $\max(\max\{x_{ij}^* : x_{ij} = c\}, \gamma_{j(c-1)})$ และค่าสูงสุดเท่ากับ $\min(\min\{ x_{ij}^* : x_{ij} = c+1\}, \gamma_{j(c+1)})$ โดยให้ $\gamma_{j1} = 0$

กำหนดให้ λ_j เป็นเวกเตอร์แถวที่ j ของ Λ สำหรับ Full conditional distribution ของ x_{ij}^* คือ การแจกแจงแบบปกติตัดปลาย (Truncated Normal distribution) ที่มีค่าเฉลี่ยเท่ากับ $\lambda_j'\phi_i$ และมีความแปรปรวนเท่ากับ 1 โดยจะคำนวณความหนาแน่นเฉพาะในช่วง $(\gamma_{j(x_{ij}-1)}, \gamma_{j(x_{ij})}]$ เมื่อ j เป็นตัวแปรเชิงอันดับ และเป็นความหนาแน่น ณ จุด x_{ij} เมื่อ j เป็นตัวแปรแบบต่อเนื่อง

Full conditional distribution ของ λ_1 คือ การแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับ $\left(L_{0j} + \psi_{jj}^{-1}N\right)^{-1}\left(L_{0j}l_{0j} + \psi_{jj}^{-1}(x_j^* - \lambda_2'\phi_i)\right)$ และมีความแปรปรวนเท่ากับ $\left(L_{0j} + \psi_{jj}^{-1}N\right)^{-1}$ ในส่วนของ λ_2 จะต้องมีสมาชิกอย่างน้อย 1 ตัวที่ถูกกำหนดเครื่องหมาย ดังนั้น Full conditional distribution ของ λ_2 คือ การแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับ $\left(L_{0j} + \psi_{jj}^{-1}\phi_i'\phi_i\right)^{-1}\left(L_{0j}l_{0j} + \psi_{jj}^{-1}\phi_i'(x_j^* - \lambda_1)\right)$ และมีความแปรปรวนเท่ากับ $\left(L_{0j} + \psi_{jj}^{-1}\phi_i'\phi_i\right)^{-1}$ (การแจกแจงปกติแบบตัดปลายสำหรับตัวที่ถูกกำหนดเครื่องหมาย)

Full conditional distribution ของ ϕ_i คือ การแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับ $\left(\mathrm{I} + \lambda_2'\Psi^{-1}\lambda_2\right)^{-1}\left(\lambda_2'\Psi^{-1}\left(x_j^* - \lambda_1\right)\right)$ และมีความแปรปรวนเท่ากับ $\left(\mathrm{I} + \lambda_2'\Psi^{-1}\lambda_2\right)^{-1}$

สำหรับ ψ_{jj} เมื่อ j เป็นตัวแปรแบบต่อเนื่อง Full conditional distribution ของ ψ_{jj} คือ การแจกแจงแกมมาแบบผกผัน ที่มีพารามิเตอร์ $\left(a_{0j} + N\right)/2$ และ $\left(b_{0j} + \left(x_j^* - \Phi\lambda_j\right)'\left(x_j^* - \Phi\lambda_j\right)\right)$ เมื่อ j เป็นตัวแปรเชิงอันดับ ψ_{jj} มีค่าเท่ากับ 1

ใช้ขั้นตอนวิธีของเทคนิค MCMC ได้แก่ การสุ่มตัวอย่างแบบกิบส์ (Gibbs Sampler) และการสุ่มตัวอย่างแบบบฮิตแอนด์รัน (Hit-and-run Sampler) ในการจำลองแล้วใช้วิธีค่าเฉลี่ยกลุ่ม (Batch means method) ประมาณค่าคะแนนเฉลี่ยและค่าคลาดเคลื่อนมาตรฐาน (SE) ซึ่งจะใช้ปันเกณฑ์ในการเปรียบเทียบประสิทธิภาพในการลู่เข้าของตัวประมาณ ในบทนี้จึงเป็นการนำเสนอวิธีดำเนินการวิจัยซึ่งผู้วิจัยได้ดำเนินการตามขั้นตอนต่อไปนี้

1. ดำเนินการด้วยวิธีการสุ่มตัวอย่างแบบกิบส์
2. ดำเนินการด้วยวิธีสุ่มตัวอย่างแบบฮิตแอนด์รัน
3. เปรียบเทียบประสิทธิภาพของการสุ่มตัวอย่างทั้ง 2 วิธี
4. สรุปผลการทดลอง

โดยขั้นตอนที่ 1-2 มีรายละเอียดของการดำเนินการดังนี้

ขั้นตอนในการดำเนินงานวิจัย

1.1 ขั้นตอนดำเนินการด้วยวิธีการสุ่มตัวอย่างแบบกิบส์

1) ทำการสร้างจุดเริ่มต้นของ γ_{jc} , x_{ij}^{*} , ϕ_i , λ_{j1} , λ_{j2} และ ψ_{jj}

2) เริ่มต้นให้จำนวนรอบ $n=1$

3) สุ่ม γ_{jc} สำหรับ j ที่เป็นตัวแปรเชิงอันดับจาก Full conditional distribution ที่มีการแจกแจงแบบสม่ำเสมอ

$$U\left(\max(\max\{x_{ij}^{*}: x_{ij}=c\},\gamma_{j(c-1)}),\min(\min\{x_{ij}^{*}: x_{ij}=c+1\},\gamma_{j(c+1)})\right)$$

4) สุ่ม x_{ij}^{*} สำหรับ j ที่เป็นตัวแปรเชิงอันดับจาก Full conditional distribution $N\left(\lambda_j'\phi_i,1\right)$ โดยจะคำนวณความหนาแน่นเฉพาะในช่วง $(\gamma_{j(x_{ij}-1)},\gamma_{j(x_{ij})}]$ โดยตัวแปร j ที่มี $C_j>2$ จะสุ่ม x_{ij}^{*} ในแต่ละกลุ่มที่ $x_{ij}=c$

5) สุ่ม ϕ_i สำหรับทุก $i=1,\ldots,N$ จาก Full conditional distribution ที่มีการแจกแจงแบบปกติ $N\left(\left(\mathrm{I}+\lambda_2'\Psi^{-1}\lambda_2\right)^{-1}\left(\lambda_2'\Psi^{-1}\left(x_j^{*}-\lambda_1\right)\right),\left(\mathrm{I}+\lambda_2'\Psi^{-1}\lambda_2\right)^{-1}\right)$

6) สุ่ม λ_{j1} สำหรับ j ที่เป็นตัวแปรเชิงอันดับจากFull conditional distribution ที่มีการแจกแจงแบบปกติ $N\left(\left(L_{0j}+\psi_{jj}^{-1}N\right)^{-1}\left(L_{0j}l_{0j}+\psi_{jj}^{-1}(x_j^{*}-\lambda_2'\phi_i)\right),\left(L_{0j}+\psi_{jj}^{-1}N\right)^{-1}\right)$

7) สุ่ม λ_{j2} สำหรับทุก $j=1,\ldots,J$ โดยกำหนดให้ λ_{12} (λ_2 ของตัวแปร courts) มีเครื่องหมายเป็นลบ จาก Full conditional distribution ที่มีการแจกแจงแบบปกติ $N\left(\left(L_{0j}+\psi_{jj}^{-1}\phi_i'\phi_i\right)^{-1}\left(L_{0j}l_{0j}+\psi_{jj}^{-1}\phi_i'(x_j^{*}-\lambda_1)\right),\left(L_{0j}+\psi_{jj}^{-1}\phi_i'\phi_i\right)^{-1}\right)$ (การแจกแจงปกติแบบตัดปลาย สำหรับตัวที่ถูกกำหนดเครื่องหมาย)

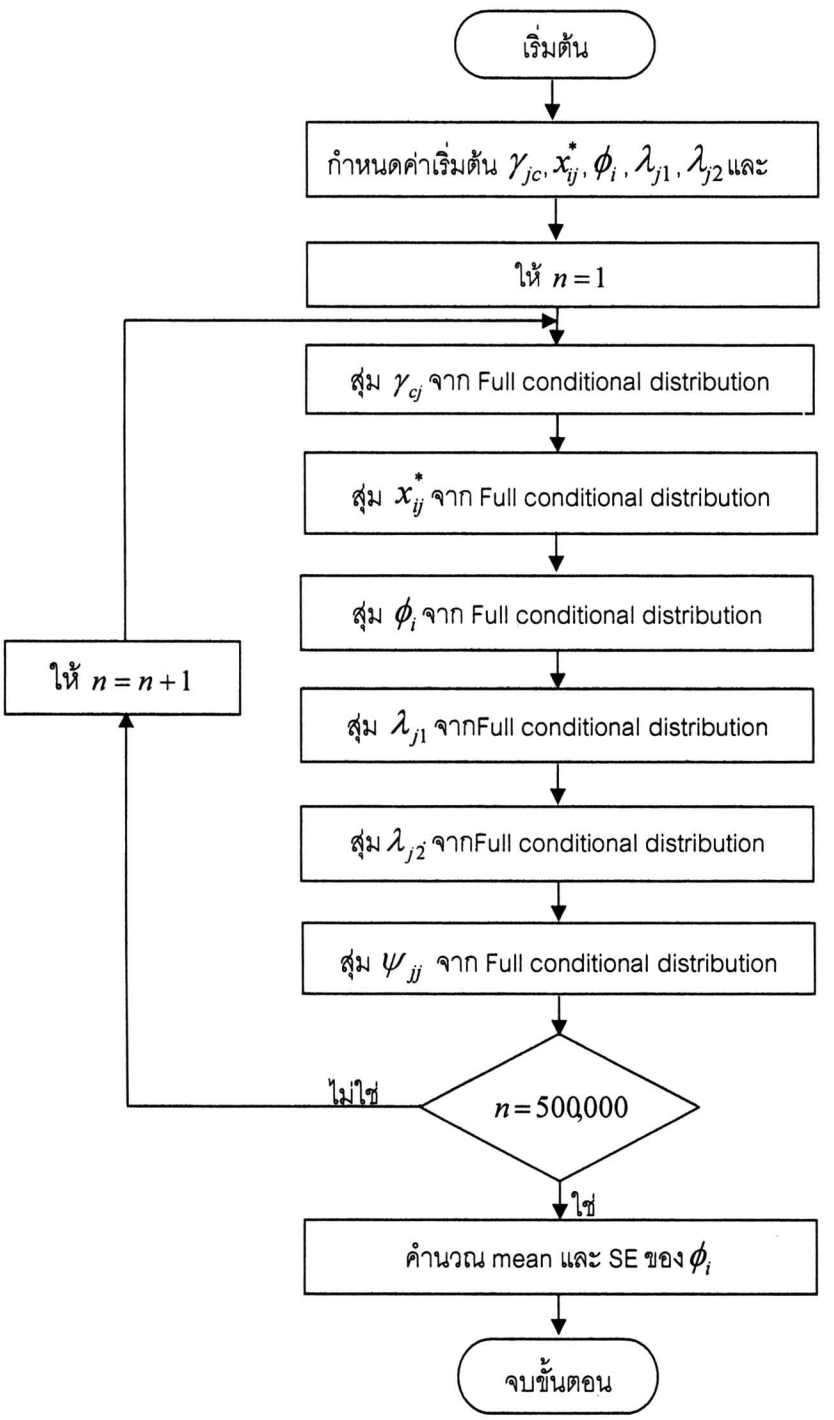
8) สุ่ม ψ_{jj} สำหรับ j ที่เป็นตัวแปรแบบต่อเนื่อง จาก Full conditional distribution ที่มีการแจกแจงแกมมาแบบผกผัน $IG\left(\left(a_{0j}+N\right)/2\ ,\left(b_{0j}+\left(x_j^*-\Phi\lambda_j\right)'\left(x_j^*-\Phi\lambda_j\right)\right)\right)$

9) ให้ $n=n+1$ ดำเนินการซ้ำจนกระทั่ง $n=500{,}000$

10) คำนวณคะแนนเฉลี่ย (mean) และค่าคลาดเคลื่อนมาตรฐาน (SE) ของ ϕ_i ด้วยวิธีค่าเฉลี่ยกลุ่ม โดยให้จำนวน batch เท่ากับ 20

ขั้นตอนดังกล่าวสามารถเขียนอยู่ในรูปที่ 3.1

รูปที่ 3.1 แสดงขั้นตอนการสุ่มตัวอย่างแบบกิบส์



3.2 ขั้นตอนดำเนินการด้วยวิธีการสุ่มตัวอย่างแบบฮิตแอนด์รัน

ในการเปรียบเทียบการสุ่มตัวอย่างแบบกิบส์กับการสุ่มตัวอย่างแบบฮิตแอนด์รันนั้น จะสุ่ม x_{ij}^* ด้วยการสุ่มตัวอย่างแบบฮิตแอนด์รัน เนื่องจาก x_{ij}^* มีการแจกแจงแบบปกติในหลายมิติที่มีขอบเขตจำกัดซึ่งสามารถหาการแจกแจงบนเส้นตรงในทิศทางของฮิตแอนด์รันได้ สำหรับพารามิเตอร์อื่นๆ การหาการแจกแจงบนเส้นตรงในทิศทางของฮิตแอนด์รันนั้นทำได้ยาก จึงยังคงใช้การสุ่มตัวอย่างด้วยวิธีกิบส์

1) ทำการสร้างจุดเริ่มต้นของ $\gamma_{jc}, x_{ij}^*, \phi_i, \lambda_{j1}, \lambda_{j2}$ และ ψ_{jj}

2) เริ่มต้นให้จำนวนรอบ $n=1$

3) สุ่ม γ_{jc} สำหรับ j ที่เป็นตัวแปรเชิงอันดับจาก Full conditional distribution ที่มีการแจกแจงแบบสม่ำเสมอ

$$U\left(\max(\max\{x_{ij}^* : x_{ij}=c\}, \gamma_{j(c-1)}), \min(\min\{x_{ij}^* : x_{ij}=c+1\}, \gamma_{j(c+1)})\right)$$

4) สุ่ม x_{ij}^* ด้วยการสุ่มตัวอย่างแบบฮิตแอนด์รัน สำหรับ j ที่เป็นตัวแปรเชิงอันดับที่มี $C_j > 2$ โดยจะสุ่ม x_{ij}^* ในแต่ละกลุ่มที่ $x_{ij}=c, c=1,2,\ldots,C_j$ พร้อมกัน

5) สุ่ม ϕ_i สำหรับทุก $i=1,\ldots,N$ จาก Full conditional distribution ที่มีการแจกแจงแบบปกติ $N\left(\left(\mathrm{I}+\lambda_2'\Psi^{-1}\lambda_2\right)^{-1}\left(\lambda_2'\Psi^{-1}\left(x_j^*-\lambda_1\right)\right), \left(\mathrm{I}+\lambda_2'\Psi^{-1}\lambda_2\right)^{-1}\right)$

6) สุ่ม λ_{j1} สำหรับ j ที่เป็นตัวแปรเชิงอันดับจากFull conditional distribution ที่มีการแจกแจงแบบปกติ $N\left(\left(L_{0j}+\psi_{jj}^{-1}N\right)^{-1}\left(L_{0j}l_{0j}+\psi_{jj}^{-1}(x_j^*-\lambda_2'\phi_i)\right), \left(L_{0j}+\psi_{jj}^{-1}N\right)^{-1}\right)$

7) สุ่ม λ_{j2} สำหรับทุก $j=1,\ldots,J$ โดยกำหนดให้ λ_{12} (λ_2 ของตัวแปร courts) มีเครื่องหมายเป็นลบ จาก Full conditional distribution ที่มีการแจกแจงแบบปกติ $N\left(\left(L_{0j}+\psi_{jj}^{-1}\phi_i'\phi_i\right)^{-1}\left(L_{0j}l_{0j}+\psi_{jj}^{-1}\phi_i'(x_j^*-\lambda_1)\right), \left(L_{0j}+\psi_{jj}^{-1}\phi_i'\phi_i\right)^{-1}\right)$ (การแจกแจงปกติแบบตัดปลาย สำหรับตัวที่ถูกกำหนดเครื่องหมาย)

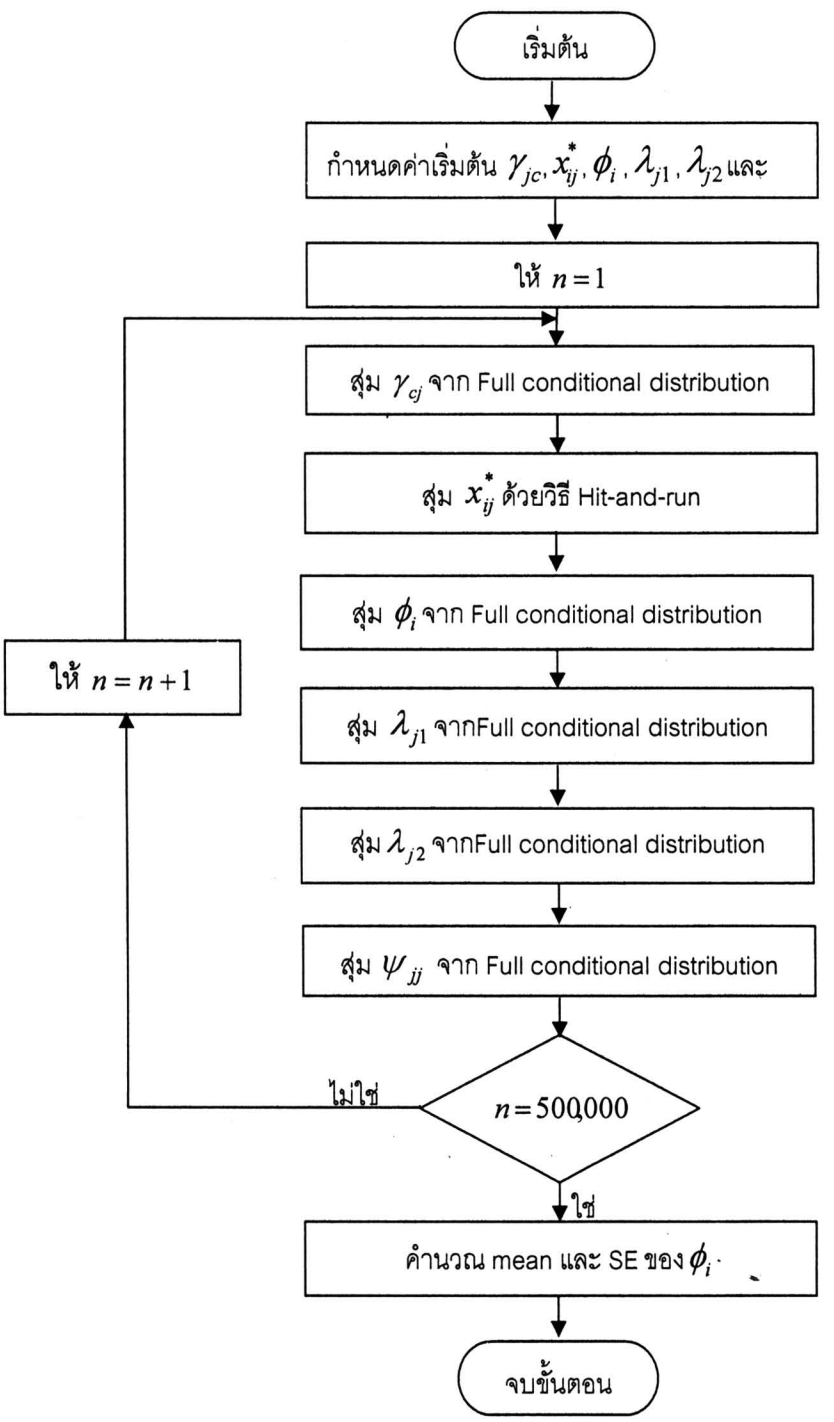
8) สุ่ม ψ_{jj} สำหรับ j ที่เป็นตัวแปรแบบต่อเนื่อง จาก Full conditional distribution ที่มีการแจกแจงแกมมาแบบผกผัน $IG\left(\left(a_{0j}+N\right)/2, \left(b_{0j}+\left(x_j^*-\Phi\lambda_j\right)'\left(x_j^*-\Phi\lambda_j\right)\right)\right)$

9) ให้ $n=n+1$ ดำเนินการซ้ำจนกระทั่ง $n=500{,}000$

10) คำนวณคะแนนเฉลี่ย (mean) และค่าคลาดเคลื่อนมาตรฐาน (SE) ของ ϕ_i ด้วยวิธีค่าเฉลี่ยกลุ่ม โดยให้จำนวน batch เท่ากับ 20

ขั้นตอนดังกล่าวสามารถเขียนอยู่ในรูปที่ 3.2

รูปที่ 3.2 แสดงขั้นตอนการสุ่มตัวอย่างแบบฮิตแอนด์รัน



ในงานวิจัยนี้ใช้โปรแกรม R 2.9.0 ในการประมวลผล ในการจัดอันดับความเสี่ยงทางการเมือง-เศรษฐกิจจะใช้คะแนนเฉลี่ย (Mean) ในการจัดอันดับ การเปรียบเทียบประสิทธิภาพระหว่างการดำเนินการด้วยการสุ่มตัวอย่างแบบฮิตแอนด์รันและการสุ่มตัวอย่างแบบกิบส์ จะใช้ค่าคลาดเคลื่อนมาตรฐานของคะแนนเฉลี่ย (ϕ_i) เป็นเกณฑ์ในการเปรียบเทียบประสิทธิภาพในการลู่เข้าของตัวประมาณ วิธีใดที่มีค่าคลาดเคลื่อนมาตรฐานน้อยกว่า ถือว่าเป็นวิธีที่มีประสิทธิภาพมากกว่า