การอนุมานเชิงสถิติโดยใช้ข้อมูลอันดับที่ไม่สมบูรณ์จากตัวแปรร่วม และการประยุกต์ในการจัดพอร์ตการลงทุน

นางสาวชุติมน สินธุประมา

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ ที่ส่งผ่านทางบัณฑิตวิทยาลัย
The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR)
are the thesis authors' files submitted through the University Graduate School.

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาสถิติ ภาควิชาสถิติ
คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2559
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

STATISTICAL INFERENCE BASED ON IMPERFECT RANKING FROM CONCOMITANT VARIABLES AND ITS APPLICATION IN PORTFOLIO SELECTION

Miss Chutimon Sindhuprama

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Statistics
Department of Statistics
Faculty of Commerce and Accountancy
Chulalongkorn University
Academic Year 2016
Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การอนุมานเชิงสถิติโดยใช้ข้อมูลอันดับที่ไม่สมบูรณ์จากตัวแปรร่วมและการประยุกต์ในการจัดพอร์ตการลงทุน
โดย	นางสาวชุติมน สินธุประมา
สาขาวิชา	สถิติ
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	รองศาสตราจารย์ ดร.เสกสรร เกียรติสุไพบูลย์

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

.....คณบดีคณะพาณิชยศาสตร์และการบัญชี

(รองศาสตราจารย์ ดร.พสุ เดชะรินทร์)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ

(อาจารย์ ดร.อัครินทร์ ไพบูลย์พานิช)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(รองศาสตราจารย์ ดร.เสกสรร เกียรติสุไพบูลย์)

.....กรรมการ

(อาจารย์ ดร.นัท กุลวานิช)

.....กรรมการภายนอกมหาวิทยาลัย

(ศาสตราจารย์ ดร.แอนโทนี เฮย์เตอร์)

ชุติมน สินธุประมา : การอนุมานเชิงสถิติโดยใช้ข้อมูลอันดับที่ไม่สมบูรณ์จากตัวแปรร่วมและการประยุกต์ในการจัดพอร์ตการลงทุน (STATISTICAL INFERENCE BASED ON IMPERFECT RANKING FROM CONCOMITANT VARIABLES AND ITS APPLICATION IN PORTFOLIO SELECTION) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: รศ. ดร.เสกสรร เกียรติสุไพบูลย์, 65 หน้า.

ค่าคาดหวังแบบมีเงื่อนไขของข้อมูลอันดับสามารถนำไปประยุกต์ใช้ในปัญหาเพื่อการตัดสินใจในกรณีที่มีข้อมูลเชิงปริมาณในอดีตกับข้อมูลปัจจุบันที่เป็นข้อมูลอันดับ อย่างไรก็ตามในทางปฏิบัติ ณ เวลาที่ต้องตัดสินใจ ข้อมูลอันดับที่นำมาใช้มักเป็นข้อมูลอันดับที่ไม่สมบูรณ์ งานวิจัยนี้จึงศึกษาผลของการอนุมานเชิงสถิติเมื่อใช้ข้อมูลอันดับที่ไม่สมบูรณ์จากตัวแปรร่วมที่มีความสัมพันธ์กับตัวแปรที่ต้องการศึกษา ผลการศึกษาพบว่า ความถูกต้องซึ่งวัดจากระยะห่างระหว่างค่าจริงกับค่าคาดหวังแบบมีเงื่อนไขเมื่อใช้ข้อมูลอันดับที่ไม่สมบูรณ์จากตัวแปรร่วมจะเพิ่มมากขึ้น หากระดับความสมบูรณ์ของข้อมูลอันดับของตัวแปรร่วมมีค่าสูง และระยะห่างของค่าในเวกเตอร์ค่าเฉลี่ยห่างกันมาก แต่หากจำนวนตัวแปรที่สนใจศึกษาและค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรร่วมมีค่าเพิ่มขึ้น ความถูกต้องดังกล่าวกลับมีค่าลดลง

การจัดพอร์ตการลงทุนเมื่อมีข้อมูลผลตอบแทนในอดีตกับข้อมูลอันดับกลุ่มหลักทรัพย์เป็นตัวอย่างหนึ่งของการประยุกต์ใช้ค่าคาดหวังแบบมีเงื่อนไขของข้อมูลอันดับ ผลการจัดพอร์ตการลงทุนเมื่อจำลองข้อมูลของ 10 หลักทรัพย์ ที่แต่ละหลักทรัพย์มีค่าสัมประสิทธิ์สหสัมพันธ์เป็น 0.5 และ 0.75 และใช้ค่าสัมประสิทธิ์การหลีกเลี่ยงความเสี่ยงเป็น 4 พบว่า เมื่อใช้ค่าคาดหวังแบบมีเงื่อนไขบนข้อมูลอันดับที่ถูกต้องของตัวแปรที่ต้องการศึกษา (Perfect Rank) ได้ค่าอรรถประโยชน์ใกล้เคียงกับเมื่อใช้ข้อมูลผลตอบแทนจริงมากที่สุด ส่วนการใช้ค่าคาดหวังแบบมีเงื่อนไขบนข้อมูลอันดับของตัวแปรร่วม (Imperfect Rank) ได้อรรถประโยชน์สูงกว่าเมื่อข้อมูลของการแจกแจงก่อน และค่าคาดหวังแบบมีเงื่อนไข *โดยนัย*บนข้อมูลอันดับของตัวแปรร่วม (Implied Rank) จะมีประสิทธิภาพเหนือกว่าการใช้ข้อมูลของการแจกแจงก่อนเมื่อตัวแปรร่วมที่มีระดับความสมบูรณ์ของข้อมูลอันดับของตัวแปรร่วมมากกว่า 0.6

กรณีศึกษาเมื่อทดลองจัดพอร์ตการลงทุนโดยใช้ข้อมูลผลตอบแทนรายวันของกลุ่มอุตสาหกรรมในตลาดหลักทรัพย์แห่งประเทศไทยระหว่างเดือนสิงหาคม พ.ศ. 2559 ถึงเดือนมกราคม พ.ศ. 2560 ร่วมกับข้อมูลอันดับที่ได้จากบทวิเคราะห์หลักทรัพย์ พบว่าเมื่อใช้ข้อมูลอันดับที่ได้จากการรวบรวมคำแนะนำให้ซื้อและขายหุ้น แล้วถ่วงน้ำหนักด้วยสัดส่วนมูลค่าของหลักทรัพย์และความแม่นยำของการทำนายในอดีตตามเกณฑ์ที่เหมาะสม สามารถได้พอร์ตการลงทุนที่มีอรรถประโยชน์สะสมสูงกว่าการใช้ข้อมูลจากการแจกแจงก่อนเพียงอย่างเดียวได้

ภาควิชา สถิติ ลายมือชื่อนิสิต .....

สาขาวิชา สถิติ ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

ปีการศึกษา 2559

5881515326 : MAJOR STATISTICS
KEYWORDS: CONCOMITANT VARIABLE / CONDITIONAL EXPECTATION / PORTFOLIO SELECTION

CHUTIMON SINDHUPRAMA: STATISTICAL INFERENCE BASED ON IMPERFECT RANKING FROM CONCOMITANT VARIABLES AND ITS APPLICATION IN PORTFOLIO SELECTION. ADVISOR: ASSOC. PROF. SEKSAN KIATSUPAIBUL, Ph.D., 65 pp.

Rank constrained conditional expectation can be utilized in various decision-making problems, particularly, when quantitative historical data and a valid ranking are available. However, in practice, the correct ranking is hard to obtain. Thus, this paper examines the efficiency of incorporating ranking information derived from concomitant variables into the statistical inference process. Our findings reveal that the accuracy, measured by the distance between the actual values and the conditional expectation based on imperfect rankings, noticeably decreases as the degree of perfect ranking increases and the gaps between the values in mean vector become wider. The number of variables and the correlation coefficient of concomitant variables, on the contrary, have negative effects on the accuracy.

In a simulation study, we apply the concept of rank constrained conditional expectation to portfolio selection problems. We consider a simulated portfolio of 10 assets. Each pair of the asset returns has the correlation coefficient equal to 0.5 and 0.75 and the risk aversion coefficient is set to 4. We explore three types of rankings, perfect rankings, imperfect rankings and implied rankings. The results indicate that the perfect ranking achieves the highest performance while the imperfect ranking leads to substantial utility gain from portfolio selection compared with the prior. For the implied ranking, its performance strongly depends on the degree of perfect ranking. In this case, it outperforms the prior distribution when the degree of perfect ranking is greater than 0.6.

In an empirical study, we collect the data from the Stock Exchange of Thailand during the period of August 2016 to January 2017, then combine the information of the historical returns with the ranking extracted from daily stock picks. We find that leveraging imperfect ranking knowledge can enhance the efficiency of portfolio optimization compared with adopting the prior parameters solely. For this data set, the aggregate stock picks perform well when the recommendations are appropriately weighted based on the proportion of their market capitalization and the quality of the data sources.

Department: Statistics
Field of Study: Statistics
Academic Year: 2016

Student's Signature .....
Advisor's Signature .....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยดีจากการดูแลและให้ความช่วยเหลือของ รศ.ดร. เสกสรร เกียรติสุไพบูลย์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ตั้งแต่รับเป็นอาจารย์ที่ปรึกษา ให้คำแนะนำในการหาหัวข้อวิทยานิพนธ์ คอยแนะนำเอกสารและแหล่งข้อมูลที่เป็นประโยชน์ ช่วยแก้ไขปัญหาต่างๆ ที่เกิดขึ้นระหว่างการดำเนินการศึกษา ตลอดจนช่วยตรวจความเรียบร้อยด้วยความเอาใจใส่ในทุกขั้นตอน ผู้เขียนขอขอบพระคุณอาจารย์เป็นอย่างยิ่ง

ขอขอบพระคุณ อ.ดร.อัครินทร์ ไพบูลย์พานิช ประธานกรรมการสอบวิทยานิพนธ์ อ.ดร. นัท กุลวานิช และ ศ.ดร.แอนโทนี เฮย์เตอร์ กรรมการสอบวิทยานิพนธ์ที่สละเวลาตรวจสอบเนื้อหา ชี้ให้เห็นข้อผิดพลาดต่างๆ รวมทั้งให้คำแนะนำที่มีประโยชน์ ทำให้งานวิจัยชิ้นนี้ถูกต้องสมบูรณ์มากขึ้น

นอกจากนี้ขอขอบพระคุณคณาจารย์ในภาควิชาสถิติทุกท่านที่ให้ความรู้ทั้งด้านวิชาการและถ่ายทอดประสบการณ์ต่างๆ ตลอดระยะเวลาของการศึกษา รวมไปถึงเจ้าหน้าที่ของภาควิชาสถิติที่ช่วยจัดทำเอกสาร และอำนวยความสะดวกในด้านต่างๆ เป็นอย่างดี

สุดท้ายนี้ผู้เขียนขอขอบคุณครอบครัว เพื่อนๆ และน้องๆ ที่ช่วยสนับสนุนและเป็นกำลังใจที่สำคัญมาโดยตลอด

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง
บทคัดย่อภาษาอังกฤษ ..... จ
กิตติกรรมประกาศ ..... ฉ
สารบัญ ..... ช
สารบัญตาราง ..... ญ
สารบัญรูป ..... ฎ
บทที่ 1 บทนำ 1
1.1 ความเป็นมาและความสำคัญของปัญหา 1
1.2 วัตถุประสงค์การวิจัย 2
1.3 สมมติฐานการวิจัย 2
1.4 นิยามและสัญลักษณ์ 2
1.5 ขอบเขตงานวิจัย 4
1.6 ประโยชน์ที่คาดว่าจะได้รับ 6
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 7
2.1 Ranked Set Sampling และตัวแปรร่วม (concomitant variable) 7
2.2 ค่าคาดหวังภายใต้เงื่อนไขของข้อมูลอันดับ 10
2.3 ทฤษฎีการจัดพอร์ตการลงทุน 11
2.4 ความไม่แน่นอนของข้อมูลอันดับ 15
2.5 การวัดประสิทธิภาพของการจัดพอร์ตการลงทุนโดยใช้ Certainty Equivalent Return 16
2.6 CAPM model และเมทริกซ์ความแปรปรวนร่วมตามตัวแบบปัจจัยเดียว (One-factor Model) 17
2.7 การมีส่วนร่วมของกลุ่มคน (Crowd Sourcing) 17

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

หน้า

2.8 การรวบรวมผลการทำนาย (Forecasting Combination) และการรวมข้อมูลอันดับ (Rank Aggregation) 18
บทที่ 3 วิธีการศึกษา 21
3.1 การศึกษาโดยการจำลองข้อมูล 21
3.1.1 ตัวแบบ (Model) 21
3.1.2 เงื่อนไขของตัวแปรที่ต้องการศึกษา 22
3.1.3 การคำนวณค่าคาดหวัง 23
3.1.4 การเปรียบเทียบค่าคาดหวัง 25
3.1.5 การจัดพอร์ตการลงทุนโดยใช้ข้อมูลจำลอง 26
3.2 การศึกษาโดยใช้ข้อมูลจริง 28
3.2.1 ลักษณะและที่มาของข้อมูล 29
3.2.2 ข้อมูลอันดับจากบทวิเคราะห์ 30
3.2.3 กระบวนการจัดอันดับ 32
3.2.4 การคำนวณผลตอบแทนดุลยภาพ (Equilibrium Returns) 35
3.2.5 ขั้นตอนการคำนวณ 35
3.2.6 การใช้ข้อมูลอันดับจากข้อมูลในอนาคต 35
3.2.7 การทดลองโดยใช้วิธีการเคลื่อนตัวอย่างการทดลอง (Rolling Sample) 36
บทที่ 4 ผลการศึกษา 38
4.1 ผลการเปรียบเทียบค่าคาดหวังจากข้อมูลที่จำลองตามเงื่อนไขต่างๆ 38
4.1.1 จำนวนตัวแปร 38
4.1.2 ระดับความสมบูรณ์ของข้อมูลอันดับของตัวแปรร่วม 39
4.1.3 ค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรร่วม 40
4.1.4 ระยะห่างระหว่างค่าในเวกเตอร์ค่าเฉลี่ย 41

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

หน้า

4.2 ผลการจัดพอร์ตการลงทุนจากข้อมูลจำลอง 42
4.3 ผลการจัดพอร์ตการลงทุนจากข้อมูลจริง 43
4.3.1 ผลเมื่อใช้ข้อมูลอันดับที่ถูกต้อง 43
4.3.2 ผลเมื่อใช้ข้อมูลอันดับจากตัวแปรร่วม 44
4.3.3 ผลการจัดพอร์ตการลงทุนสำหรับกลุ่มอุตสาหกรรมย่อย 54
บทที่ 5 บทสรุปและข้อเสนอแนะ 58
5.1 อภิปรายและสรุปผล 58
5.2 ข้อเสนอแนะ 59
รายการอ้างอิง 62
ประวัติผู้เขียนวิทยานิพนธ์ 65

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญตาราง

ตารางที่ 3.1 จำนวนรายการคำแนะนำในช่วง 6 เดือน แยกตามบทวิเคราะห์ 32

ตารางที่ 4.1 ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนที่รวบรวมโดยให้น้ำหนักกับคำแนะนำแต่ละรายการเท่าๆ กัน กับผลตอบแทนของวันทำการถัดไป และผลตอบแทนสิ้นวันนั้นๆ ... 46

ตารางที่ 4.2 ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนที่รวบรวมโดยถ่วงน้ำหนักคำแนะนำด้วยสัดส่วนมูลค่าของหุ้นในกลุ่มอุตสาหกรรมกับกับผลตอบแทนของวันทำการถัดไปและผลตอบแทนสิ้นวันนั้นๆ 48

ตารางที่ 4.3 ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนที่รวบรวมโดยถ่วงน้ำหนักด้วยระดับความรุนแรงของคำแนะนำกับผลตอบแทนสิ้นวัน 49

ตารางที่ 4.4 ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนที่รวบรวมโดยประเมินความแม่นยำของแหล่งข้อมูลย้อนหลัง 10 วันกับผลตอบแทนสิ้นวัน 51

ตารางที่ 4.5 ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนที่รวบรวมโดยประเมินความแม่นยำของแหล่งข้อมูลย้อนหลัง 20 วันกับผลตอบแทนสิ้นวัน 52

ตารางที่ 4.6 ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนที่รวบรวมโดยประเมินความแม่นยำของแหล่งข้อมูลย้อนหลัง 30 วันกับผลตอบแทนสิ้นวัน 53

ตารางที่ 4.7 ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนที่รวบรวมโดยถ่วงน้ำหนักคำแนะนำด้วยสัดส่วนมูลค่าของหุ้นในกลุ่มอุตสาหกรรมย่อยกับผลตอบแทนของวันทำการถัดไปและผลตอบแทนสิ้นวันนั้นๆ 56

ตารางที่ 4.8 จำนวนและสัดส่วนของคำแนะนำจำแนกตามกลุ่มอุตสาหกรรมย่อย 57

สารบัญรูป

รูปที่ 1.1 ความสัมพันธ์ของตัวแปรและข้อมูลอันดับที่ได้จาก Ranking Function 3

รูปที่ 1.2 ขอบเขตงานวิจัยและเนื้อหาที่เกี่ยวข้อง 4

รูปที่ 2.1 หลักการสุ่มตัวอย่างแบบ Ranked Set Sampling 7

รูปที่ 2.2 อัตราผลตอบแทนและความเสี่ยงที่เหมาะสมตามหลักการจัดพอร์ตการลงทุนของ Markowitz 12

รูปที่ 2.3 หลักการจัดพอร์ตการลงทุนตาม Black-Litterman Model 13

รูปที่ 2.4 ผลของระยะห่างของค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานต่อความข้อผิดพลาดของข้อมูลอันดับ 16

รูปที่ 3.1 ขั้นตอนการศึกษาความแตกต่างระหว่างค่าคาดหวังที่มีเงื่อนไขของข้อมูลอันดับของตัวแปรที่สนใจศึกษากับเมื่อใช้ข้อมูลอันดับจากตัวแปรร่วมในเงื่อนไขต่างๆ 26

รูปที่ 3.2 ขั้นตอนการศึกษาประสิทธิภาพของการจัดพอร์ตการลงทุนเมื่อใช้ข้อมูลจริง 28

รูปที่ 3.3 รายละเอียดกลุ่มหลักทรัพย์ตามการแบ่งของตลาดหลักทรัพย์แห่งประเทศไทย 29

รูปที่ 3.4 ที่มาของข้อมูลที่จะนำไปวิเคราะห์ 30

รูปที่ 3.5 ตัวอย่างข้อมูลที่ได้จากบทวิเคราะห์หลักทรัพย์ 31

รูปที่ 3.6 ข้อมูลจริงจากแหล่งต่างๆ ที่นำมาคำนวณหาค่าคาดหวังแบบมีเงื่อนไข กรณีถ่วงน้ำหนักด้วยสัดส่วนมูลค่าหุ้นในกลุ่มอุตสาหกรรม 33

รูปที่ 3.7 ข้อมูลจริงจากแหล่งต่างๆ ที่นำมาประเมินความแม่นยำของคำแนะนำในอดีต และใช้คำนวณหาค่าคาดหวังแบบมีเงื่อนไข 34

รูปที่ 3.8 รูปแสดงการใช้ข้อมูลผลตอบแทนในอดีตและในอนาคตมาคำนวณค่าคาดหวัง 36

รูปที่ 3.9 การทดลองโดยใช้วิธีการเคลื่อนตัวอย่างการทดลอง (Rolling Sample) 37

รูปที่ 4.1 ระยะห่างระหว่างค่าตัวแปรที่สนใจศึกษากับค่าคาดหวังในกรณีต่างๆ เมื่อจำนวนตัวแปรเปลี่ยนแปลงไป 39

รูปที่ 4.2 ระยะห่างระหว่างค่าตัวแปรที่สนใจศึกษากับค่าคาดหวังในกรณีต่างๆ เมื่อระดับความสมบูรณ์ของข้อมูลอันดับของตัวแปรร่วมเปลี่ยนแปลงไป 40

รูปที่ 4.3 ระยะห่างระหว่างค่าตัวแปรที่สนใจศึกษากับค่าคาดหวังในกรณีต่างๆ เมื่อค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรร่วมเปลี่ยนแปลงไป 41

รูปที่ 4.4 ระยะห่างระหว่างค่าของตัวแปรที่สนใจศึกษากับค่าคาดหวังในกรณีต่างๆ เมื่อระยะห่างของค่าในเวคเตอร์ค่าเฉลี่ยเปลี่ยนแปลงไป ณ ระดับความสมบูรณ์ของข้อมูลอันดับของตัวแปรร่วม เท่ากับ 0.5 และจำนวนตัวแปรที่สนใจศึกษามี 10 ตัวแปร..... 42

รูปที่ 4.5 อรรถประโยชน์ที่ได้รับจากการลงทุนในสัดส่วนที่คำนวณได้จากการใช้ค่าผลตอบแทนแบบต่างๆ..... 43

รูปที่ 4.6 อรรถประโยชน์และอรรถประโยชน์สะสมที่ได้รับจากการลงทุนเมื่อใช้ข้อมูลอันดับที่สมบูรณ์เทียบกับการใช้ข้อมูลจากการแจกแจงก่อน 44

รูปที่ 4.7 อรรถประโยชน์และอรรถประโยชน์สะสมที่ได้รับจากการลงทุนเมื่อใช้ข้อมูลอันดับที่รวบรวมโดยให้น้ำหนักกับคำแนะนำแต่ละรายการเท่าๆกัน โดยประเมินประสิทธิภาพจากผลตอบแทนสิ้นวันทำการถัดไป..... 45

รูปที่ 4.8 อรรถประโยชน์และอรรถประโยชน์สะสมที่ได้รับจากการลงทุนเมื่อใช้ข้อมูลอันดับที่รวบรวมโดยให้น้ำหนักกับคำแนะนำแต่ละรายการเท่าๆ กัน โดยประเมินประสิทธิภาพจากผลตอบแทนสิ้นวันทำการนั้นๆ 46

รูปที่ 4.9 อรรถประโยชน์และอรรถประโยชน์สะสมที่ได้รับจากการลงทุนเมื่อใช้ข้อมูลอันดับที่รวบรวมโดยถ่วงน้ำหนักคำแนะนำด้วยสัดส่วนมูลค่าของหุ้นในกลุ่มอุตสาหกรรม..... 47

รูปที่ 4.10 อรรถประโยชน์และอรรถประโยชน์สะสมที่ได้รับจากการลงทุนเมื่อใช้ข้อมูลอันดับที่รวบรวมโดยถ่วงน้ำหนักด้วยระดับความรุนแรงของคำแนะนำ..... 49

รูปที่ 4.11 อรรถประโยชน์และอรรถประโยชน์สะสมที่ได้รับจากการลงทุนเมื่อใช้ข้อมูลอันดับที่รวบรวมโดยประเมินความแม่นยำของแหล่งข้อมูลย้อนหลัง 10 วันทำการ..... 51

รูปที่ 4.12 อรรถประโยชน์และอรรถประโยชน์สะสมที่ได้รับจากการลงทุนเมื่อใช้ข้อมูลอันดับที่รวบรวมโดยประเมินความแม่นยำของแหล่งข้อมูลย้อนหลัง 20 วันทำการ..... 52

รูปที่ 4.13 อรรถประโยชน์และอรรถประโยชน์สะสมที่ได้รับจากการลงทุนเมื่อใช้ข้อมูลอันดับที่รวบรวมโดยประเมินความแม่นยำของแหล่งข้อมูลย้อนหลัง 30 วันทำการ..... 53

รูปที่ 4.14 อรรถประโยชน์และอรรถประโยชน์สะสมที่ได้รับจากการลงทุนในกลุ่มอุตสาหกรรมย่อยเมื่อใช้ข้อมูลอันดับที่สมบูรณ์เทียบกับการใช้ข้อมูลจากการแจกแจงก่อน..... 54

รูปที่ 4.15 อรรถประโยชน์และอรรถประโยชน์สะสมที่ได้รับจากการลงทุนเมื่อใช้ข้อมูลอันดับที่รวบรวมโดยถ่วงน้ำหนักคำแนะนำด้วยสัดส่วนมูลค่าของหุ้นในกลุ่มอุตสาหกรรมย่อย ... 55

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในการวิเคราะห์ทางสถิติส่วนมากมักเป็นการพัฒนาเทคนิคและเครื่องมือสำหรับการวิเคราะห์ข้อมูลเชิงปริมาณ เนื่องจากข้อมูลประเภทนี้ถือว่าเป็นข้อมูลที่สมบูรณ์ สามารถวัด บอกขนาดและเปรียบเทียบความแตกต่างของค่าที่วัดได้ สามารถใช้เครื่องมือพื้นฐานทางคณิตศาสตร์มาคำนวณและจัดการกับข้อมูลได้ง่าย อย่างไรก็ตามในหลายกรณีการเก็บข้อมูลประเภทดังกล่าวยังมีข้อจำกัดบางประการ เช่น เมื่อการวัดค่าของข้อมูลทำได้ยาก ใช้ต้นทุนสูง หรืออาจต้องใช้เวลานานในการเก็บรวบรวมข้อมูล ส่งผลให้ข้อมูลอันดับ (ranking) กลายเป็นข้อมูลอีกประเภทหนึ่งที่มีความสำคัญและนิยมใช้กันมาก เนื่องจากวิธีการได้ข้อมูลไม่ยุ่งยากซับซ้อนเท่าข้อมูลเชิงปริมาณ รวมทั้งต้นทุนและเวลาในการหาข้อมูลดังกล่าวยังต่ำกว่าด้วย ยกตัวอย่างเช่น ในทางการตลาดก็นิยมพิจารณาโดยการจัดลำดับ เช่น อันดับสินค้าขายดี การจัดอันดับสินค้าที่ผู้บริโภคน่าจะสนใจซื้อ ด้านการศึกษาก็มีการจัดอันดับของมหาวิทยาลัยที่ดีของโลก หรือแม้กระทั่งข้อมูลทางการแพทย์ในกรณีที่การตรวจวัดค่าทางวิทยาศาสตร์มีต้นทุนการตรวจสูงหรืออาจต้องรอผลการตรวจเป็นเวลานาน บางครั้งแพทย์จะใช้การประเมินอาการโดยเป็นข้อมูลอันดับเช่นกัน จะเห็นได้ว่าการวิเคราะห์ข้อมูลอันดับร่วมกับข้อมูลเชิงปริมาณเป็นปัญหาที่น่าสนใจและสามารถนำไปประยุกต์ใช้ได้ในหลายสาขาวิชา

อย่างไรก็ตาม ในสถานการณ์จริงการที่จะได้ข้อมูลอันดับที่สมบูรณ์ (perfect rank) เพื่อนำมาวิเคราะห์ร่วมกับข้อมูลเชิงปริมาณ เช่น การอนุมานร่วมกับแจกแจงก่อน (prior) มักเป็นไปได้ยาก จึงเกิดแนวคิดที่ใช้ข้อมูลอันดับของตัวแปรร่วมที่มีความสัมพันธ์กับตัวแปรที่ต้องการศึกษาแทน ในกรณีที่ไม่สามารถวัดค่าของตัวแปรที่สนใจได้โดยตรง ยกตัวอย่างเช่น ในการศึกษาความเป็นพิษของดินในพื้นที่อันตราย การเข้าไปเก็บตัวอย่างและนำข้อมูลมาวิเคราะห์มีความเสี่ยงและใช้ต้นทุนสูง จึงอาศัยการจัดอันดับโดยดูจากสภาพภายนอก เช่น สีของดินแทน แล้วนำมาจัดลำดับ เป็นต้น ซึ่งแนวคิดดังกล่าวมาจากในการศึกษาเรื่อง Ranked Set Sampling โดยจะเรียกตัวแปรที่เกี่ยวข้องและใช้ข้อมูลอันดับของตัวแปรนี้เป็นตัวแทนตัวแปรที่สนใจศึกษา ว่าตัวแปรร่วม (concomitant variable) ซึ่งมีผู้พัฒนาและศึกษาโมเดลของตัวแปรร่วมและตัวแปรที่สนใจศึกษาไว้หลายแบบด้วยกัน

งานวิจัยชิ้นนี้ต้องการศึกษาการหาค่าคาดหวังอย่างมีเงื่อนไขเมื่อทราบการลักษณะของข้อมูลในอดีต และมีข้อมูลอันดับ ณ เวลานั้น (side information) มาประกอบการวิเคราะห์ โดยจะอาศัยโมเดลของตัวแปรร่วมและตัวแปรที่สนใจศึกษาจากงานวิจัยเกี่ยวกับ Ranked Set Sampling มาต่อ

ยอดร่วมกับการหาค่าคาดหวังอย่างมีเงื่อนไขของข้อมูลอันดับ โดยจะศึกษาทั้งในเชิงทฤษฎีว่าค่าคาดหวังเมื่อใช้ข้อมูลอันดับที่ไม่สมบูรณ์จากตัวแปรร่วมจะแตกต่างจากการใช้ข้อมูลอันดับที่สมบูรณ์จากตัวแปรที่สนใจศึกษาอย่างไรบ้าง โดยพิจารณาในกรณีที่จำนวนข้อมูล ลักษณะความสัมพันธ์ของตัวแปรร่วม และระยะห่างของค่าเฉลี่ยแต่ละตัวแปรแตกต่างกัน

นอกจากนี้ เพื่อให้เห็นถึงการนำข้อสรุปเชิงทฤษฎีไปใช้ในสถานการณ์จริง ส่วนที่สองของงานวิจัยนี้จะนำหลักการข้างต้นไปประยุกต์ใช้กับการจัดพอร์ตการลงทุน ตามทฤษฎีของ Markowitz (1952) และ Black and Litterman (1991) ที่กล่าวถึงการหาค่าคาดหวังเมื่อมีข้อมูลเพิ่มเติม (side information) โดยงานของ Chiarawongse et al. (2012) ได้แสดงให้เห็นแล้วว่า การใช้ข้อมูลอันดับทำให้การจัดพอร์ตการลงทุนมีประสิทธิภาพมากขึ้น สำหรับงานวิจัยชิ้นนี้จะศึกษากรณีที่ข้อมูลเพิ่มเติมดังกล่าวเป็นข้อมูลอันดับที่ได้จากตัวแปรร่วมแทน

1.2 วัตถุประสงค์การวิจัย

(1) เพื่อศึกษาความแตกต่างของค่าคาดหวังแบบมีเงื่อนไขเชิงอันดับเมื่อใช้ตัวแปรสนใจศึกษาเปรียบเทียบกับการใช้ตัวแปรร่วม (concomitant variable)

(2) เพื่อประยุกต์ใช้ข้อมูลอันดับของตัวแปรร่วมในการจัดพอร์ตการลงทุน

1.3 สมมติฐานการวิจัย

(1) ความถูกต้องของค่าคาดหวังแบบมีเงื่อนไขของข้อมูลอันดับที่ไม่สมบูรณ์จากตัวแปรร่วม ซึ่งวัดจากระยะห่างระหว่างค่าจริงกับค่าคาดหวังแบบมีเงื่อนไข จะมีค่าเพิ่มขึ้นเมื่อจำนวนตัวแปรที่สนใจศึกษาลดลง ระดับความสมบูรณ์ของข้อมูลอันดับของตัวแปรร่วมมีค่าสูงขึ้น ค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรร่วมมีค่าน้อยลง และระยะห่างของค่าในเวกเตอร์ค่าเฉลี่ยเพิ่มขึ้น

(2) การใช้ข้อมูลอันดับแม้จะเป็นข้อมูลจากตัวแปรร่วมภายใต้เงื่อนไขที่เหมาะสม ก็ยังทำให้ประสิทธิภาพในการจัดพอร์ตการลงทุนเพิ่มมากขึ้นเมื่อเทียบกับการใช้ข้อมูลจากการแจกแจงก่อน (prior) เพียงอย่างเดียว

1.4 นิยามและสัญลักษณ์

เนื่องจากงานวิจัยแต่ละเรื่องมีการใช้สัญลักษณ์ที่แตกต่างกัน เพื่อความเข้าใจง่าย จึงขอกำหนดสัญลักษณ์ที่จะใช้ให้เหมือนกันตลอดทั้งงานวิจัยนี้

Y แทน เวกเตอร์ตัวแปรที่สนใจศึกษา

X แทน เวกเตอร์ตัวแปรร่วม (Concomitant variable)

ε แทน เวกเตอร์ค่าความคลาดเคลื่อน

$R(Y)$ แทน ข้อมูลอันดับของตัวแปรที่สนใจศึกษา (Rank Y)

$R(X)$ แทน ข้อมูลอันดับของตัวแปรร่วม (Rank X)

N แทน จำนวนตัวแปรที่สนใจศึกษา

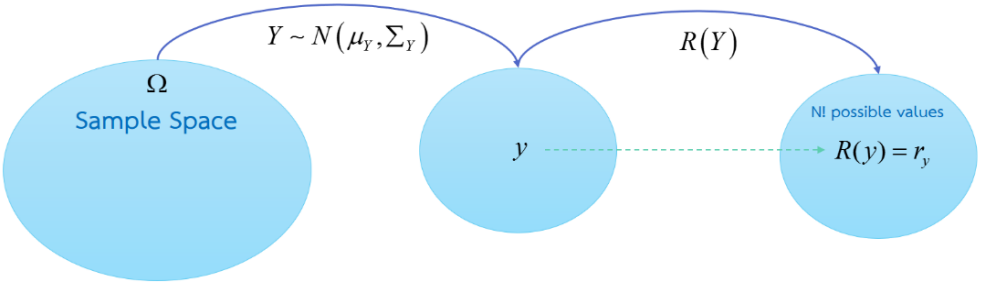
โดย Y, X และ ε เป็นตัวแปรสุ่มที่กำหนดให้มีการแจกแจงแบบปกติหลายตัวแปร (multivariate normal) มีเวกเตอร์ค่าเฉลี่ยขนาด $N \times 1$ และเมทริกซ์ความแปรปรวนและความแปรปรวนร่วมขนาด $N \times N$ ตามสัญลักษณ์ดังต่อไปนี้

$$Y \sim N(\mu_Y, \Sigma_Y)$$
$$X \sim N(\mu_X, \Sigma_X)$$
$$\varepsilon \sim N(\mu_\varepsilon, \Sigma_\varepsilon)$$

และให้ X และ ε เป็นอิสระกัน เว้นแต่จะมีการกำหนดเป็นอย่างอื่น ซึ่งจะระบุไว้ในส่วนนั้นๆ

ในส่วนของข้อมูลอันดับ กำหนดให้ R แทน Ranking function ซึ่งแสดงความสัมพันธ์ (mapping) ระหว่างค่าในเวกเตอร์ของตัวแปรสุ่ม กับลำดับที่เกิดจากการเรียงค่าของตัวแปรสุ่มนั้นจากน้อยไปมาก ตามนิยามของ Savage (1957) ที่ว่า ข้อมูลอันดับของตัวเลขใดๆ จำนวน N ค่า $y_1, y_2, ..., y_N$ สามารถเขียนแทนได้ด้วย $r_y = (r_{y1}, r_{y2}, ..., r_{yN})$ เมื่อ r_{yi} แทนอันดับของ y_i หรืออีกนัยหนึ่ง คือ จำนวนสมาชิกในเวกเตอร์ y ที่มีค่าน้อยกว่าหรือเท่ากับค่า y_i นั้นๆ โดยค่าของเวกเตอร์ r_y มีค่าที่เป็นไปได้ทั้งหมด $N!$ แบบ

แผนภาพตามรูปที่ 1.1 แสดงให้เห็นว่า ตัวแปรสุ่ม Y ถูกสุ่มมากจากปริภูมิตัวอย่าง (Ω) โดยมีการแจกแจงแบบปกติหลายตัวแปร $Y \sim N(\mu_Y, \Sigma_Y)$ และ $R(Y)$ จะเป็นตัวแปรสุ่มแบบ nominal random variable ที่เกิดจากการแปลงค่าของตัวแปรสุ่ม Y ให้เป็นข้อมูลอันดับโดยใช้ ranking function (R) ค่าที่เป็นไปได้ของ $R(Y)$ มีทั้งหมด $N!$ แบบ จากภาพสำหรับเวกเตอร์ของข้อมูล y ซึ่งได้จากตัวแปรสุ่ม Y เมื่อแปลงเป็นข้อมูลอันดับ $R(y)$ จะใช้สัญลักษณ์ r_y ก็คือหนึ่งในค่าที่เป็นได้ทั้งหมด $N!$ ค่าของ $R(Y)$ นั่นเอง



รูปที่ 1.1 ความสัมพันธ์ของตัวแปรและข้อมูลอันดับที่ได้จาก Ranking Function

สำหรับค่าคาดหวังที่จะศึกษา จะให้นิยามและใช้สัญลักษณ์ ดังนี้

$E[Y]$ แทน ค่าคาดหวังของการแจกแจงก่อน (prior)

$E[Y \mid R(Y) = r_y]$ แทน ค่าคาดหวังแบบมีเงื่อนไขบนข้อมูลอันดับที่ถูกต้องของตัวแปรที่ต้องการศึกษา (Perfect Rank)

$E[Y \mid R(X) = r_x]$ แทน ค่าคาดหวังแบบมีเงื่อนไขบนข้อมูลอันดับของตัวแปรร่วม (Imperfect Rank)

$E[Y \mid R(Y) = r_x]$ แทน ค่าคาดหวังแบบมีเงื่อนไข*โดยนัย*บนข้อมูลอันดับของตัวแปรร่วม (Implied Rank)

โดยจะกล่าวถึงรายละเอียด และข้อแตกต่างของค่าคาดหวังแต่ละแบบต่อไปในหัวข้อวิธีการศึกษา

ในการศึกษาส่วนที่สองที่ศึกษาเกี่ยวกับการคำนวณผลตอบแทนและการจัดพอร์ตการลงทุน จะใช้สัญลักษณ์ให้สอดคล้องกัน ดังนี้

Y แทน เวกเตอร์อัตราผลตอบแทนจากการลงทุนในหลักทรัพย์หรือกลุ่มหลักทรัพย์ที่สนใจ

X แทน เวกเตอร์ตัวแปรร่วม (Concomitant variable) เช่น อัตราส่วนทางการเงิน หรือตัวชี้วัดของหุ้นแต่ละตัวที่มีความสัมพันธ์กับอัตราผลตอบแทน สำหรับงานวิจัยนี้จะใช้คะแนน (scoring) ที่รวบรวมได้จากบทวิเคราะห์มาสร้างเป็นข้อมูลอันดับของตัวแปรร่วม

ε แทน เวกเตอร์ค่าความคลาดเคลื่อนของตัวแปรร่วมที่แตกต่างจากอัตราผลตอบแทน

r_y แทน ข้อมูลอันดับของผลตอบแทนจริงของหุ้น (perfect rank)

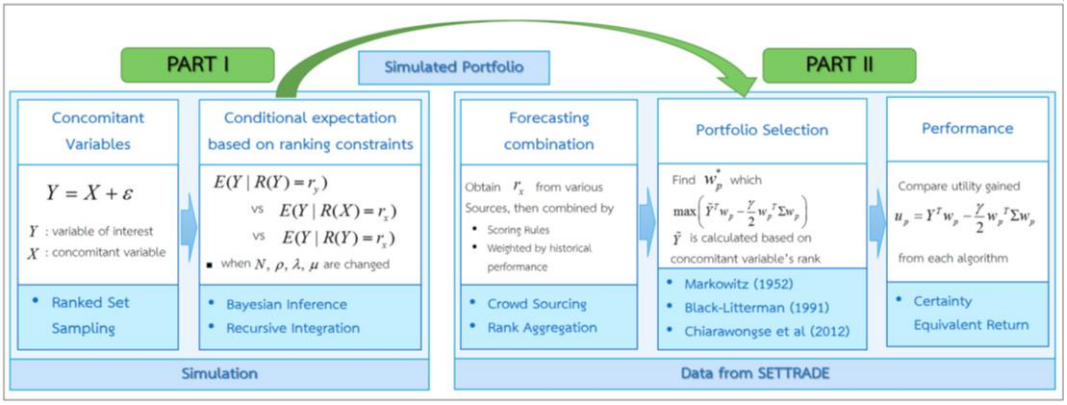
r_x แทน ข้อมูลอันดับของตัวแปรร่วม ในงานวิจัยนี้คืออันดับของหุ้นที่ได้จากบทวิเคราะห์

N แทน จำนวนหุ้นหรือกลุ่มหลักทรัพย์ที่สนใจ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

1.5 ขอบเขตงานวิจัย

เนื้อหางานวิจัยแบ่งออกเป็น 2 ส่วนหลัก แสดงได้ดังรูป



รูปที่ 1.2 ขอบเขตงานวิจัยและเนื้อหาที่เกี่ยวข้อง

ส่วนแรกเป็นการศึกษาเชิงทฤษฎีและการจำลองข้อมูลเพื่อศึกษาความแตกต่างของค่าคาดหวังอย่างมีเงื่อนไขของข้อมูลอันดับ (rank constraints) เมื่อใช้ข้อมูลอันดับของตัวแปรที่สนใจศึกษาเปรียบเทียบกับข้อมูลจากตัวแปรร่วม (concomitant variable) โดยจะศึกษาผลลัพธ์เมื่อจำนวนข้อมูล ระดับความสัมพันธ์ของตัวแปรร่วม ระดับความสมบูรณ์ของข้อมูลอันดับของตัวแปรร่วม และระยะห่างของค่าในเวกเตอร์ค่าเฉลี่ยแตกต่างกัน

งานวิจัยนี้จะใช้การจำลองข้อมูลเพื่อศึกษาการใช้ตัวแปรร่วมในการอนุมานทางสถิติ โดยเริ่มจากการกำหนดโครงสร้างความสัมพันธ์ของตัวแปรที่สนใจศึกษากับตัวแปรร่วม ซึ่งมีผู้ศึกษาและพัฒนาโมเดลไว้หลายรูปแบบเพื่อใช้ในเรื่อง Ranked Set Sampling จากนั้นจะนำรูปแบบความสัมพันธ์ดังกล่าว มาใช้คำนวณหาค่าคาดหวังโดยเปรียบเทียบความแตกต่างของค่าคาดหวังเมื่อใช้ข้อมูลอันดับที่ถูกต้องของตัวแปรที่สนใจศึกษา เทียบกับเมื่อใช้ข้อมูลอันดับของตัวแปรร่วม ในการวิจัยส่วนนี้จะใช้ทฤษฎีการอนุมานแบบเบส์ และคำนวณค่าคาดหวังโดยใช้วิธีปริพันธ์เวียนเกิด (Recursive Integration)

สำหรับเงื่อนไขที่จะศึกษา แบ่งออกเป็นกรณีที่จำนวนข้อมูล (N) เป็น 5, 10 และ 50 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรร่วม (ρ) เป็น 0, 0.5 และ 0.9 และให้ระดับความสมบูรณ์ของข้อมูลอันดับของตัวแปรร่วม (λ) มีค่าเท่ากับ 0.25, 0.5 และ 0.9 จากนั้นเปรียบเทียบค่าคาดหวังใน 3 กรณี โดยเปรียบเทียบระยะห่างซึ่งวัดตามวิธีของ Euclidean Distance ระหว่างค่าของตัวแปรที่สนใจศึกษาที่จำลองได้ในแต่ละรอบ กับค่าคาดหวังแบบมีเงื่อนไขที่คำนวณได้ในกรณีต่างๆ ประกอบด้วย (1) ค่าคาดหวังแบบมีเงื่อนไขบนข้อมูลอันดับที่ถูกต้องของตัวแปรที่ต้องการศึกษา (Perfect Rank) (2) ค่าคาดหวังแบบมีเงื่อนไขบนข้อมูลอันดับของตัวแปรร่วม (Imperfect Rank) และ (3) ค่าคาดหวังแบบมีเงื่อนไขโดยนัยบนข้อมูลอันดับของตัวแปรร่วม (Implied Rank) ซึ่งจะอธิบายรายละเอียดของค่าคาดหวังแต่ละแบบในหัวข้อถัดไป

นอกจากนี้ ยังจำลองข้อมูลเพื่อศึกษาผลเมื่อระยะห่างของค่าเฉลี่ยในเวกเตอร์ค่าเฉลี่ยแตกต่างกัน ซึ่งจะพิจารณาเมื่อจำนวนข้อมูล (N) เป็น 10 และความสมบูรณ์ของข้อมูลอันดับของตัวแปรร่วม (λ) มีค่าในระดับปานกลาง เท่ากับ 0.5 แล้วจึงจำลองข้อมูลให้ค่าในเวกเตอร์ค่าเฉลี่ยมีระยะห่างเท่าๆ กัน โดยกำหนดให้ระยะห่างนี้มีค่าเป็น 0, 0.5, 1, 2 และ 3 ตามลำดับ

สุดท้ายเพื่อให้เข้าใจผลของระดับความสมบูรณ์ของข้อมูลอันดับกับประสิทธิภาพของการจัดพอร์ตการลงทุนก่อนที่จะคำนวณจากข้อมูลจริงในการศึกษาส่วนต่อไป ผู้วิจัยจึงได้จำลองข้อมูลเพิ่มเติม ในกรณีที่จำนวนตัวแปรที่สนใจศึกษา (N) เป็น 10 และใช้ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรร่วม (ρ) เป็น 0.5 และ 0.75 ซึ่งเป็นระดับที่ใกล้เคียงกับค่าจริงในตลาดหลักทรัพย์ จากนั้นจึงทดลองจัดพอร์ตการลงทุน โดยเปรียบเทียบประสิทธิภาพโดยเมื่อใช้ข้อมูลอันดับที่ถูกต้องกับข้อมูลอันดับของตัวแปรร่วม เมื่อระดับความสมบูรณ์ของข้อมูลอันดับมีค่าแตกต่างกัน ผลการจำลองที่ได้จะ

ทำให้ทราบด้วยว่า ระดับความสมบูรณ์ของข้อมูลอันดับของตัวแปรร่วม (λ) ต้องมีค่าเป็นเท่าใด ที่ยังทำให้การจัดพอร์ตการลงทุนมีประสิทธิภาพมากกว่าการใช้ข้อมูลจากการแจกแจงก่อนเพียงอย่างเดียว

ส่วนที่สองของงานวิจัยจะศึกษาผลการจัดพอร์ตการลงทุนเมื่อใช้ข้อมูลจริง ในที่นี้จะใช้ข้อมูลผลตอบแทนในอดีตของกลุ่มอุตสาหกรรมในตลาดหลักทรัพย์แห่งประเทศไทยมาจัดพอร์ตการลงทุนร่วมกับข้อมูลอันดับที่ถูกต้อง (ศึกษาโดยใช้ข้อมูลอันดับในเวลา t+1 มาคำนวณ) และข้อมูลอันดับของตัวแปรร่วมภายใต้เงื่อนไขต่างๆ ซึ่งได้จากข้อมูลอันดับที่รวบรวมได้จากความคิดเห็น หรือการวิเคราะห์ของบริษัทหลักทรัพย์ที่เผยแพร่จริง โดยรวบรวมข้อมูลบทวิเคราะห์หุ้นจากหลายแหล่ง ซึ่งตามหลักการของ crowd sourcing เชื่อว่าจะให้ผลการทำนายที่ถูกต้องมากขึ้น แล้วจึงนำข้อมูลที่ได้จะนำมาปรับให้มีความถูกต้องสมบูรณ์มากขึ้น โดยประยุกต์ใช้เกณฑ์การให้คะแนนแบบต่างๆ

เมื่อได้ข้อมูลอันดับข้างต้นแล้วจะนำข้อมูลดังกล่าวมาทดลองจัดพอร์ตการลงทุน ตามทฤษฎีของ Markowitz (1952) และ Black and Litterman (1991) โดยใช้โมเดลที่คล้ายกับงานศึกษาของ Chiarawongse et al. (2012) จากนั้นจึงวัดประสิทธิภาพของการจัดพอร์ตการลงทุนโดยใช้ค่าอรรถประโยชน์ตามหลักการ certainty equivalent return เพื่อเปรียบเทียบผลที่ได้จากวิธีการต่างๆ

ข้อมูลที่ใช้คำนวณในส่วนนี้จะใช้ข้อมูลจากเว็บไซต์ของตลาดหลักทรัพย์แห่งประเทศไทย เป็นข้อมูลรายวันย้อนหลัง 6 เดือน ตั้งแต่เดือนสิงหาคม พ.ศ. 2559 ถึงเดือนมกราคม พ.ศ. 2560 โดยจะศึกษาหลักทรัพย์เป็นกลุ่มอุตสาหกรรม ซึ่งแบ่งออกเป็น 8 กลุ่มอุตสาหกรรมหลัก และ 28 กลุ่มอุตสาหกรรมย่อย ตามเกณฑ์ของตลาดหลักทรัพย์แห่งประเทศไทย

1.6 ประโยชน์ที่คาดว่าจะได้รับ

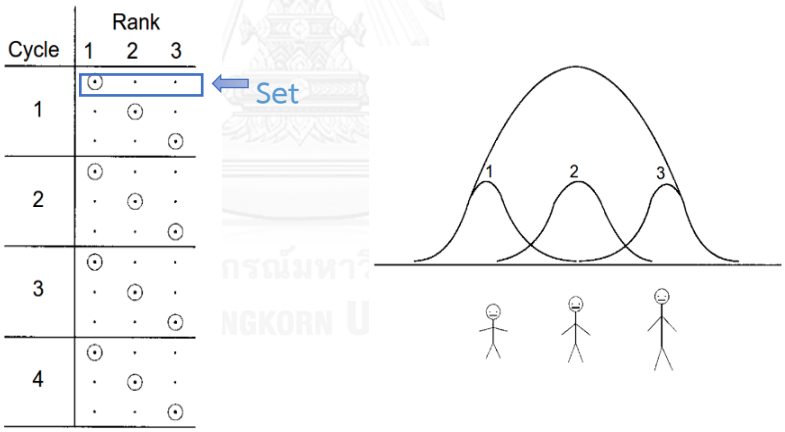
ผลการศึกษาจะแสดงให้เห็นถึงประสิทธิภาพของการใช้ข้อมูลอันดับของตัวแปรร่วมในการประมาณค่าคาดหวังแบบมีเงื่อนไขเชิงอันดับว่าแตกต่างจากการใช้ข้อมูลจากตัวแปรที่สนใจศึกษามากน้อยเพียงใด และมีเงื่อนไขใดบ้างที่มีผลต่อขนาดความแตกต่างดังกล่าว นอกจากนี้ การศึกษาโดยการประยุกต์ใช้ข้อมูลที่มีอยู่ในการจัดพอร์ตการลงทุน ทำให้เห็นข้อแตกต่างระหว่างข้อมูลจริงที่หาได้กับข้อมูลที่สมบูรณ์ตามทฤษฎี รวมทั้งยังได้เห็นถึงกระบวนการแก้ไขปัญหาของข้อมูลจริง เช่น วิธีการจัดการและรวบรวมข้อมูล หรือ การจัดการกับข้อมูลที่ไม่สมบูรณ์ด้วยวิธีการต่างๆ อีกด้วย

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 Ranked Set Sampling และตัวแปรร่วม (concomitant variable)

การสำรวจและเก็บข้อมูลใช้วิธีที่เรียกว่า Ranked Set Sampling คิดขึ้นครั้งแรกโดย McIntyre (1952) จนถูกนำมาใช้อย่างแพร่หลายในงานวิจัยหลายด้าน เช่น ด้านสิ่งแวดล้อม ด้านป่าไม้ และการศึกษามลพิษในเวลาต่อมา Patil et al. (2006) ได้สรุปขั้นตอนการเก็บข้อมูลตามหลัก Ranked Set Sampling ไว้เป็นแผนภาพที่เข้าใจได้ง่ายตามรูปที่ 2.1 ว่า ตามหลัก Ranked Set Sampling จะเริ่มจากการสุ่มตัวอย่างขนาด m หน่วยมาจากประชากร จากนั้นนำชุดตัวอย่าง (set) ที่สุ่มมาดังกล่าวมาเรียงลำดับ (ranking) แล้วจึงวัดค่าของข้อมูลที่มีอันดับต่างๆตามที่กำหนดในแต่ละรอบ โดยในหนึ่งรอบ (cycle) จะมีชุดตัวอย่าง m ชุดเช่นกัน อย่างเช่นกรณี m=3 และสุ่มตัวอย่างทั้งหมด 4 รอบ จะวัดค่าของข้อมูลจริงเฉพาะตำแหน่งที่เป็นสัญลักษณ์ ⊙ ตามแผนภาพด้านล่างซ้าย



รูปที่ 2.1 หลักการสุ่มตัวอย่างแบบ Ranked Set Sampling

ส่วนภาพด้านขวา แสดงถึงลักษณะของประชากรที่ถูกสุ่มมาเป็นตัวอย่าง จะเห็นได้ว่า การสุ่มตัวอย่างด้วยวิธีนี้จะทำให้ได้ตัวอย่างที่ประกอบด้วยตัวแทนของประชากรในแต่ละลักษณะที่ต้องการได้ดีกว่าการสุ่มตัวอย่างอย่างง่าย (Simple Random Sampling - SRS) และในทางทฤษฎีก็ได้รับการพิสูจน์แล้วว่าวิธีการดังกล่าวทำให้ได้ค่าสถิติที่มีประสิทธิภาพมากกว่า โดยดูได้จากค่า Relative Precision (RP) เนื่องจาก ค่าความแปรปรวนของค่าเฉลี่ยของวิธีนี้จะต่ำกว่าแบบ SRS เสมอ

$$1 \le RP = \frac{\text{Variance of sample average with SRS}}{\text{Variance of sample average with Ranked Set Sampling}} \le \frac{m+1}{2}$$

จากหลักการของ Ranked Set Sampling ข้างต้น ในสถานการณ์จริงหลายกรณีที่เราไม่สามารถสังเกตหรือวัดค่าที่แน่นอนของตัวแปรที่สนใจศึกษาได้ แต่สามารถวัดค่าสังเกตหรือจัดลำดับของตัวแปรสุ่มอื่นที่มีความสัมพันธ์กับตัวแปรที่สนใจศึกษาได้ จึงมีผู้คิดวิธีการประยุกต์ใช้ตัวแปรร่วม (concomitant variable) มาช่วยในขั้นตอนการจัดอันดับ โดยตัวแปรร่วมดังกล่าวจะเป็นตัวแปรที่มีความสัมพันธ์กับตัวแปรที่ต้องการศึกษา แต่จะสามารถวัดค่าหรือจัดอันดับได้ง่ายและใช้ต้นทุนต่ำกว่า เช่น เมื่อตัวแปรที่ต้องการศึกษาคือปริมาณสารพิษในดิน (the concentration of volatile organic solvents) ซึ่งวัดค่าได้ยากและมีต้นทุนการเก็บข้อมูลสูง แต่เราทราบว่าค่าดังกล่าวมีความสัมพันธ์เป็นสัดส่วนกับค่า total organic halides (TOX) ซึ่งวัดค่าได้ง่ายและต้นทุนต่ำกว่า จึงจัดอันดับของตัวอย่างตามค่า total organic halides (TOX) ซึ่งถือเป็นตัวแปรร่วมแทน หรือ ในงานศึกษาด้านพฤกษศาสตร์เกี่ยวกับ vegetation concentration on the ground ทำโดยการจัดอันดับโดยใช้ค่าของ intensity of near-infrared electromagnetic radiation แทน เป็นต้น นอกจากนี้ในบางกรณีจะจัดอันดับจากประสบการณ์ของผู้เชี่ยวชาญในด้านนั้นๆ โดยไม่ได้มีการวัดค่าเชิงปริมาณของตัวแปรร่วม

ในงานวิจัยนี้จึงจะอ้างอิงโมเดลความสัมพันธ์ของตัวแปรที่สนใจกับตัวแปรร่วมจากงานวิจัยด้าน Ranked Set Sampling เป็นหลัก เนื่องจากลักษณะของข้อมูลเป็นข้อมูลอันดับเหมือนกัน โดยการใช้ตัวแปรร่วมในการศึกษามีการกำหนดโครงสร้างความสัมพันธ์ไว้แตกต่างกัน ดังต่อไปนี้

- Dell and Clutter (1972) ได้เสนอโมเดล

  $$X_i = Y_i + \varepsilon_i$$

  โดยกำหนดให้ $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ ข้อดีของโมเดลข้างต้น คือ เป็นโมเดลที่ไม่ซับซ้อน แต่ก็มีผู้วิจารณ์ว่าโมเดลดังกล่าวไม่สามารถอธิบายสถานการณ์ที่ซับซ้อนได้ เช่น ผลของความคลาดเคลื่อนเมื่อจำนวนข้อมูลเพิ่มขึ้น หรือ ผลเมื่อข้อมูลมีความสัมพันธ์กันในรูปแบบอื่นๆ

- David and Levine (1972) เสนอโมเดลในเวลาไล่เลี่ยกับ Dell and Cluster โดยโมเดลนี้จะอยู่ในรูปทั่วไปมากขึ้น คือ

  $$X_i = g(Y_i, \varepsilon_i)$$

  อย่างไรก็ตาม ในงานวิจัยดังกล่าวก็ตั้งข้อสมมติอย่างง่ายให้ $X_i = Y_i + \varepsilon_i$ และให้ X_i, ε_i เป็นอิสระจากกัน (mutually independent) และ $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ จะเห็นว่า

ถึงแม้จะปรับรูปแบบให้อยู่ในรูปทั่วไป แต่ด้วยเงื่อนไขและรูปแบบ $g(\cdot)$ ที่ศึกษา ทำให้ผลที่ได้ไม่ได้แตกต่างจากงานของ Dell and Cluster มากนัก

- Stokes (1977) กำหนดให้ concomitant variable ของตัวแปรที่สนใจศึกษา เป็นตัวแปรสุ่มที่สามารถวัดค่าได้อย่างถูกต้องโดยใช้ต้นทุนต่ำหรือไม่มีต้นทุน จะได้ว่า

$$\mathrm{E}[Y \mid X] = \mu_Y + \frac{\rho\sigma_Y}{\sigma_X}(X - \mu_X) \quad \text{และ} \quad \mathrm{var}[Y \mid X] = \sigma_Y^2\left(1-\rho^2\right)$$

โดย Stokes แสดงให้เห็นว่าเมื่อตัวแปรที่สนใจศึกษามีความสัมพันธ์กับตัวแปรร่วมมากขึ้น ความแม่นยำของค่าสถิติที่ได้ก็จะสูงขึ้นด้วย

- Chen et al. (2004) ได้สรุปความสัมพันธ์ของฟังก์ชันความน่าจะเป็นเมื่อใช้ตัวแปรร่วมไว้ดังนี้

$$f_{[r]}(y) = \int f_{Y|X_{(r)}}(y \mid x) g_{(r)}(x)dx$$

เมื่อ $[r]$ แทน ข้อมูลอันดับเมื่อใช้ตัวแปรร่วม

(r) แทน ข้อมูลอันดับของตัวแปรที่สนใจศึกษา

$g_{(r)}(x)$ แทน ฟังก์ชันความน่าจะเป็น (marginal density function) ของ $x_{(r)}$

และจะได้ $f(y) = \frac{1}{k}\sum_{i=1}^{k} f_{[r]}(y)$

จากตัวอย่างข้างต้น จะเห็นว่าการวิเคราะห์และสร้างโมเดลเกี่ยวกับตัวแปรอีกตัวที่เกี่ยวข้องกับตัวแปรที่สนใจศึกษามีมานานแล้ว และวิธีการศึกษาก็จะแตกต่างกันออกไปตามลักษณะของข้อมูลที่ได้จากตัวแปรที่สนใจและค่าสังเกต

นอกจากโมเดลข้างต้นที่แสดงความสัมพันธ์ของตัวแปรที่สนใจศึกษากับตัวแปรร่วมแล้ว ยังมีการอธิบาย Ranking Error แบบอื่นๆ อีก เช่น Bohn and Wolfe (1994) ที่อธิบายความคลาดเคลื่อนของข้อมูลอันดับ โดยการสร้างเมทริกซ์ของความน่าจะเป็นที่อันดับที่ได้จากการสังเกตจะแตกต่างไปจากอันดับที่ถูกต้อง โดยค่า p_{ij} แต่ละตัวในเมทริกซ์จะแสดงถึง โอกาสที่ข้อมูลในตำแหน่ง i ที่ถูกต้องจะถูกจัดไว้ในอันดับที่ j ดังนั้น ในกรณีที่อันดับถูกต้องสมบูรณ์ (perfect rank) จะได้ $p_{ij} = 1$ เมื่อ $i = j$ และ $p_{ij} = 0$ ถ้า $i \neq j$ โดยเรียกเมทริกซ์นี้ว่า doubly stochastic matrix

วิธี Adaptive Ranked Set Sampling หรือ Multi-layer Ranked Set Sampling ตาม Chen and Shen (2003) แต่วิธีการดังกล่าวเน้นสำหรับการประยุกต์ใช้เพื่อการสำรวจตัวอย่าง แล้วคำนวณค่าสถิติและเปรียบเทียบประสิทธิภาพของค่าที่ได้จากการเก็บตัวอย่างแบบ Ranked Set Sampling กับ SRS มากกว่าการอธิบายความสัมพันธ์ของตัวแปรที่สนใจศึกษากับตัวแปรร่วม

งานศึกษาของ Chen et al. (2006) ได้ศึกษาความถูกต้องในการจัดอันดับข้อมูลสำหรับ Ranked Set Sampling เมื่อใช้ตัวแปรร่วมโดยใช้ข้อมูลจริงจาก National Health and Nutrition Examination (NHANES III) ปี 1988-1994 กำหนดให้ตัวแปรที่สนใจศึกษาคือข้อมูลอันดับของค่าดัชนีมวลกาย (Body Mass Index -BMI) โดยทดลองใช้ตัวแปรร่วมหลายตัว เช่น สัดส่วนระหว่างเอวเทียบกับสะโพก ความยาวเส้นรอบวงแขน ความยาวเส้นรอบสะโพก ซึ่งตัวแปรแต่ละตัวมีระดับความสัมพันธ์กับตัวแปรที่สนใจศึกษาแตกต่างกัน จากนั้นทำการสุ่มข้อมูลตัวอย่างตามจำนวนที่ต้องการ และเก็บข้อมูลอันดับแต่ละรอบ สุดท้ายสรุปออกมาเป็นเมทริกซ์ความน่าจะเป็น ผลที่ได้ก็เป็นไปตามที่คาดไว้คือ ยิ่งตัวแปรร่วม หรือ ชุดของตัวแปรร่วม มีความสัมพันธ์กับตัวแปรที่ต้องการศึกษามาก ความน่าจะเป็นที่ข้อมูลอันดับจะตรงกันก็จะมากตามไปด้วย

สำหรับงานวิจัยฉบับนี้ จะศึกษาตัวแปรร่วมที่มีลักษณะเฉพาะ คือเป็นข้อมูลอันดับ ณ เวลาใดเวลาหนึ่ง โดยจะถือว่าเป็นข้อมูล side information เชิงคุณภาพ เช่น ข้อมูลอันดับของผลตอบแทนราคาหุ้น ณ เวลาใดเวลาหนึ่ง จึงจะอ้างอิงความสัมพันธ์ของตัวแปรจากงานศึกษาที่เกี่ยวข้องกับ Ranked Set Sampling เป็นหลัก โดยจะนำมาประกอบกับข้อมูลผลตอบแทนในอดีตและหาค่าคาดหวังอย่างมีเงื่อนไขของผลตอบแทนภายใต้เงื่อนไขของข้อมูลอันดับที่มี

ข้อแตกต่างอีกประการหนึ่ง คือ ในเรื่อง Ranked Set Sampling จะกำหนดให้ตัวแปรที่สนใจและตัวแปรร่วมแต่ละตัวมีการแจกแจงแบบปกติ และสุ่มตัวอย่างจากการแจกแจงดังกล่าวอย่างเป็นอิสระต่อกัน เนื่องจากในบริบทของ Ranked Set Sampling การสุ่มตัวอย่างแต่ละชุดจะเป็นอิสระจากกัน แต่ในงานวิจัยนี้จะนำไปประยุกต์ใช้กับการจัดพอร์ตการลงทุน ซึ่งข้อมูลผลตอบแทนของหุ้นควรพิจารณาให้มีการแจกแจงแบบปกติหลายตัวแปร (multivariate normal) เพราะผลตอบแทนของหุ้นหรือกลุ่มอุตสาหกรรมมักมีความสัมพันธ์กัน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2.2 ค่าคาดหวังภายใต้เงื่อนไขของข้อมูลอันดับ

การคำนวณค่าคาดหวังเมื่อมีเงื่อนไขของข้อมูลอันดับ $\tilde{Y} = E\left(Y \mid R(Y) = r_y\right)$ นิยามโดย

$\tilde{Y} = E\left(Y \mid Y \in \mathbb{R}\right)$ เมื่อ $\mathbb{R} = \left\{y \in \Re^N : Ay \le b\right\}$

การคำนวณค่าดังกล่าวจะซับซ้อน โดยเฉพาะเมื่อ N มีค่ามาก การจำลองข้อมูลและการคำนวณจะทำได้ยากเนื่องจากถูกจำกัดด้วยเงื่อนไขที่มากขึ้น มีผู้ศึกษาวิธีการคำนวณหาค่าคาดหวังโดยมีเงื่อนไขของข้อมูลอันดับดังกล่าวไว้หลายวิธี เช่น Chiarawongse et al. (2012) ใช้วิธีการหาค่าคาดหวังโดยการสุ่มตัวอย่างแบบ Polar Metropolis Hit-and-Run หรืออาจหาค่าคาดหวังโดยใช้วิธีปริพันธ์เวียนบังเกิด (recursive integration) ตาม Hayter (2006) และ Kiatsupaibul et al. (2017) ซึ่งจะใช้วิธี recursive integration ในการคำนวณค่า moments ภายใต้เงื่อนไขของข้อมูลที่มี เช่น เมื่อทราบ

ข้อมูลอันดับ โดยวิธีดังกล่าว สามารถใช้ได้ทั้งในกรณีที่ตัวแปรที่สนใจเป็นอิสระต่อกัน และประยุกต์ใช้ในกรณีที่ตัวแปรที่สนใจมีความสัมพันธ์กัน โดยมีการแจกแจงร่วมเป็นแบบปกติหลายตัวแปรได้ด้วย

Kiatsupaibul et al. (2017) ยังได้ศึกษาผลของค่าคาดหวังเมื่อทราบข้อมูลอันดับไว้ว่า เมื่อเปรียบเทียบระหว่างการใช้ข้อมูลอันดับที่ใช้สอดคล้องกับการแจกแจง prior ในลักษณะที่เรียกว่าเป็น reinforcing ranking นั้น กับข้อมูลอันดับที่ตรงข้ามกับการแจกแจง prior ที่เรียกว่า opposing ranking ในกรณีที่ $N = 5$ พบว่า การใช้ opposing ranking จะทำให้ค่าคาดหวังของ posterior ที่ได้นั้น มีอันดับแตกต่างจากค่าคาดหวังเดิม โดยจะปรับอันดับใหม่ให้สอดคล้องกับข้อมูลอันดับที่เพิ่มเข้ามา และค่าส่วนเบี่ยงเบนมาตรฐานที่ได้จะลดลงกว่าของ prior

2.3 ทฤษฎีการจัดพอร์ตการลงทุน

ทฤษฎีกลุ่มหลักทรัพย์ของ Markowitz

พัฒนาขึ้นโดยนักเศรษฐศาสตร์ชื่อ Harry Markowitz ในปี 1952 โดยกำหนดให้ อรรถประโยชน์ที่ผู้ลงทุนจะได้รับจากการลงทุนในกลุ่มหลักทรัพย์จะขึ้นอยู่กับส่วนประกอบ 2 ส่วนด้วยกัน คือ อัตราผลตอบแทนที่ได้รับจากกลุ่มหลักทรัพย์ (r_p) และความเสี่ยง (σ_p^2) ซึ่งวัดจากความแปรปรวนของอัตราผลตอบแทน บางครั้งจึงถูกเรียกว่าเป็น “Mean-variance portfolio selection model” โดยผลตอบแทนและความเสี่ยงสามารถคำนวณได้จาก

$$r_p = \tilde{Y}^T w$$

$$\sigma_p^2 = w^T \Sigma w$$

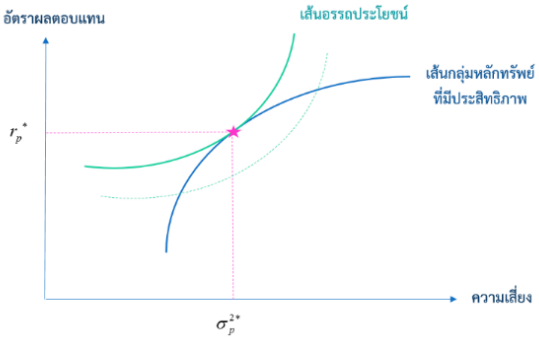
เมื่อ $\tilde{Y}$ คือ ค่าคาดหวังของผลตอบแทนแต่ละหลักทรัพย์

w คือ สัดส่วนการลงทุนในแต่ละหลักทรัพย์

ตามหลักเศรษฐศาสตร์ ผู้ลงทุนจะได้รับอรรถประโยชน์สูงขึ้น เมื่อได้รับผลตอบแทนมากขึ้น และอรรถประโยชน์จะลดลงเมื่อความเสี่ยงของกลุ่มหลักทรัพย์ดังกล่าวเพิ่มขึ้น เส้นอรรถประโยชน์ของผู้ลงทุนจึงมีลักษณะเป็นเส้นโค้งหงาย โดยทุกจุดบนเส้นอรรถประโยชน์เดียวกันจะแสดงสัดส่วนของผลตอบแทนและความเสี่ยงที่ทำให้ผู้ลงทุนได้รับอรรถประโยชน์เท่ากัน เส้นอรรถประโยชน์ของนักลงทุนแต่ละคนจะมีรูปร่างแตกต่างกัน ขึ้นอยู่กับว่าผู้ลงทุนรายนั้นจะมีอรรถประโยชน์ลดลงมากน้อยเพียงใดเมื่อต้องรับความเสี่ยงเพิ่มขึ้น (ถ้าต้องการหลีกเลี่ยงความเสี่ยงมาก เส้นอรรถประโยชน์จะมีความชันมากด้วย) ซึ่งวัดได้จากค่าสัมประสิทธิ์การหลีกเลี่ยงความเสี่ยง (Risk Aversion Coefficient) แทนด้วยสัญลักษณ์ γ ถ้ามีค่ามากแสดงว่าต้องการหลีกเลี่ยงความเสี่ยงมากนั่นเอง

รูปที่ 2.2 แสดงให้เห็นว่าผู้ลงทุนจะเลือกลงทุนในกลุ่มหลักทรัพย์ที่ให้อรรถประโยชน์สูงสุดภายใต้เงินลงทุนที่มี โดยจะเลือกลงทุนในกลุ่มหลักทรัพย์ที่มีประสิทธิภาพ ซึ่งแสดงได้ด้วยเส้นกลุ่มหลักทรัพย์ที่มีประสิทธิภาพ (Markowitz Efficient Frontier) กล่าวคือ ทุกจุดบนเส้นนี้จะไม่มีพอร์ตการลงทุน

อื่นที่ให้อัตราผลตอบแทนสูงกว่า ณ ระดับความเสี่ยงเท่ากัน และไม่มีพอร์ตการลงทุนที่มีความเสี่ยงต่ำกว่า ณ ระดับอัตราผลตอบแทนเท่ากัน จากกราฟด้านล่าง จุดดุลยภาพ r_p^* และ σ_p^{2*} จะเป็นจุดที่เส้นอรรถประโยชน์ของผู้ลงทุนสัมผัสกับเส้นกลุ่มหลักทรัพย์ที่มีประสิทธิภาพภายใต้เงินลงทุนหนึ่งๆ พอดี โดย ณ จุดนี้ ผู้บริโภคจะได้รับอรรถประโยชน์สูงสุดจากการลงทุนภายใต้เงินลงทุนที่มี



รูปที่ 2.2 อัตราผลตอบแทนและความเสี่ยงที่เหมาะสมตามหลักการจัดพอร์ตการลงทุนของ Markowitz

โดยผู้ลงทุนจะหาพอร์ตการลงทุนที่มีสัดส่วนการลงทุน w_p^* ที่ได้รับอรรถประโยชน์จากผลตอบแทนเมื่อปรับด้วยความเสี่ยงจากการลงทุนมากที่สุด คือ

$$\max_{w_p^*} \left(u_p = \tilde{Y}^T w_p - \frac{\gamma}{2} w_p{}^T \Sigma w_p \right)$$

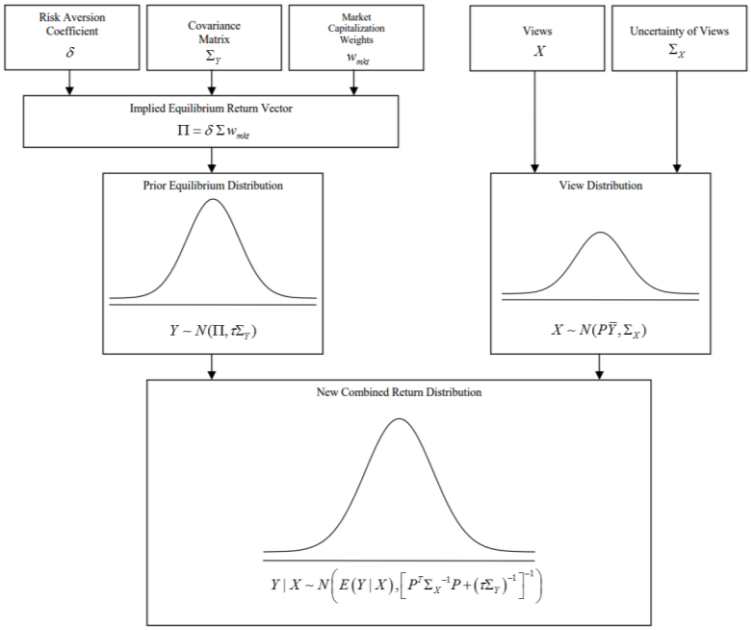
ซึ่งจะอธิบายโดยละเอียดอีกครั้งในหัวข้อการวัดประสิทธิภาพของการจัดพอร์ตการลงทุน (Certainty Equivalent of Return)

ทฤษฎีการจัดพอร์ตการลงทุนของ Black-Litterman model

ตั้งชื่อตาม Fisher Black และ Robert Litterman ที่ได้พัฒนาโมเดลนี้ขึ้นตั้งแต่ปี ค.ศ. 1990 โดยได้ใช้หลักการของ Capital Asset Pricing Model (CAPM) และ Markowitz's mean-variance optimization model ร่วมกัน โมเดลนี้มีจุดเด่นที่ใช้ข้อมูลประกอบ (unique views) มามีส่วนในการคำนวณผลตอบแทนจากการลงทุน โดยอาศัยวิธีการของเบส์ในการรวมความเห็นของนักลงทุนกับค่าของผลตอบแทนจากดุลยภาพของตลาดที่ผ่านมา (prior distribution)

ข้อมูลที่นำมาใช้คำนวณผลตอบแทนตาม Black-Litterman จะประกอบด้วยข้อมูล 2 ส่วนหลักๆ คือ การแจกแจงของค่าคาดหวังของผลตอบแทน (prior) ซึ่งพิจารณาได้ทั้งจากดุลยภาพของตลาดและจากข้อมูลในอดีต รวมกับข้อมูลอีกส่วนที่เป็นข้อมูลประกอบ (side information) ที่ได้จากความคิดเห็นของนักลงทุน (subjective views) Black-Litterman กำหนดให้ข้อมูลดังกล่าวอยู่ในรูปความสัมพันธ์เชิงเส้นของผลตอบแทนของหุ้นที่จะนำมาจัดพอร์ตการลงทุน (linear combination of

returns) จากนั้นจึงนำข้อมูลดังกล่าวมาประมาณค่าเพื่อหาการแจกแจง posterior นอกจากนี้ ยังต้องระบุความเชื่อมั่นต่อข้อมูลที่ใช้เพื่อคำนวณหาความเสี่ยง โดยใช้หลักของ Markowitz's mean-variance optimization โดย Idzorek (2007) ได้สรุปแผนภาพการคำนวณตาม Black-Litterman model ไว้ดังรูป



รูปที่ 2.3 หลักการจัดพอร์ตการลงทุนตาม Black-Litterman Model

กำหนดให้ Y คือ ผลตอบแทนของหลักทรัพย์จำนวน N หลักทรัพย์ โดยให้มีการแจกแจงแบบปกติ (multivariate normal) ที่มีค่าเฉลี่ย $\bar{Y}$ และมีเมทริกซ์ความแปรปรวนร่วมเป็น $\tau\Sigma_Y$ และให้ X คือ ความคิดเห็น (view) เกี่ยวกับผลตอบแทนของหลักทรัพย์ โดยให้มีการแจกแจงแบบแบบปกติที่มีค่าเฉลี่ย $P\bar{Y}$ เมื่อ P เป็นเมทริกซ์ขนาด $K \times N$ และมีเมทริกซ์ความแปรปรวนร่วมเป็น Σ_X

ข้อกำหนดข้างต้น จะได้การแจกแจง posterior ที่มีเงื่อนไขของ view ดังต่อไปนี้

$$f(Y \mid X) \propto f(X \mid Y) f(Y)$$
$$\propto \exp\left(-\frac{1}{2}\left[(X-PY)^T \Sigma_X^{-1}(X-PY) + (Y-\bar{Y})^T (\tau\Sigma_Y)^{-1}(Y-\bar{Y})\right]\right)$$

และสามารถหาค่าคาดหวังได้จาก

$$E(Y \mid X) = \left[P^T\Sigma_X^{-1}P + (\tau\Sigma_Y)^{-1}\right]^{-1}\left[P^T\Sigma_X^{-1}X + (\tau\Sigma_Y)^{-1}\bar{Y}\right]$$

งานของ Chiarawongse et al. (2012) แสดงให้เห็นว่า เมื่อประยุกต์ใช้ Black-Litterman model กับข้อมูลประกอบเชิงอันดับทำให้ผลลัพธ์ของการจัดพอร์ตการลงทุนที่ได้มีประสิทธิภาพมากขึ้น หลังจากนั้นสุพัตรา เพ็ชรน้ำขาว (2015) ศึกษาการจัดพอร์ตการลงทุนตามงานของ

Chiarawongse et al. (2012) โดยใช้ข้อมูลของผลตอบแทนรายเดือน 10 อุตสาหกรรมในประเทศสหรัฐอเมริกา พบว่า การใช้ข้อมูลประกอบเชิงอันดับทำให้ประสิทธิภาพดีขึ้นจริง สอดคล้องกับงานของปริยากร มณีกูล (2015) ที่แสดงให้เห็นว่าประสิทธิภาพของการจัดพอร์ตการลงทุนจะต่ำลงเมื่อค่าสัมประสิทธิสหสัมพันธ์ของหลักทรัพย์มีค่าสูงขึ้น

การคำนวณผลตอบแทนดุลยภาพ (Equilibrium Return)

การจัดพอร์ตการลงทุนตามหลักของ Black-Litterman Model นั้น ข้อมูลหนึ่งที่สำคัญในการคำนวณ คือ การหาพารามิเตอร์ของเวกเตอร์ค่าเฉลี่ยสำหรับการแจกแจงก่อน (prior) ซึ่งแทนด้วยสัญลักษณ์ Π ตามรูปที่ 2.3 ซึ่งการคำนวณค่าดังกล่าวทำได้หลายวิธี Satchell and Scowcroft (2003) กล่าวถึงข้อดีและข้อเสียของการคำนวณแต่ละวิธีไว้ดังนี้

- ใช้ค่าเฉลี่ยผลตอบแทนในอดีต (historical sample mean) วิธีนี้คำนวณได้งมีงานวิจัยหลายงานที่พิสูจน์แล้วว่าการใช้ค่าเฉลี่ยของผลตอบแทนในอดีตไม่สามารถใช้เป็นผลตอบแทนดุลยภาพที่เหมาะสมได้ เนื่องจากค่าที่ได้จะมีความคลาดเคลื่อนสูง Idzorek (2007) แสดงให้เห็นผลของการคำนวณโดยใช้ผลตอบแทนในอดีตของสินทรัพย์ประเภทต่างๆ พบว่า ค่าที่ได้มีความแปรปรวนสูงมาก นอกจากนี้ความคลาดเคลื่อนจะไม่ลดลงแม้ว่าจะเพิ่มจำนวนหรือความถี่ของข้อมูล เพราะโดยธรรมชาติของข้อมูลผลตอบแทน เมื่อเราพิจารณาผลตอบแทนที่มีความถี่เพิ่มขึ้น ความแปรปรวนของผลตอบแทนที่ได้จะสูงขึ้นตามไปด้วย แม้ว่าจะมีแนวคิดที่เสนอให้ปรับค่าเฉลี่ย เช่น Stein estimators แต่วิธีดังกล่าวก็ยังคงยากที่จะนำไปใช้จริงในทางปฏิบัติ
- ใช้วิธี Reverse Optimization โดยวิธีนี้จะคำนวณหาผลตอบแทนดุลยภาพจากสมการ

  $$\Pi = \delta \Sigma w_{mkt}$$

  โดยค่าคงที่ δ ที่ใช้ปรับค่า (scaling factor) ในสมการข้างต้น มีวิธีคำนวณได้หลายแบบ ในงานวิจัยด้านการเงิน เช่น Satchell and Scowcroft (2003) และ Idzorek (2007) จะใช้คำนวณ Risk Aversion Coefficient จากสูตร

  $$\delta = \frac{Risk\ \ Premium}{Market\ Varinace}$$

  ค่า Risk Premium คำนวณจากค่าคาดหวังของผลตอบแทนเฉลี่ยของตลาดลบด้วยผลตอบแทนจากสินทรัพย์ที่ปราศจากความเสี่ยง ส่วน Market Variance สามารถหาจาก

  $$\sigma_{mkt}^2 = w_{mkt}^T \Sigma w_{mkt}$$

2.4 ความไม่แน่นอนของข้อมูลอันดับ

ในความเป็นจริงเป็น ณ เวลาที่ตัดสินใจลงทุน เราจะยังไม่มีข้อมูลอันดับของผลตอบแทนของหุ้นที่สมบูรณ์ (perfect rank) มาใช้ในการจัดพอร์ตการลงทุน ดังนั้น จึงจำเป็นต้องพิจารณาความไม่แน่นอนที่จะเกิดขึ้นอันเนื่องมาจากการที่ข้อมูลที่ใช้ไม่สมบูรณ์ วิธีการพิจารณาความไม่แน่นอนของผลตอบแทนสามารถทำได้หลายวิธี ขึ้นอยู่กับทฤษฎีที่สนใจและพื้นฐานที่มาของข้อมูล

ยกตัวอย่างเช่น ในงานวิจัยของ Chiarawongse et al. (2012) ได้กล่าวถึงกรณีที่ข้อมูลอันดับ (qualitative views) ที่ได้อาจไม่เท่ากับอันดับที่ถูกต้อง โดยเสนอว่า

$$\mathrm{P}(Y \mid \mathfrak{R} = \mathbb{R}) = K\,\mathrm{P}(Y \mid Y \in \mathbb{R}) + (1-K)\,\mathrm{P}(Y)$$

เมื่อ $\mathfrak{R}$ แทนตัวแปรสุ่มของ observed view และ K คือ ความน่าจะเป็นที่ view จะถูกต้อง หรือเป็น confidence on views คำนวณได้จาก

$$K = \frac{\mathrm{P}(Y \mid \mathfrak{R} = \mathbb{R}) - \mathrm{P}(Y)}{\mathrm{P}(Y \mid Y \in \mathbb{R}) - \mathrm{P}(Y)}$$

และจะได้ว่า $\tilde{Y} = K\hat{Y} + (1-K)\bar{Y}$

โดย $\tilde{Y} = \mathrm{E}(Y \mid \mathfrak{R} = \mathbb{R})$, $\hat{Y} = \mathrm{E}(Y \mid Y \in \mathbb{R})$, $\bar{Y} = \mathrm{E}(Y)$ และ $\mathbb{R} = \left\{ y \in \Re^N : Ay \le b \right\}$

สมการข้างต้น แสดงถึงความสัมพันธ์ของค่าคาดหวังที่ได้จากข้อมูลประกอบเชิงอันดับที่ไม่สมบูรณ์ (imperfect view) ที่เกิดจากความสัมพันธ์เชิงเส้นที่เป็น convex combination ระหว่างค่าคาดหวังที่ได้จากข้อมูลประกอบเชิงอันดับที่ถูกต้อง ค่าคาดหวังของผลตอบแทนจากการแจกแจงก่อน (prior) เมื่อไม่ได้นำข้อมูลประกอบมาพิจารณา โดย Chiarawongse et al. (2012) ได้จำลองข้อมูลและพบว่าเมื่อค่า K มีค่ามากขึ้น (ค่าใกล้ 1) ค่า certainty equivalent of the payoff ซึ่งใช้วัดประสิทธิภาพการจัดพอร์ตการลงทุนจะยิ่งแตกต่างกันมากขึ้นเมื่อจำนวนหุ้นแตกต่างกัน งานของสุพัตรา เพ็ชรน้ำขาว (2015) และปริยากร มณีกูล (2015) ศึกษาผลของค่า K ต่อการจัดพอร์ตการลงทุนไว้ว่าการเลือกค่าความเชื่อมั่นให้เหมาะสมมีส่วนช่วยเพิ่มประสิทธิภาพของการจัดพอร์ตการลงทุนได้

ในการงานวิจัยนี้ที่เลือกใช้ข้อมูลอันดับของตัวแปรร่วม ก็จำเป็นต้องพิจารณาความผิดพลาดที่เกิดจากการที่ข้อคิดเห็น (subjective views) ของผู้เชี่ยวชาญที่ทำนายอันดับของผลตอบแทนอาจไม่ถูกต้องทั้งหมด (imperfect rank) โดยจะต้องพิจารณาส่วนของความผิดพลาด (view error) ร่วมด้วย แต่กระบวนการวิเคราะห์จะแตกต่างจากการกำหนดค่าความเชื่อมั่น (K) ตามวิธีใน Chiarawongse et al. (2012) เนื่องจากลักษณะของข้อมูลประกอบที่จะนำมาวิเคราะห์แตกต่างกัน ซึ่งงานวิจัยนี้จะศึกษาว่าหากตัวแปรร่วมและตัวแปรที่สนใจศึกษามีความสัมพันธ์กันในลักษณะและเงื่อนไขที่ต่างกันแล้ว ค่าความไม่แน่นอนหรือค่าความผิดพลาดที่เกิดขึ้นจะเป็นอย่างไร

Lee et al. (2014) ประยุกต์ใช้แนวคิดของ Thurstone's model มาใช้ในการอธิบายความไม่แน่นอนของการจัดอันดับของแต่ละคนไว้ว่า หากให้แต่ละคนจัดอันดับข้อมูล จะเกิดข้อผิดพลาดที่เรียกว่า "subject-level uncertainty" ที่มีการแจกแจงแบบปกติ ที่มีค่าเฉลี่ย μ_j และส่วนเบี่ยงเบนมาตรฐาน σ_j อธิบายได้ดังรูป

สมมติให้ข้อมูลที่ถูกต้อง (Latent Ground Truth) มีค่าเฉลี่ยที่แท้จริงเป็น μ_1, μ_2, μ_3 และ μ_4 ตามลำดับ

ถ้าให้คนมาจัดอันดับข้อมูล จะเกิดความไม่แน่นอนของตำแหน่งค่าเฉลี่ยที่มีการแจกแจงแบบ $N(\mu_j, \sigma_j^2)$ จากนั้นแต่ละคนจะจัดอันดับที่ของข้อมูลโดยการสุ่มค่าจากแต่ละการแจกแจงมาจัดอันดับ ซึ่งอันดับที่ได้ก็จะแตกต่างกัน ตามรูปร่างการแจกแจงของแต่ละคน

รูปที่ 2.4 ผลของระยะห่างของค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานต่อความข้อผิดพลาดของข้อมูลอันดับ

จากรูปด้านบนจะเห็นได้ชัดว่าอันดับที่ได้จะแตกต่างจากค่าจริงมากน้อยแค่ไหนจะขึ้นอยู่กับค่าส่วนเบี่ยงเบนมาตรฐาน σ_j และระยะห่างระหว่าง μ แต่ละค่า

2.5 การวัดประสิทธิภาพของการจัดพอร์ตการลงทุนโดยใช้ Certainty Equivalent Return

ในการเปรียบเทียบประสิทธิภาพของพอร์ตการลงทุนแต่ละแบบเทียบกัน โดยดูจากค่าอรรถประโยชน์ที่ผู้ลงทุนจะได้รับ ซึ่งตามหลักของ Markowitz พบว่า อรรถประโยชน์ที่ได้รับจะขึ้นอยู่กับองค์ประกอบ 2 ส่วนด้วยกัน คือ ผลตอบแทนและความเสี่ยง ตามสมการ

$$u_p = \tilde{Y}^T w_p - \frac{\gamma}{2} w_p{}^T \Sigma w_p$$

เมื่อ u_p คือ อรรถประโยชน์ที่ได้รับจากการลงทุนในพอร์ตการลงทุน

$\tilde{Y}$ คือ ค่าคาดหวังของผลตอบแทนแต่ละหลักทรัพย์

w_p คือ อัตราส่วนการลงทุนในแต่ละหลักทรัพย์

Σ คือ เมทริกซ์ความแปรปรวนและความแปรปรวนร่วมของหลักทรัพย์

γ คือ ค่าสัมประสิทธิ์การหลีกเลี่ยงความเสี่ยง (Risk Aversion Coefficient)

ในสมการข้างต้น พจน์แรกแสดงถึงอรรถประโยชน์ที่จะได้รับจากผลตอบแทนจากการลงทุน และปรับด้วยพจน์หลังที่แสดงถึงอรรถประโยชน์ที่ลดลงเมื่อความเสี่ยงเพิ่มขึ้น สำหรับค่าสัมประสิทธิ์การหลีกเลี่ยงความเสี่ยงนั้น สามารถคำนวณได้จากค่า Arrow-Pratt risk aversion index ตามรายละเอียดใน Pratt (1964) และ Levy and Levy (1991) โดย Ang (2014) อธิบายไว้ว่า นักลงทุนทั่วไปจะมีค่านี้อยู่ระหว่าง 1 ถึง 10 ซึ่งค่ามากแสดงว่านักลงทุนรายนี้พยายามหลีกเลี่ยงความเสี่ยงมากด้วย

2.6 CAPM model และเมทริกซ์ความแปรปรวนร่วมตามตัวแบบปัจจัยเดียว (One-factor Model)

เนื่องจากการคำนวณค่าคาดหวังด้วยวิธีปริพันธ์เวียนบังเกิด (recursive integration) เมทริกซ์ความแปรปรวนร่วม ต้องอยู่ในรูปตัวแบบปัจจัยเดียว (One-factor Model) ซึ่งสามารถใช้ CAPM model มาช่วยในการประมาณค่าดังกล่าวได้ โดย

$$\Sigma_Y = \begin{bmatrix} \text{var}(Y_1) & \text{cov}(Y_1,Y_2) & \dots & \text{cov}(Y_1,Y_N) \\ \text{cov}(Y_2,Y_1) & \text{var}(Y_2) & \dots & \text{cov}(Y_2,Y_N) \\ \dots & \dots & \dots & \dots \\ \text{cov}(Y_N,Y_1) & \text{cov}(Y_N,Y_2) & \dots & \text{var}(Y_N) \end{bmatrix}$$

โดยค่า $\text{cov}(Y_i,Y_j)$ สามารถประมาณได้จาก $\beta_i\beta_j\,\text{var}(Y_m)$
เมื่อ Y_i คือ ผลตอบแทนของหลักทรัพย์ i และ Y_m คือ ผลตอบแทนของตลาด
ซึ่งค่า β_i หาได้จากสมการถดถอยระหว่าง Y_i กับ Y_m ดังนี้ $Y_i = \beta_0 + \beta_i Y_m + \varepsilon_i$ ค่า β_i นี้จะแสดงความสัมพันธ์ของผลตอบแทนระหว่างหลักทรัพย์ที่สนใจกับผลตอบแทนของทั้งตลาด

จากนั้นใช้สมการตาม Kiatsupaibul et al. (2017) ที่กำหนดให้

$$Y_i = \bar{Y}_i + \rho_i M + \sqrt{\sigma_i^2 - \rho_i^2}\,Z_i \quad ; \quad 1 \le i \le N$$

เมื่อ M และ Z_i เป็นตัวแปรสุ่มที่มีการแจกแจงปกติมาตรฐานที่เป็นอิสระกัน (independent standard normal)

2.7 การมีส่วนร่วมของกลุ่มคน (Crowd Sourcing)

แนวคิดเรื่อง Crowd prediction หรือที่รู้จักกันในชื่อว่า the wisdom of crowds มีการศึกษาและพิสูจน์ในงานวิจัยหลายด้านว่า เมื่อรวมผลการทำนายจากหลายแหล่งจะทำให้ผลการทำนายมีประสิทธิภาพมากขึ้น กล่าวคือ ค่าเฉลี่ยที่ได้จะใกล้เคียงกับค่าจริงและมีค่าความแปรปรวนต่ำ เช่น การทำนายสภาพอากาศ การทำนายราคาตลาด (market prediction) เป็นต้น

การทำนายราคาหุ้นก็เคยมีผู้ศึกษาไว้ว่า crowd sourcing สามารถทำนายราคาได้ใกล้เคียงกับค่าจริง ยกตัวอย่างเช่น งานศึกษาของ Wang et al. (2015) ที่ศึกษาข้อมูลที่ได้จากเว็บไซต์ SeekingAlpha และ StockTwits ซึ่งเป็น crowdsourcing investment service ที่มีผู้ใช้จำนวนมาก โดยการสร้างเครื่องมือในการวิเคราะห์ข้อความที่ผู้ใช้เข้ามาแสดงความคิดเห็นเกี่ยวกับราคาหุ้น โดยจะบันทึกว่าเป็นความคิดเห็นที่เป็นบวกหรือเป็นลบต่อราคาหุ้นนั้น แล้วมาคำนวณเป็น sentimental score แล้วนำค่าดังกล่าวไปใช้สร้างกลยุทธ์การลงทุน ซึ่งพบว่าผลตอบแทนจากการลงทุนดีกว่าผลตอบแทนเฉลี่ยของตลาด อย่างไรก็ตาม มีรายละเอียดบางประเด็นที่น่าสนใจ เช่น การกำหนดช่วงระยะเวลาที่จะนำข้อมูลมาใช้ในการพยากรณ์ (time window) ซึ่งข้อมูลแต่ละแหล่งก็มีช่วงเวลาการตอบสนองต่อการเปลี่ยนแปลงของราคาที่แตกต่างกัน หรือการเปรียบเทียบผลการทำนายระหว่างการใช้ข้อมูลจากผู้ใช้ทั้งหมดกับการเลือกใช้ข้อมูลจากกลุ่มผู้เชี่ยวชาญเฉพาะที่สามารถทำนายได้ใกล้เคียงกับค่าจริง ซึ่งงานวิจัยของ Wang et al. (2015) ชิ้นนี้ได้ศึกษากระบวนการมีส่วนร่วมและการระบุระดับความเชี่ยวชาญ (Expertized Identification) ของผู้ใช้ไว้ด้วย

งานวิจัยของ Davis-Stober et al. (2014) ได้กล่าวถึงเงื่อนไขที่ทำให้การใช้ crowd prediction มีประสิทธิภาพไว้อย่างน่าสนใจ คือ การที่กลุ่มคนจะสามารถทำนายค่าได้อย่างมีประสิทธิภาพ หากกลุ่มคนดังกล่าวมีลักษณะที่แตกต่างกัน โดยยกตัวอย่างว่า การเพิ่มความคิดเห็นของคนที่อาจจะไม่ได้เป็น expert แต่มีเห็นความแตกต่างจากคนส่วนใหญ่ (negatively uncorrelated) ก็สามารถช่วยให้ผลการทำนายดีขึ้นได้ นอกจากนี้ยังได้ข้อสรุปว่าหากคนในกลุ่มมี bias ที่มีความสัมพันธ์กัน การเลือกเฉพาะคนที่มีความสามารถในการทำนายมาบางส่วน (select only highly skilled judges) จะให้ผลที่แม่นยำกว่าการใช้ค่าเฉลี่ยของกลุ่มคนทั้งหมด นอกจากนี้ยังมีการศึกษาวิธีการเพิ่มประสิทธิภาพของผลการทำนายและการพิจารณาลักษณะการแจกแจงของ forecast error และ asymmetric risk ที่เกิดขึ้น โดยสามารถศึกษาเพิ่มเติมได้จากงานของ Menezes et al. (2000)

สำหรับการหาค่าคาดหวังตามโมเดลของ Black-Litterman ก็จะเห็นได้ว่าประสิทธิภาพของผลลัพธ์ที่ได้จากโมเดลข้างต้น ขึ้นอยู่กับความถูกต้องของข้อมูลอันดับของผลตอบแทน ณ เวลาที่สนใจ การรวบรวมข้อมูลของตัวแปรร่วมหรือความคิดเห็นจากหลายแหล่งจึงเป็นทางเลือกที่น่าสนใจในการลดความผิดพลาด

2.8 การรวบรวมผลการทำนาย (Forecasting Combination) และการรวมข้อมูลอันดับ (Rank Aggregation)

เมื่อข้อมูลที่จะนำมาใช้วิเคราะห์ไม่ได้มาจากแหล่งเดียว ก่อนจะนำข้อมูลไปวิเคราะห์ในโมเดลจึงจำเป็นต้องมีการรวมข้อมูลดังกล่าวให้เป็นข้อมูลอันดับอันเดียว โดยกระบวนการรวมข้อมูลทำได้หลายวิธี ขึ้นอยู่กับลักษณะของข้อมูลและวัตถุประสงค์ของการใช้ข้อมูล โดยปกติแล้วการรวมข้อมูลอันดับ

แบ่งออกเป็น 2 ลักษณะตามวัตถุประสงค์ของผลลัพธ์ที่ต้องการ คือ การรวมข้อมูลเพื่อให้ได้ผลของข้อมูลอันดับที่สะท้อนถึงข้อมูลทั้งหมดที่มีให้มากที่สุด (best reflects the ranking inputs) เช่น การรวมข้อมูลความชอบ (rating) ของสินค้าต่างๆ และการรวมข้อมูลเพื่อให้ตรงกับข้อเท็จจริงมากที่สุด เช่น การทำนายสภาพอากาศ การทำนายราคา กรณีหลังนี้จะสามารถประเมินได้ว่าผู้ทำนายสามารถทำนายได้ถูกต้องแม่นยำมากน้อยแค่ไหนเมื่อเทียบกับค่าจริง (ground truth) ซึ่งวิธีการรวมข้อมูลอันดับจะแตกต่างกันตามวัตถุประสงค์ เช่น การสร้าง objective scoring function โดยการเปลี่ยนข้อมูลที่มีให้เป็นลำดับคะแนน จากนั้นจึงเรียงลำดับข้อมูลที่สนใจตามลำดับคะแนนดังกล่าว หรืออาจใช้วิธีการเปรียบเทียบข้อมูลทีละคู่ และหาผลลัพธ์ที่ทำให้จำนวน Disagreement pairs น้อยที่สุด เป็นต้น

เนื่องจากข้อมูล (input) จากบทวิเคราะห์หุ้นที่จะนำมาใช้ในงานวิจัยนี้ ไม่ได้เป็นข้อมูลอันดับโดยตรง กล่าวคือ ในบทวิเคราะห์แต่ละอันไม่ได้จัดอันดับหุ้นที่น่าลงทุน แต่จะเป็นการวิเคราะห์และเลือกหุ้นที่น่าลงทุน (แนะนำให้ซื้อ) 1-15 ตัวต่อบทวิเคราะห์ ดังนั้น จึงไม่สามารถใช้หลักการของ Rank Aggregation โดยตรงได้ อย่างไรก็ตาม สามารถประยุกต์ใช้หลักการให้คะแนน (scoring) และแนวคิด pairwise comparison บางส่วนมาช่วยในการให้คะแนนและจัดอันดับได้ โดยรายละเอียดการคิดคะแนนที่เป็นที่นิยมและสามารถนำมาประยุกต์ใช้ในงานวิจัยนี้ได้ มีดังนี้

Historical Performance Score

หากมีข้อมูลความถูกต้องของผลการทำนายในอดีต เราสามารถหาค่าเฉลี่ยถ่วงน้ำหนัก โดยให้น้ำหนักกับข้อมูลที่มาจากแหล่งที่สามารถทำนายได้ถูกต้องในอดีตมาก อย่างเช่นในงานของ Wang et al. (2015) ที่ได้กล่าวไปแล้วในหัวข้อ crowd sourcing ว่าการเลือกใช้ข้อมูลจากกลุ่มผู้เชี่ยวชาญเฉพาะจะสามารถทำนายได้ใกล้เคียงกับค่าจริงมากขึ้น สอดคล้องกับงานของ Hill and Ready-Campbell (2011) ศึกษากระบวนการให้คะแนนแก่ผู้เชี่ยวชาญในการทำนายเพื่อจะหาค่าเฉลี่ยแบบถ่วงน้ำหนักของราคาหุ้น โดยใช้ข้อมูลจากเว็บไซต์ Motley Fool CAPS โดยเปรียบเทียบผลการจัดพอร์ตการลงทุนเมื่อใช้ข้อมูลจากผู้ใช้ทั้งหมดเทียบกับผลตอบแทนของ S&P500 และเมื่อใช้กระบวนการถ่วงน้ำหนักข้อมูลตามระดับความสามารถในการพยากรณ์ในอดีต โดยผลการศึกษาพบว่า the wisdom of experts in crowds ทำให้ผลตอบแทนโดยรวม และผลตอบแทนหลังปรับความเสี่ยงซึ่งวัดจาก sharp ratio ดีขึ้นอย่างมีนัยสำคัญ

สำหรับวิธีการให้คะแนนความถูกต้องสามารถทำได้หลายวิธีขึ้นอยู่กับข้อมูลที่มี เช่น หากเป็นข้อมูลเชิงปริมาณก็จะดูจากค่า Mean Square Error หรือ Correlation ได้ แต่ถ้าเป็นข้อมูลอันดับสามารถใช้ค่าสถิติที่ใช้วัดระยะห่างของข้อมูลอันดับ เช่น Kendall tau distance แทน เป็นต้น

Time Score

เนื่องจากคำแนะนำให้ซื้ออาจไม่มีผลต่อราคาในวันนั้นๆทันที การให้น้ำหนักคะแนนตามระยะเวลา เช่น การถ่วงน้ำหนักคำแนะนำโดยให้น้ำหนักกับวันปัจจุบันที่พิจารณามากที่สุด และให้น้ำหนักกับวันก่อนหน้านั้นลดลงตามลำดับ Antunovich and Sarkar (2006) ได้ศึกษาผลของ stock picks ต่อราคาของหุ้นในตลาด NASDAQ พบว่า หุ้นที่ได้รับคำแนะนำให้ซื้อจะมีราคาสูงขึ้นในวันนั้นๆ อย่างเห็นได้ชัด โดยหุ้นที่มี media exposure มากกว่า ก็จะมีผลให้ราคาสูงมากกว่าหุ้นตัวอื่นด้วย

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 3
วิธีการศึกษา

ในบทนี้จะอธิบายวิธีการศึกษาซึ่งประกอบด้วย 2 ส่วนหลัก คือ *ส่วนแรก*ศึกษาโดยการจำลองข้อมูลเพื่อแสดงให้เห็นผลของปัจจัยต่างๆ ที่มีต่อค่าคาดหวังแบบมีเงื่อนไขของข้อมูลอันดับ รวมทั้งทดลองจัดพอร์ตการลงทุนจากข้อมูลที่จำลองได้ *ส่วนที่สอง*เป็นการทดลองใช้ข้อมูลจริงที่รวบรวมจากบทวิเคราะห์หลักทรัพย์มาจัดพอร์ตการลงทุนและเปรียบเทียบประสิทธิภาพที่ได้เมื่อใช้วิธีการรวมข้อมูลแบบต่างๆ

3.1 การศึกษาโดยการจำลองข้อมูล

งานศึกษาส่วนแรกจะใช้การจำลองข้อมูลตามโมเดลที่กำหนดเพื่อศึกษาความแตกต่างระหว่างค่าของตัวแปรที่สนใจศึกษากับคาดหวังที่มีเงื่อนไขของข้อมูลอันดับ โดยเปรียบเทียบเมื่อใช้ข้อมูลอันดับของตัวแปรที่สนใจศึกษากับเมื่อใช้ข้อมูลอันดับจากตัวแปรร่วม ภายใต้เงื่อนไขต่างๆ กัน ตามรายละเอียดดังต่อไปนี้

3.1.1 ตัวแบบ (Model)

งานวิจัยนี้กำหนดโมเดลความสัมพันธ์โดยให้ค่าของตัวแปรที่สนใจศึกษา (Y) เป็นผลรวมของตัวแปรร่วม (X) กับค่าความคลาดเคลื่อน (ε) ทุกตัวแปรมีการแจกแจงแบบปกติหลายตัวแปร (multivariate normal) ที่ถูกปรับค่าให้อยู่ในรูปแบบมาตรฐาน เพื่อให้สอดคล้องกับการนำผลการศึกษาไปประยุกต์ใช้กับข้อมูลจริงในกรณีที่ค่าของตัวแปรที่สนใจศึกษา (Y) และค่าของตัวแปรร่วม (X) มีหน่วยแตกต่างกัน นอกจากนี้ยังกำหนดให้ X และ ε เป็นอิสระกันด้วย

จากเงื่อนไขข้างต้น ตัวแปรในโมเดลจะมีความสัมพันธ์ตามหลักการของ Multivariate normal distribution with a product correlation structure และตามตัวแบบปัจจัยเดียว (One factor model) ที่อธิบายไว้ใน Tong (1990), เสกสรร เกียรติสุไพบูลย์ (2012) และ Kiatsupaibul et al. (2017) ดังนี้

$$Y = \lambda X + \sqrt{1-\lambda^2}\varepsilon \qquad (3.1)$$

โดย Y คือ เวกเตอร์ของตัวแปรที่สนใจศึกษาที่ปรับค่าให้เป็นค่ามาตรฐาน

X คือ เวกเตอร์ของตัวแปรร่วมซึ่งปรับค่าเป็นค่ามาตรฐาน โดยมีค่าเฉลี่ยเป็นศูนย์ และ

มีเมทริกซ์ความแปรปรวนร่วมเป็นเมทริกซ์ขนาด $N \times N$ ที่สมาชิกทุกตัวบนเส้นทแยงมุมมีค่าเป็น 1 และสมาชิกนอกเส้นทแยงมุมมีค่าตามค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรร่วมแต่ละตัว (ρ)

ε คือ เวกเตอร์ของค่าความคลาดเคลื่อน ในที่นี้จะมีการแจกแจงแบบปกติ มีค่าเฉลี่ยเป็นศูนย์ และมีเมทริกซ์ความแปรปรวนร่วมเป็น Identity Matrix ขนาด $N \times N$

λ คือ ค่าคงที่แสดงระดับความสมบูรณ์ของข้อมูลอันดับของตัวแปรร่วม (degree of perfect ranking)

ค่า λ มีค่าระหว่าง 0 ถึง 1 โดยหาก λ มีค่ามากจะแสดงว่าข้อมูลอันดับมีความสมบูรณ์มาก ซึ่งก็คือ ข้อมูลอันดับของตัวแปรร่วม (r_x) จะมีค่าสอดคล้องกับข้อมูลอันดับของตัวแปรที่ต้องการศึกษา (r_y) และมีความหมายโดยนัยว่า ตัวแปรร่วมและตัวแปรที่สนใจศึกษามีความสัมพันธ์กันในระดับสูงด้วย

นอกจากนี้ การพิจารณาค่า λ ก็เปรียบเสมือนว่าเรากำลังพิจารณาน้ำหนักที่ให้กับตัวแปรร่วมและความค่าความผิดพลาด ในแง่การตีความหมายจะสอดคล้องกับค่าความเชื่อมั่นใน Chiarawongse et al. (2012) คือ ถ้าค่าระดับความสมบูรณ์ของข้อมูลอันดับของตัวแปรร่วม λ มากขึ้น ก็ควรจะให้น้ำหนักกับค่าของตัวแปรร่วมมากขึ้นและให้น้ำหนักกับเทอมของค่าความผิดพลาดน้อยลงนั่นเอง

นอกเหนือจากโมเดลข้างต้น งานวิจัยนี้สนใจศึกษาผลของระยะห่างระหว่างค่าในเวกเตอร์ค่าเฉลี่ยด้วย ผู้วิจัยจึงปรับโมเดลที่จะศึกษาโดยบวกเวกเตอร์ค่าเฉลี่ยเพิ่มเข้าไปในสมการ 3.1 ดังนี้

$$Y = \lambda X + \sqrt{1-\lambda^2}\varepsilon + \mu$$

เมื่อ μ คือ เวกเตอร์ค่าเฉลี่ย จากนั้นปรับสัญลักษณ์ที่ใช้เพื่อความสะดวกในการอ้างอิงในลำดับถัดไปเป็น

$$Y = V + \sqrt{1-\lambda^2}\varepsilon \qquad (3.2)$$

เมื่อ $V = \lambda X + \mu$ และในกรณีนี้จะถือว่าตัวแปร V เป็นตัวแปรร่วม

3.1.2 เงื่อนไขของตัวแปรที่ต้องการศึกษา

การจำลองข้อมูลจะแบ่งเป็น 2 ส่วน คือ ส่วนแรกจะจำลองข้อมูลโดยกำหนดให้ตัวแปรมีความสัมพันธ์กันตามสมการ (3.1) ที่ยังไม่นำค่าในค่าเฉลี่ยมาพิจารณา (ในส่วนนี้ให้เวกเตอร์ค่าเฉลี่ยของตัวแปรที่สนใจศึกษาเป็นศูนย์) แล้วศึกษาผลของปัจจัยต่างๆ ต่อไปนี้

(1) จำนวนตัวแปรที่สนใจ (N)

(2) ระดับความสมบูรณ์ของข้อมูลอันดับของตัวแปรร่วม (λ)

(3) ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรร่วม (ρ)

โดยจะจำลองข้อมูลตามเงื่อนไข ดังต่อไปนี้

ปัจจัยที่ต้องการศึกษา	ค่าที่จะใช้ในการจำลอง
จำนวนตัวแปร (N)	5, 10 และ 50
ระดับความสมบูรณ์ของข้อมูลอันดับของตัวแปรร่วม (λ)	0.25, 0.5 และ 0.9
ค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรร่วม (ρ)	0, 0.5 และ 0.9

หลังจากนั้นจึงจะรวมระยะห่างระหว่างค่าในเวกเตอร์ค่าเฉลี่ยเข้ามา โดยจำลองข้อมูลตามสมการที่ (3.2) สำหรับการจำลองข้อมูลในส่วนนี้ และเลือกศึกษากรณีที่

- จำนวนตัวแปร (N) มีค่าเป็น 10
- ระดับความสมบูรณ์ของข้อมูลอันดับของตัวแปรร่วม (λ) เท่ากับ 0.5 เนื่องจากต้องการศึกษาว่าหากตัวแปรร่วมและตัวแปรที่สนใจศึกษามีความสัมพันธ์กันในระดับปานกลางแล้ว หากค่าเฉลี่ยของตัวแปรแต่ละตัวในเวกเตอร์ค่าเฉลี่ยมีค่าแตกต่างกันแล้วจะทำให้ความแตกต่างของค่าคาดหวังแต่ละแบบลดลงมากน้อยเพียงใด

กำหนดให้ระยะห่างของค่าเฉลี่ยแต่ละค่าห่างเท่าๆ กันก่อน โดยศึกษาเปรียบเทียบเมื่อระยะห่างนี้มีค่าเป็น 0, 0.5, 1, 2 และ 3 ตามลำดับ

3.1.3 การคำนวณค่าคาดหวัง

หลังจากจำลองข้อมูลตามเงื่อนไขข้างต้นแล้ว จึงคำนวณค่าคาดหวังที่มีเงื่อนไขของข้อมูลประกอบเชิงอันดับ ในงานวิจัยนี้จะใช้วิธีปริพันธ์เวียนเกิด (Recursive Integration)

กรณีที่ 1 ใช้ข้อมูลอันดับที่ถูกต้องของตัวแปรที่สนใจศึกษา (Perfect Rank)

คำนวณค่าคาดหวังได้จากการแจกแจงต่อไปนี้

$$f\left(Y \mid Y \in \mathbb{R}\right) \propto I_{\mathbb{R}} \exp\left(-\frac{1}{2}\left[\left(y-\bar{y}\right)^{T} \Sigma^{-1}\left(y-\bar{y}\right)\right]\right)$$

เมื่อ $\mathbb{R} = \left\{y \in \Re^{N} : Ay \le b\right\}$

$$\hat{y} = E\left(Y \mid Y \in \mathbb{R}\right) = \frac{1}{K} \int_{x \in \mathbb{R}} x \exp\left(\frac{1}{2}\left[\left(x-\bar{y}\right)^{T} \Sigma^{-1}\left(x-\bar{y}\right)\right]\right) dx$$

เมื่อ $K = \int_{x\in\mathbb{R}} \exp\left(\frac{1}{2}\left[(x-\bar{y})^T \Sigma^{-1}(x-\bar{y})\right]\right)dx$

กรณีที่ 2 ใช้ข้อมูลอันดับของตัวแปรร่วม

- *เมื่อใช้โมเดลที่ยังไม่รวมผลของระยะห่างของค่าในเวกเตอร์ค่าเฉลี่ย*

กำหนดให้ตัวแปรร่วม (X) และค่าความผิดพลาด (ε) เป็นอิสระต่อกัน และตัวแปรแต่ละตัวมีการแจกแจงแบบปกติหลายตัวแปร (multivariate normal) และความสัมพันธ์ของตัวแปรร่วมและตัวแปรที่สนใจศึกษา เป็นไปตามสมการที่ (3.1) แล้วจะคำนวณหาได้ค่าคาดหวังที่มีเงื่อนไขเชิงอันดับได้จาก

$$\begin{aligned} E\left[Y \mid R(X)=r_x\right] &= E\left[\lambda X + \sqrt{1-\lambda^2}\varepsilon \mid R(X)=r_x\right] \\ &= \lambda E\left[X \mid R(X)=r_x\right] + \sqrt{1-\lambda^2}E\left[\varepsilon \mid R(X)=r_x\right] \\ &= \lambda E\left[X \mid R(X)=r_x\right] + \sqrt{1-\lambda^2}E\left(\varepsilon\right) \end{aligned}$$

จะเห็นว่า เมื่อกำหนดให้ $E(\varepsilon)=0$ แล้ว จะได้ว่า

$$E\left[Y \mid R(X)=r_x\right] = \lambda E\left[X \mid R(X)=r_x\right] \qquad (3.3)$$

เปรียบเสมือนว่าเราสามารถอนุมานทางสถิติโดยใช้เพียงข้อมูลของตัวแปรร่วมอย่างเดียวได้ อย่างไรก็ตาม ในสถานการณ์จริง เราอาจทราบเพียงข้อมูลอันดับของตัวแปรร่วม แต่ไม่ทราบการรูปแบบการแจกแจงที่จะมาคำนวณค่าคาดหวังตามสมการ (3.3) ได้ ดังนั้น งานวิจัยชิ้นนี้จึงเพิ่มการศึกษา โดยจะทดลองคำนวณค่าคาดหวังกรณีที่กำหนดให้ใช้ข้อมูลอันดับของตัวแปรร่วมในการคำนวณร่วมกับข้อมูลการแจกแจงก่อนของตัวแปรที่สนใจจศึกษา โดยจะเรียกค่าคาดหวังในกรณีดังกล่าวว่าเป็น ค่าคาดหวังแบบมีเงื่อนไข*โดยนัย*บนข้อมูลอันดับของตัวแปรร่วม (Implied Rank) แทนด้วยสัญลักษณ์ $E\left[Y \mid R(Y)=r_x\right]$

โดยสรุปแล้ว ค่าคาดหวังที่จะคำนวณและเปรียบเทียบกันจะประกอบด้วย

(1) ค่าคาดหวังแบบมีเงื่อนไขบนข้อมูลอันดับที่ถูกต้องของตัวแปรที่ต้องการศึกษา (Perfect Rank)

$$E\left[Y \mid R(Y)=r_y\right]$$

(2) ค่าคาดหวังแบบมีเงื่อนไขบนข้อมูลอันดับของตัวแปรร่วม (Imperfect Rank)

$$E\left[Y \mid R(X)=r_x\right]$$

(3) ค่าคาดหวังแบบมีเงื่อนไข*โดยนัย*บนข้อมูลอันดับของตัวแปรร่วม (Implied Rank)

$$E\left[Y \mid R(Y) = r_x\right]$$

- *เมื่อใช้โมเดลที่รวมผลของระยะห่างของค่าในเวกเตอร์ค่าเฉลี่ย*

หากพิจารณาระยะห่างของค่าในเวกเตอร์ค่าเฉลี่ย ตามสมการ (3.2) โดยตัวแปรร่วมที่จะพิจารณาคือ $V = \lambda X + \mu$ จะได้ค่าคาดหวังแบบมีเงื่อนไขบนข้อมูลอันดับของตัวแปรร่วม (Imperfect Rank)

$$\begin{aligned} E\left[Y \mid R(V) = r_v\right] &= E\left[V + \sqrt{1-\lambda^2}\varepsilon \mid R(V) = r_v\right] \\ &= E\left[V \mid R(V) = r_v\right] + \sqrt{1-\lambda^2}E\left[\varepsilon \mid R(V) = r_v\right] \end{aligned}$$

ภายใต้ข้อกำหนดที่ว่าตัวแปรร่วม (V) และค่าความผิดพลาด (ε) เป็นอิสระต่อกัน และ $E(\varepsilon) = 0$ แล้ว จะได้ว่า

$$E\left[Y \mid R(V) = r_v\right] = E\left[V \mid R(V) = r_v\right] \tag{3.4}$$

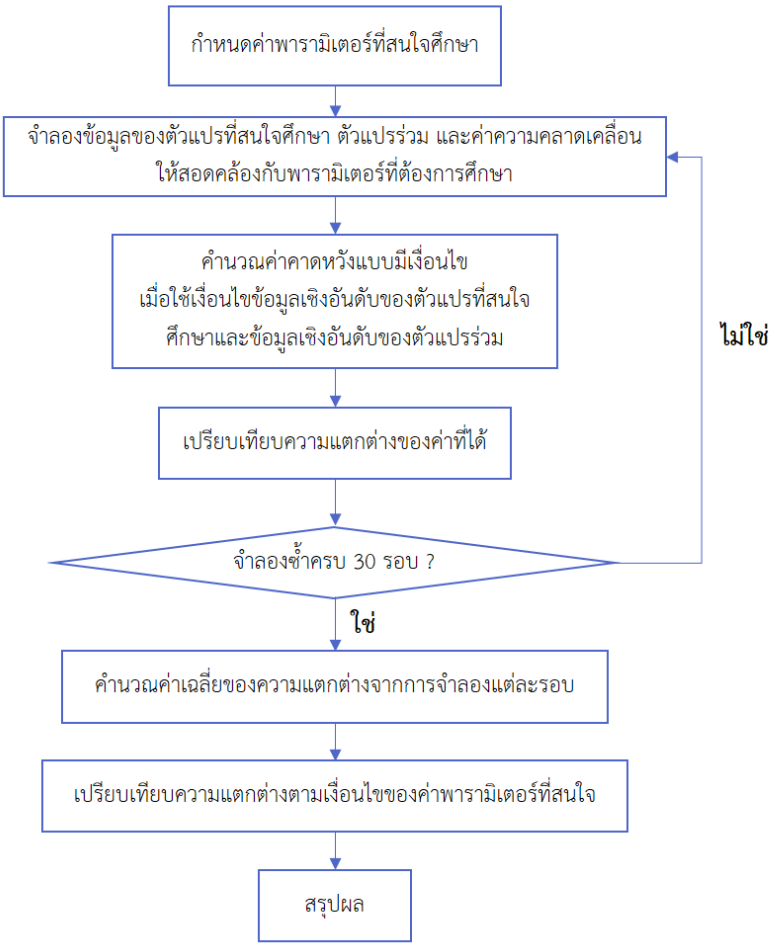
และค่าคาดหวังแบบมีเงื่อนไข*โดยนัย*บนข้อมูลอันดับของตัวแปรร่วม (Implied Rank) ในกรณีนี้ให้แทนด้วยสัญลักษณ์ $E\left[Y \mid R(Y) = r_v\right]$

3.1.4 การเปรียบเทียบค่าคาดหวัง

การเปรียบเทียบผลต่างของค่าคาดหวังที่ได้ในแต่ละกรณีจะใช้ค่า Euclidean distance เนื่องจากค่าคาดหวังแบบมีเงื่อนไขของข้อมูลอันดับที่ได้จะเป็นเวกเตอร์ ดังนั้นการพิจารณาความแตกต่างของค่าในเวกเตอร์ดังกล่าว สามารถดูจาก Euclidean distance ซึ่งคำนวณตามสมการด้านล่างนี้

$$d(y, \hat{y}) = \sqrt{(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + \ldots + (y_n - \hat{y}_n)^2} \tag{3.5}$$

สำหรับการจำลองแต่ละเงื่อนไขจะเปรียบเทียบระยะห่างระหว่างค่าของตัวแปรที่สนใจศึกษา (y) กับค่าคาดหวัง ($\hat{y}$) ที่คำนวณได้กรณีที่เป็น Perfect Rank, Imperfect Rank และ Implied Rank ตามลำดับ โดยในแต่ละเงื่อนไขจะจำลองข้อมูล 30 รอบ แล้วคำนวณค่าเฉลี่ยของระยะห่างที่ได้มาเปรียบเทียบกัน วิธีการจำลองสามารถแสดงได้ตามรูปที่ 3.1



รูปที่ 3.1 ขั้นตอนการศึกษาความแตกต่างระหว่างค่าคาดหวังที่มีเงื่อนไขของข้อมูลอันดับของตัวแปรที่สนใจศึกษากับเมื่อใช้ข้อมูลอันดับจากตัวแปรร่วมในเงื่อนไขต่างๆ

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

3.1.5 การจัดพอร์ตการลงทุนโดยใช้ข้อมูลจำลอง

เพื่อให้เข้าใจความสัมพันธ์ของระดับความสมบูรณ์ของข้อมูลอันดับของตัวแปรร่วม (λ) ที่มีต่อประสิทธิภาพของการจัดพอร์ตการลงทุนก่อนที่จะไปทดลองกับข้อมูลจริง ผู้วิจัยจึงทดลองจำลองข้อมูล โดยเลือกใช้เงื่อนไขที่ใกล้เคียงกับข้อมูลจริงในตลาดหลักทรัพย์ ดังนี้

- จำนวนตัวแปรที่สนใจศึกษา (N) เป็น 10 เนื่องจากใกล้เคียงกับจำนวนกลุ่มอุตสาหกรรมในตลาดหลักทรัพย์จริง รวมทั้งยังเป็นจำนวนที่ไม่มากเกินไปที่นักลงทุนน่าจะสามารถจัดอันดับได้อย่างถูกต้องพอสมควร
- ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรร่วม มีค่าเป็น 0.5 และ 0.75 ซึ่งใกล้เคียงกับระดับความสัมพันธ์ของกลุ่มอุตสาหกรรมจริงของตลาด

วิธีการศึกษาเริ่มจาก

(1) จำลองข้อมูลในกรณีต่อไปนี้

ปัจจัยที่ต้องการศึกษา	ค่าที่จะใช้ในการจำลอง
จำนวนตัวแปร (N)	10
ระดับความสมบูรณ์ของข้อมูลอันดับของตัวแปรร่วม (λ)	0.25, 0.5, 0.75 และ 0.9
ค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรร่วม (ρ)	0.5 และ 0.75
ระยะห่างของค่าในเวกเตอร์ค่าเฉลี่ย	0

(2) หาค่าคาดหวังโดยมีเงื่อนไขของข้อมูลอันดับทั้ง 3 วิธี เช่นเดียวกับในงานศึกษาส่วนแรก

(3) นำค่าดังกล่าวมาจัดพอร์ตการลงทุน โดยหาน้ำหนักการลงทุนที่เหมาะสมตามหลักของ Markowitz และ Black-Litterman คือ

$$\max_{w^*} u_p = y^T w_p - \frac{\gamma}{2} {w_p}^T \Sigma w_p$$

งานวิจัยนี้ใช้ค่าสัมประสิทธิ์การหลีกเลี่ยงความเสี่ยง (Risk Aversion Coefficient) เท่ากับ 4 เป็นค่าที่แสดงว่านักลงทุนยอมรับความเสี่ยงได้ในระดับปานกลาง

(4) เปรียบเทียบค่าเฉลี่ยของอรรถประโยชน์ที่ได้จากการจำลอง 30 รอบ

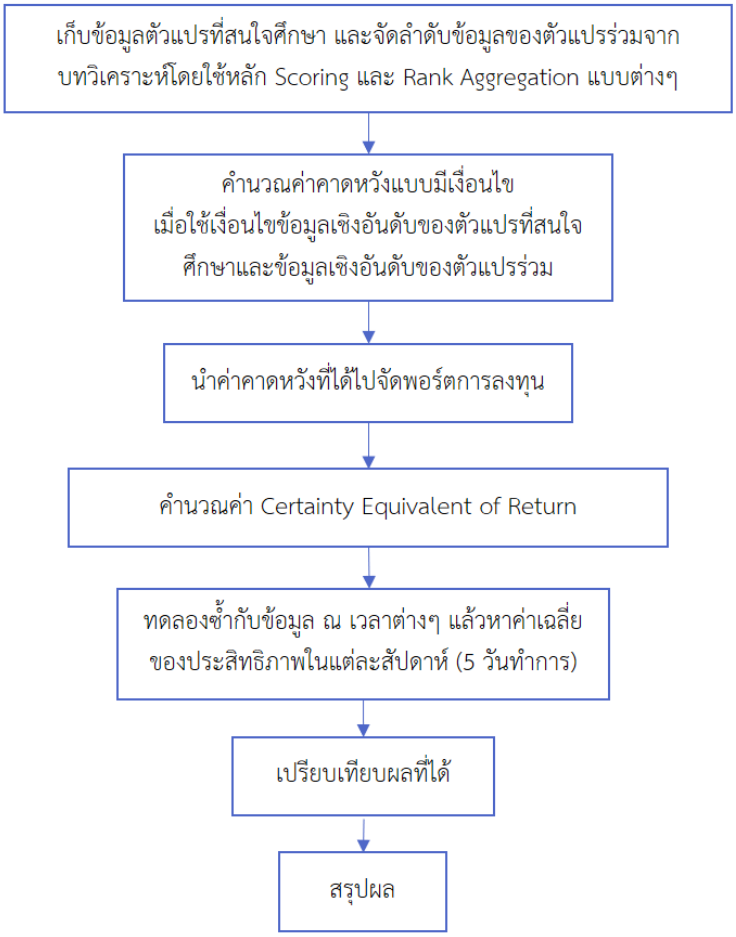
โดยนอกจากจะเปรียบเทียบที่ได้จากค่าคาดหวังทั้ง 3 แบบข้างต้นแล้ว ผู้วิจัยได้คำนวณค่าอรรถประโยชน์เพิ่มเติมอีก 2 กรณี คือ กรณีที่ใช้ผลตอบแทนจริง (Clairvoyance) ซึ่งแม้ว่าในทางปฏิบัติการที่จะทราบผลตอบแทนที่ถูกต้องเป็นไปได้ยากมาก แต่เราคำนวณค่าดังกล่าวเพื่อใช้อ้างอิงค่าอรรถประโยชน์สูงสุดที่เกิดขึ้นได้ภายใต้เงื่อนไขที่กำหนด (Upper-bound) และกรณีสุดท้ายใช้ค่าจากการแจกแจงก่อน ซึ่งพิจารณาเฉพาะข้อมูลในอดีตโดยไม่ใช้ข้อมูลอันดับมาประกอบการตัดสินใจ รวมเป็น 5 กรณี สรุปได้ดังนี้

1. ผลตอบแทนจริง (Clairvoyance)
2. ค่าคาดหวังแบบมีเงื่อนไขบนข้อมูลอันดับที่สมบูรณ์ของผลตอบแทน (Perfect Rank)
3. ค่าคาดหวังแบบมีเงื่อนไขบนข้อมูลอันดับของตัวแปรร่วม (Imperfect Rank)
4. ค่าคาดหวังแบบมีเงื่อนไขโดยนัยของข้อมูลอันดับของตัวแปรร่วม (Implied Rank)
5. ค่าเฉลี่ยของผลตอบแทนในอดีต (Prior)

ผลการศึกษาโดยใช้ข้อมูลจำลองในส่วนนี้จะทำให้เห็นว่าตัวแปรร่วมและตัวแปรที่สนใจศึกษาควรมีความสัมพันธ์กันในระดับใดจึงจะทำให้การจัดพอร์ตการลงทุนมีประสิทธิภาพมากกว่าการใช้ข้อมูลจากการแจกแจงก่อน (Prior) เพียงอย่างเดียว

3.2 การศึกษาโดยใช้ข้อมูลจริง

ในส่วนที่สองนี้จะทดลองจัดพอร์ตการลงทุน โดยใช้ข้อมูลผลตอบแทนของกลุ่มอุตสาหกรรมในตลาดหลักทรัพย์แห่งประเทศไทยกับข้อมูลอันดับที่ได้จากการรวบรวมข้อมูลคำแนะนำการซื้อขายหุ้น (Stock Picks) จากบทวิเคราะห์ของบริษัทหลักทรัพย์ต่างๆ ที่เผยแพร่เป็นประจำทุกวัน นำมาถ่วงน้ำหนักและจัดอันดับด้วยวิธีการต่างๆ แล้วเปรียบเทียบประสิทธิภาพที่ได้ ขั้นตอนการศึกษาในส่วนนี้แสดงได้ดังรูป



รูปที่ 3.2 ขั้นตอนการศึกษาประสิทธิภาพของการจัดพอร์ตการลงทุนเมื่อใช้ข้อมูลจริง

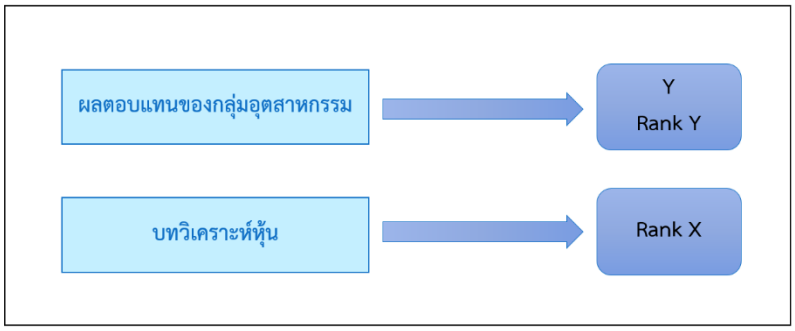
3.2.1 ลักษณะและที่มาของข้อมูล

งานวิจัยนี้จะพิจารณาหุ้นเป็นกลุ่มอุตสาหกรรม เนื่องจากหุ้นในตลาดหุ้นไทยในปัจจุบันมีมากกว่า 500 ตัว หากพิจารณาหุ้นรายตัว แล้วมาจัดอันดับ โอกาสที่อันดับจะตรงกับอันดับจริงจะเป็นไปได้ยาก และอาจเห็นผลการศึกษาไม่ชัดเจน นอกจากนี้ เหตุผลอีกประการคือ ในขั้นตอนต่อไป จะศึกษาโดยการรวบรวมบทวิเคราะห์หุ้น เพื่อมาใช้เป็นตัวแทนของการใช้ข้อมูลอันดับของตัวแปรร่วม แต่ในแต่ละวันจะมีหุ้นเพียงไม่กี่ตัวที่จะได้รับคำแนะนำให้ซื้อ นอกจากนี้ หากเป็นการวิเคราะห์จากข่าวที่มีผลกระทบทั้งกลุ่มอุตสาหกรรม เช่น กรณีรัฐบาลออกมาตรการกระตุ้นเศรษฐกิจ “ช็อปช่วยชาติ” หรือกรณีที่มีการลดกำลังการผลิตน้ำมัน หุ้นหลักๆ ในกลุ่มอุตสาหกรรมที่เกี่ยวข้องจะมีแนวโน้มที่จะถูกแนะนำให้ซื้อหลายตัว ดังนั้น เพื่อให้การวิเคราะห์ทำได้ง่ายขึ้น จึงจะวิเคราะห์หุ้นเป็นรายอุตสาหกรรม ตามที่ตลาดหลักทรัพย์แห่งประเทศไทยได้แบ่งหุ้นออกเป็น 8 กลุ่มอุตสาหกรรมหลัก และ 28 กลุ่มอุตสาหกรรมย่อย ดังต่อไปนี้

1	AGRO	**เกษตรและอุตสาหกรรมอาหาร**
	AGRI	ธุรกิจการเกษตร
	FOOD	อาหารและเครื่องดื่ม
2	CONSUMP	**สินค้าอุปโภคบริโภค**
	FASHION	แฟชั่น
	HOME	ของใช้ในครัวเรือนและสำนักงาน
	PERSON	ของใช้ส่วนตัวและเวชภัณฑ์
3	FINCIAL	**ธุรกิจการเงิน**
	BANK	ธนาคาร
	FIN	เงินทุนและหลักทรัพย์
	INSUR	ประกันภัยและประกันชีวิต
4	INDUS	**สินค้าอุตสาหกรรม**
	AUTO	ยานยนต์
	IMM	วัสดุอุตสาหกรรมและเครื่องจักร
	PAPER	กระดาษและวัสดุการพิมพ์
	PETRO	ปิโตรเคมีและเคมีภัณฑ์
	PKG	บรรจุภัณฑ์
	STEEL	เหล็ก
5	PROPCON	**อสังหาริมทรัพย์และก่อสร้าง**
	CONMAT	วัสดุก่อสร้าง
	PROP	พัฒนาอสังหาริมทรัพย์
	PF&REIT	กองทุนรวมอสังหาริมทรัพย์และกองทรัสต์เพื่อการลงทุนในอสังหาริมทรัพย์
	CONS	บริการรับเหมาก่อสร้าง
6	RESOURC	**ทรัพยากร**
	ENERG	พลังงานและสาธารณูปโภค
	MINE	เหมืองแร่
7	SERVICE	**บริการ**
	COMM	พาณิชย์
	HELTH	การแพทย์
	MEDIA	สื่อและสิ่งพิมพ์
	PROF	บริการเฉพาะกิจ
	TOURISM	การท่องเที่ยวและสันทนาการ
	TRANS	ขนส่งและโลจิสติกส์
8	TECH	**เทคโนโลยี**
	ETRON	ชิ้นส่วนอิเล็กทรอนิกส์
	ICT	เทคโนโลยีสารสนเทศและการสื่อสาร

รูปที่ 3.3 รายละเอียดกลุ่มหลักทรัพย์ตามการแบ่งของตลาดหลักทรัพย์แห่งประเทศไทย

ข้อมูลที่ใช้คำนวณเป็นข้อมูลรายวันย้อนหลัง 6 เดือน ตั้งแต่เดือนสิงหาคม พ.ศ. 2559 ถึงเดือนมกราคม พ.ศ. 2560 โดยจะประกอบด้วยข้อมูล 2 ส่วนตามรูปที่ 3.4 คือ ข้อมูลผลตอบแทนของกลุ่มอุตสาหกรรมใช้ข้อมูลที่เผยแพร่ในเว็บไซต์ของตลาดหลักทรัพย์แห่งประเทศไทย (www.settrade.com) เพื่อนำมาหาการแจกแจงก่อนและใช้ข้อมูล ณ เวลา t+1 มาเป็นข้อมูลอันดับที่ถูกต้อง อีกส่วนเป็นข้อมูลที่รวบรวมได้จากบทวิเคราะห์ นำมาใช้เป็นข้อมูลอันดับของตัวแปรร่วม ซึ่งตามหลักการของ Crowd Sourcing แล้ว ค่าความคาดเคลื่อนจะน้อยลงเมื่อเรารวบรวมความคิดเห็นและการทำนายจากหลายแหล่งเข้าด้วยกัน โดยจะกล่าวถึงรายละเอียดในหัวข้อถัดไป



รูปที่ 3.4 ที่มาของข้อมูลที่จะนำไปวิเคราะห์

3.2.2 ข้อมูลอันดับจากบทวิเคราะห์

ผู้วิจัยได้รวบรวมคำแนะนำการซื้อขายหุ้นในแต่ละวันซึ่งจากบทวิเคราะห์ภาวะตลาดของบริษัทหลักทรัพย์ต่างๆ ที่เผยแพร่ผ่านเว็บไซต์ www.settrade.com โดยในแต่ละวันจะมีบริษัทหลักทรัพย์ลงบทความไว้ประมาณ 15-25 บทความต่อวัน แบ่งเป็นบทวิเคราะห์หุ้นน่าซื้อในรอบเช้าและรอบบ่าย ในบทวิเคราะห์จะระบุหุ้นที่น่าซื้อจากการวิเคราะห์เชิงเทคนิค และการวิเคราะห์จากข่าวประจำวัน ในที่นี้จะเลือกเก็บข้อมูลเฉพาะคำแนะนำเพื่อซื้อขาย*ระยะสั้น*เท่านั้น ซึ่งในแต่ละบทวิเคราะห์จะระบุหุ้นที่แนะนำให้ซื้อประมาณ 1-15 ตัวต่อบทความ ตามตัวอย่างในรูปที่ 3.5 โดยมากมักไม่ได้ระบุเป็นอันดับแต่จะระบุเป็นรายชื่อหุ้นเท่านั้น จึงต้องนำข้อมูลดังกล่าวมารวบรวมและปรับให้เป็นข้อมูลอันดับก่อนนำไปคำนวณต่อไป

รูปที่ 3.5 ตัวอย่างข้อมูลที่ได้จากบทวิเคราะห์หลักทรัพย์

งานวิจัยนี้จะรวบรวมข้อมูลโดยเลือกเฉพาะบทวิเคราะห์ที่มีการเผยแพร่เป็นประจำสม่ำเสมอ เพื่อให้สามารถประเมินความแม่นยำของการเลือกหุ้นในอดีตได้ ในที่นี้ใช้ข้อมูลที่รวบรวมได้เป็นบทวิเคราะห์จำนวน 22 รายงาน ที่จัดทำโดยบริษัทหลักทรัพย์ 18 แห่ง (บางบริษัทจะออกบทวิเคราะห์ภาคเช้าและภาคบ่ายเป็นคนละรายงานกัน) ข้อมูลที่จัดเก็บเป็นไฟล์บทวิเคราะห์จำนวน 2,547 ไฟล์ จากนั้นจึงดึงข้อมูลว่าแต่ละไฟล์มีคำแนะนำอย่างไรบ้าง

ข้อมูลที่เก็บประกอบด้วย ชื่อหุ้น และประเภทของคำแนะนำ เช่น บางบทวิเคราะห์จะบอกรายละเอียดว่าเป็นการแนะนำให้ซื้อ ขาย ถือ ทยอยสะสม หรือหากเป็นข่าวในบทวิเคราะห์ก็จะระบุว่าข่าวนี้เป็นผลดีหรือผลเสียต่อหุ้นตัวใดบ้าง ซึ่งระดับและทิศทางของคำแนะนำที่แตกต่างกันนี้จะเก็บรวบรวมและนำไปใช้ประโยชน์ในการคำนวณและจัดอันดับในขั้นตอนต่อไปด้วย และอย่างที่ได้กล่าวไปแล้วว่า คำแนะนำที่เก็บข้อมูลนี้จะเลือกเฉพาะคำแนะนำในระยะสั้นเท่านั้น

จากนั้นจึงมาเตรียมข้อมูลให้พร้อมต่อการวิเคราะห์ โดยจะตัดคำแนะนำในกรณีที่หุ้นยังไม่ซื้อขายในตลาดหลักทรัพย์ออก (คำแนะนำให้ซื้อหุ้น IPO) เนื่องจากหุ้นเหล่านี้จะยังไม่มีราคาและมูลค่าอ้างอิงในอดีต ท้ายที่สุดแล้วได้รายการคำแนะนำทั้งหมด 20,924 รายการ (รายละเอียดตามตารางที่ 3.1) ที่จะนำไปผ่านกระบวนการจัดอันดับเพื่อให้ได้เป็นข้อมูลอันดับของตัวแปรร่วมในขั้นต่อไป

ตารางที่ 3.1 จำนวนรายการคำแนะนำในช่วง 6 เดือน แยกตามบทวิเคราะห์

บริษัทหลักทรัพย์	ชื่อรายงาน	จำนวนรายการคำแนะนำ
AIRA Securities	Market Outlook	779
ASL Securities	ASL Daily Market Trend	4,272
Asia Plus Securities	Equity Talks	365
Asia Wealth Securities	Daily Market Outlook	589
Country Group Securities	แนวโน้มตลาดวันนี้	222
Capital Nomuar Securities	Equity Play of the day	264
DBS Vickers Securities	Daily Trading Focus	688
Finansa Syrus	Thailand Daily Focus	712
KGI Securities	Thailand in focus	1,235
Krungsri Securities	Money Wizard	1,192
KTB Securities	Morning Bell	1,570
KT ZMICO Securities	ZOOM	1,821
LH Securities	Trends & Turns	569
MayBank KimEng	หยิบเงิน หยิบทอง	482
MayBank KimEng	Afternoon Note	73
MayBank KimEng	Tiger Intrend Morning	213
Phillip Capital	ครบเครื่องการลงทุนกับฟิลลิป	483
Phillip Capital	ส่องตลาด...ภาคบ่าย	507
RHB Securities	Market News Thailand	1,535
Thanachart Securities	The NEW Early Bird	2,376
Thanachart Securities	Daily Trading Strategy PM	581
UOB KayHian securities	Retail Market Monitor	397
รวม		**20,924**

CHULALONGKORN UNIVERSITY

3.2.3 กระบวนการจัดอันดับ

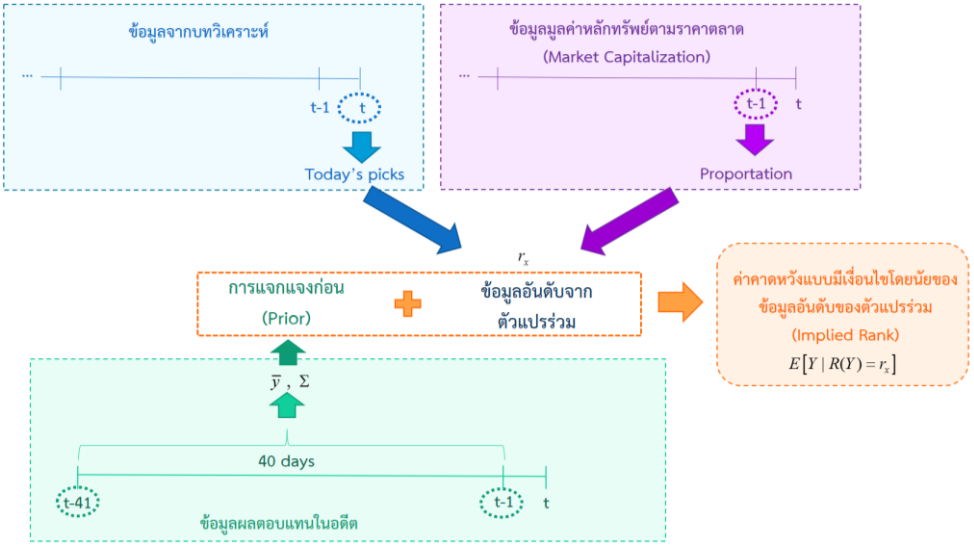
เพื่อให้ได้ข้อมูลอันดับของตัวแปรร่วมไปใช้ในการคำนวณค่าคาดหวังและนำไปทดลองจัดพอร์ตการลงทุน เราจึงต้องสร้างวิธีการให้คะแนนและถ่วงน้ำหนักรายการข้อมูลที่รวบรวมมาได้ โดยเกณฑ์การให้คะแนนที่แตกต่างกัน ก็จะทำให้ได้ผลลัพธ์ที่มีระดับความสัมพันธ์กับผลตอบแทนต่างกันตามไปด้วย วิธีการรวมคะแนนและถ่วงน้ำหนักที่จะใช้ในงานวิจัยนี้ประกอบด้วย

<u>วิธีที่ 1</u> ให้คะแนนตามคำแนะนำแต่ละรายการ

เป็นการนับคะแนนอย่างง่ายตามจำนวนข้อมูลที่ปรากฏ หุ้นตัวใดตัวหนึ่งได้รับคำแนะนำด้านบวกจะได้ 1 คะแนน และรายการที่มีคำแนะนำด้านลบมีคะแนนเป็น -1 คะแนน แนวคิดของวิธีการนี้คือหากมีการแนะนำให้ซื้อหุ้นในกลุ่มอุตสาหกรรมใด ผลตอบแทนของกลุ่มอุตสาหกรรมนั้นจะเพิ่มมากขึ้นตามไปด้วย

<u>วิธีที่ 2</u> ให้คะแนนตามความสำคัญของหุ้นในกลุ่มอุตสาหกรรมนั้นๆ

หุ้นตัวที่ได้รับคำแนะนำให้ซื้อหากคิดเป็นสัดส่วนต่อหุ้นทั้งกลุ่มอุตสาหกรรมแล้วมีสัดส่วนไม่มาก การที่หุ้นตัวดังกล่าวมีราคาสูงขึ้นอาจไม่ได้ทำให้ผลตอบแทนของทั้งกลุ่มอุตสาหกรรมเพิ่มขึ้น จึงถ่วงน้ำหนักคะแนนในวิธีที่ 1 ด้วยสัดส่วนของมูลค่าหุ้นดังกล่าว (Market Capitalization) ต่อมูลค่าหลักทรัพย์ทั้งกลุ่มอุตสาหกรรม โดยในที่นี้จะใช้ข้อมูลสัดส่วนมูลค่าของวันก่อนหน้ามาใช้ถ่วงน้ำหนักส่วนประกอบของข้อมูลที่นำมาใช้คำนวณแสดงได้ตามรูปที่ 3.6



รูปที่ 3.6 ข้อมูลจริงจากแหล่งต่างๆ ที่นำมาคำนวณหาค่าคาดหวังแบบมีเงื่อนไข กรณีถ่วงน้ำหนักด้วยสัดส่วนมูลค่าหุ้นในกลุ่มอุตสาหกรรม

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

<u>วิธีที่ 3</u> ให้คะแนนตามตามระดับความรุนแรงของคำแนะนำ

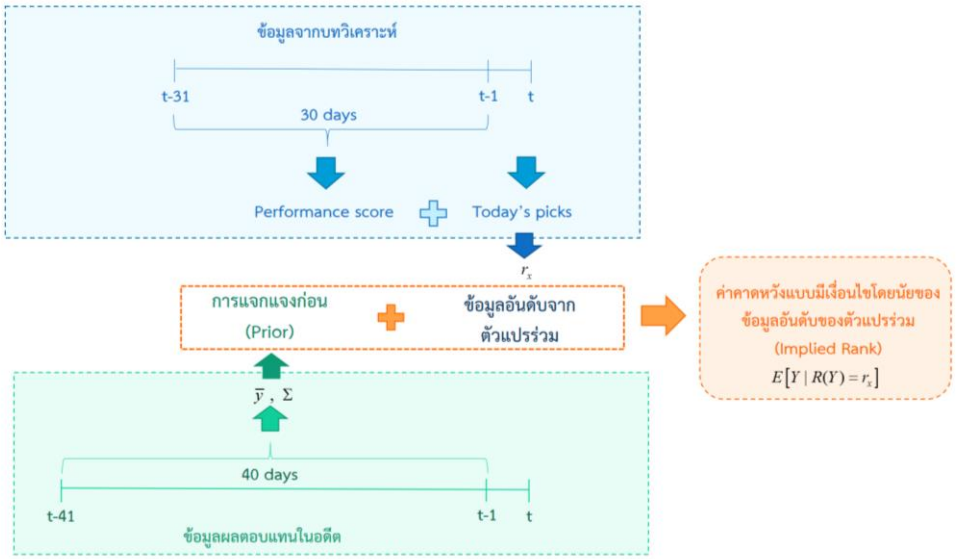
บางบทวิเคราะห์จะให้น้ำหนักกับคำแนะนำแตกต่างกัน เช่น มีการแนะนำให้ซื้อ ทยอยซื้อ สะสม ถือ หรือแนะนำให้ขาย ซึ่งเราสามารถใช้ข้อมูลส่วนนี้มาเป็นประโยชน์ในการถ่วงน้ำหนักได้ แต่จะถ่วงน้ำหนักสำหรับคำแนะนำแต่ละชนิดด้วยค่าเท่าไหร่นั้น ต้องดูตามความเหมาะสม โดยอาจต้องทดลองหลายๆ ค่าแล้วดูว่าค่าชุดใดให้ข้อมูลอันดับที่ใกล้เคียงกับผลตอบแทนจริงมากกว่ากัน

<u>วิธีที่ 4</u> ให้คะแนนตามตามความแม่นยำของการทำนายในอดีต

เนื่องจากมีงานวิจัยก่อนหน้านี้แสดงให้เห็นว่า การนำข้อมูลความแม่นยำของการทำนายในอดีตมาพิจารณา จะทำให้ผลการทำนายมีประสิทธิภาพสูงขึ้น ดังนั้น จึงจะทดลองใช้ข้อมูลผลการวิเคราะห์ในอดีตของแต่ละบริษัทหลักทรัพย์มาพิจารณาร่วมด้วย

วิธีประเมินความแม่นยำของการทำนายทำได้หลายวิธี แต่ในงานวิจัยนี้เลือกใช้ค่าสัมประสิทธิ์สหสัมพันธ์ (Pearson correlation coefficient) เป็นเกณฑ์ เนื่องจากค่าดังกล่าวสามารถบอกได้ทั้งระดับความสัมพันธ์และทิศทางด้วย โดยเราจะถือว่าแหล่งข้อมูลที่มีความสัมพันธ์กับผลตอบแทนน้อยๆ (ดูจากค่าสัมบูรณ์ของค่าสัมประสิทธิ์สหสัมพันธ์) ไม่ควรนำมาใช้ในการพิจารณา ในที่นี้จะใช้ค่าที่ใกล้เคียงศูนย์มาถ่วงน้ำหนัก ส่วนแหล่งข้อมูลที่มีความสัมพันธ์กับผลตอบแทนมาก ไม่ว่าจะเป็นความสัมพันธ์ด้านบวกหรือด้านลบ กล่าวคือ กรณีที่ค่าสัมประสิทธิ์สหสัมพันธ์มีค่าเป็นลบ อีกนัยหนึ่งคือแหล่งข้อมูลที่ทำนายตรงข้ามกับข้อมูลผลตอบแทนตลอด ก็ยังถือว่าเป็นข้อมูลที่มีประโยชน์ในการพิจารณา แต่เราคูณคำแนะนำจากแหล่งนั้นด้วยค่าลบ เพื่อปรับค่าให้สอดคล้องกับผลตอบแทนที่สนใจ

ประเด็นสำคัญอีกประการหนึ่งคือ เราควรประเมินความแม่นยำย้อนหลังไปมากน้อยเพียงใด งานศึกษานี้จึงจะทดลองประเมิน โดยจะประเมินจากผลการทำนาย 10, 20 และ 30 วันย้อนหลัง แล้วเปรียบเทียบผลที่ได้ โดยข้อมูลส่วนต่างๆ ที่นำมาใช้คำนวณหาค่าคาดหวังสรุปได้ตามรูปที่ 3.7 ซึ่งเป็นตัวอย่างกรณีที่ประเมินด้วยข้อมูลย้อนหลังไป 30 วันทำการ



รูปที่ 3.7 ข้อมูลจริงจากแหล่งต่างๆ ที่นำมาประเมินความแม่นยำของคำแนะนำในอดีต และใช้คำนวณหาค่าคาดหวังแบบมีเงื่อนไข

สุดท้ายจะทดลองใช้วิธีการจัดอันดับที่กล่าวมาข้างต้นหลายวิธีมาประกอบกัน เพื่อพยายามสร้างวิธีการรวบรวมข้อมูลและถ่วงน้ำหนักเพื่อให้ได้ข้อมูลอันดับที่ใกล้เคียงกับอันดับที่ถูกต้องมากที่สุด โดยการใช้หลักการคิดคะแนน (Scoring) ที่แตกต่างกันจะให้ข้อมูลอันดับที่แตกต่างกันไปด้วย

3.2.4 การคำนวณผลตอบแทนดุลยภาพ (Equilibrium Returns)

ในงานวิจัยนี้จะใช้วิธี Reverse Optimization เนื่องจากค่าที่ได้จากวิธีนี้จะผันผวนน้อยกว่าการใช้ค่าเฉลี่ยจากผลตอบแทนในอดีต (historical sample mean) การคำนวณทำได้ตามสมการ

$$Equilibrium\ Return\ =\ \frac{\mu_{mkt}}{w_{mkt}^{T}\Sigma w_{mkt}}\Sigma w \qquad (3.6)$$

สำหรับค่าเฉลี่ยของผลตอบแทนของตลาด (μ_{mkt}) หาจากค่าเฉลี่ยของผลตอบแทนของตลาดหลักทรัพย์แห่งประเทศไทย ข้อมูลที่นำมาคำนวณได้จาก SETSMART เป็นข้อมูลระหว่างเดือนสิงหาคม พ.ศ. 2556 ถึงเดือนกรกฎาคม พ.ศ. 2559 และสัดส่วนของกลุ่มหลักทรัพย์ในตลาด (w_{mkt}) หาได้จากมูลค่าหลักทรัพย์ตามราคาตลาด (Market Capitalization) ของกลุ่มหลักทรัพย์ที่สนใจเทียบกับมูลค่าของตลาด

3.2.5 ขั้นตอนการคำนวณ

เริ่มจากการเก็บข้อมูลจริงของตัวแปรที่ต้องการศึกษาและตัวแปรร่วม โดยการจำลองจะทำตามขั้นตอน ดังนี้

(1) ใช้ข้อมูลในอดีตมาคำนวณหาการแจกแจง prior โดยหาค่า mean และ variance-covariance matrix ดังนี้ จาก CAPM model

(2) หาค่าคาดหวังโดยมีเงื่อนไขของข้อมูลอันดับ

(3) นำค่าคาดหวังดังกล่าวมาจัดพอร์ตการลงทุน และประเมินประสิทธิภาพ โดยจะเทียบในกรณีที่

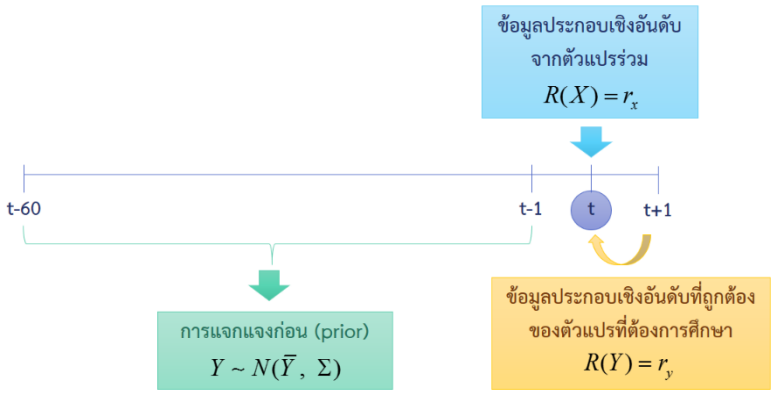
- ใช้เฉพาะข้อมูลของการแจกแจงก่อน Prior ที่ได้จากการคำนวณ Equilibrium Return และจาก One-factor Model
- ใช้ข้อมูลอันดับที่ถูกต้อง (perfect rank) ที่ได้จากการจัดลำดับผลตอบแทนที่เกิดขึ้นจริงในตลาด ณ เวลา t+1
- ใช้ข้อมูลอันดับของตัวแปรร่วม (Implied rank) ที่ได้จากการรวบรวมและจัดอันดับจากคำแนะนำให้ซื้อ-ขายหลักทรัพย์

3.2.6 การใช้ข้อมูลอันดับจากข้อมูลในอนาคต

เนื่องจากเราต้องการเปรียบเทียบผลจากการใช้ค่าคาดหวังบนข้อมูลอันดับที่ถูกต้อง กับข้อมูลอันดับของตัวแปรร่วม ที่จะมีส่วนของค่าความคลาดเคลื่อนประกอบ แต่ในความเป็นจริง ณ เวลาที่

Chulalongkorn University

ตัดสินใจลงทุน เราจะยังไม่ทราบข้อมูลอันดับที่ถูกต้อง ดังนั้น เพื่อให้สามารถศึกษาค่าดังกล่าวได้ จึงจำเป็นต้องใช้ข้อมูลในอนาคต ณ เวลา t+1 มาคำนวณแทน ดังรูป



รูปที่ 3.8 รูปแสดงการใช้ข้อมูลผลตอบแทนในอดีตและในอนาคตมาคำนวณค่าคาดหวัง

3.2.7 การทดลองโดยใช้วิธีการเคลื่อนตัวอย่างการทดลอง (Rolling Sample)

ในการประเมินประสิทธิภาพของจัดพอร์ตการลงทุนจะใช้ข้อมูลรายวัน โดย ณ วันที่พิจารณา ให้

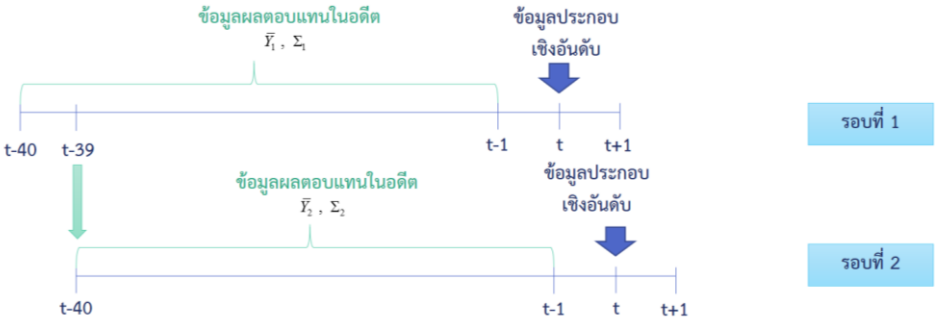
(1) คำนวณพารามิเตอร์ $\bar{Y}, \Sigma$ ของการแจกแจง prior จากข้อมูลผลตอบแทนรายวันของแต่ละกลุ่มอุตสาหกรรมย้อนหลัง 40 วันทำการ

(2) จัดอันดับกลุ่มอุตสาหกรรมโดยใช้ข้อมูลผลตอบแทนในอนาคต ณ เวลา t+1 เพื่อใช้เป็นข้อมูลประกอบเชิงดันดับที่ถูกต้อง (r_y)

(3) จัดอันดับกลุ่มอุตสาหกรรมโดยใช้ข้อมูลจากบทวิเคราะห์หุ้นที่รวบรวมมาได้ และให้คะแนนตามเกณฑ์ต่างๆ ที่กล่าวมาแล้ว เพื่อใช้เป็นข้อมูลประกอบเชิงดันดับ (r_x)

(4) คำนวณหาค่าคาดหวังแบบมีเงื่อนไขของข้อมูลอันดับ แล้วนำค่าดังกล่าวมาใช้หาค่า w_p^* ที่ได้รับอรรถประโยชน์สูงสุด

จากนั้นให้เลื่อนวันที่พิจารณาไปข้างหน้าอีก 1 วัน แล้วทำการคำนวณเช่นเดียวกัน ขั้นสุดท้ายให้นำค่าที่ได้ในแต่ละสัปดาห์ (5 วันทำการ) มาหาค่าเฉลี่ย เพื่อเปรียบเทียบผล โดยประเมินประสิทธิภาพจากค่า Certainty Equivalent of Return (CEQ) ที่ได้จากแต่ละวิธีต่อไป



รูปที่ 3.9 การทดลองโดยใช้วิธีการเคลื่อนตัวอย่างการทดลอง (Rolling Sample)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 4
ผลการศึกษา

4.1 ผลการเปรียบเทียบค่าคาดหวังจากข้อมูลที่จำลองตามเงื่อนไขต่างๆ

ผลการศึกษาส่วนแรกได้การจำลองข้อมูล แล้วเปรียบเทียบค่าเฉลี่ยของระยะห่าง (Euclidean Distance) ระหว่างค่าคาดหวังในกรณีต่างๆ กับค่าตัวแปรที่สนใจศึกษาที่จำลองได้ในแต่ละรอบ โดยศึกษาผลในกรณีที่เงื่อนไขต่อไปนี้เปลี่ยนแปลงไป

4.1.1 จำนวนตัวแปร

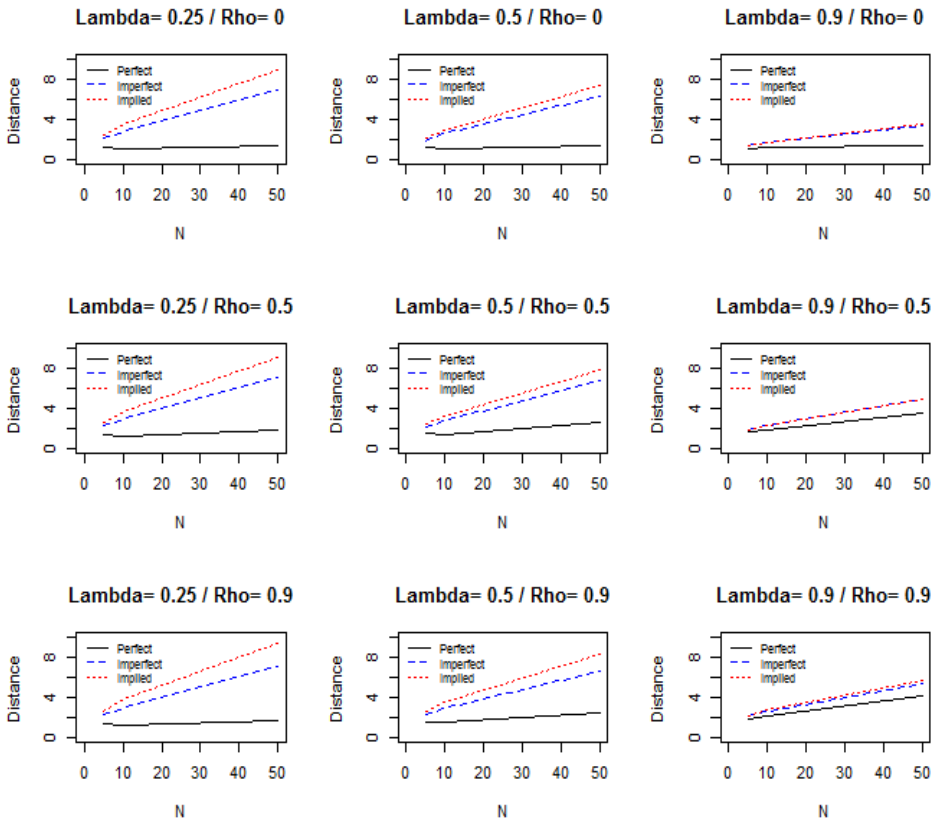
เมื่อจำลองข้อมูลโดยกำหนดให้จำนวนตัวแปร (N) ที่สนใจศึกษามีค่าแตกต่างกัน ในที่นี้ให้มีค่าเป็น 5, 10 และ 50 ตามลำดับ ผลของระยะห่างเฉลี่ยระหว่างค่าของตัวแปรที่สนใจศึกษากับค่าคาดหวังภายใต้เงื่อนไขแบบต่างๆ เมื่อจำนวนตัวแปรแตกต่างกันแสดงได้ตามรูปที่ 4.1 โดยกราฟย่อยแต่ละกราฟจะเป็นการพิจารณาเมื่อค่าระดับความสมบูรณ์ข้อมูลอันดับของตัวแปรร่วม (λ) และค่าสัมประสิทธิ์สหสัมพันธ์ (ρ) แตกต่างกัน ตามที่ระบุไว้ในแต่ละกราฟ จากผลการจำลองพบข้อสรุปที่น่าสนใจในประเด็นต่อไปนี้

ประการแรก ค่าคาดหวังแบบมีเงื่อนไขบนข้อมูลอันดับที่สมบูรณ์ (Perfect Rank) จะมีค่าใกล้เคียงกับค่าจริงที่สุด รองลงมาเป็นค่าคาดหวังที่คำนวณบนข้อมูลอันดับของของตัวแปรร่วม (Imperfect Rank) และค่าคาดหวังแบบมีเงื่อนไขโดยนัยบนข้อมูลอันดับของตัวแปรร่วม (Implied Rank) มีค่าแตกต่างจากค่าของตัวแปรที่สนใจศึกษามากที่สุด และเป็นเช่นนี้สำหรับทุกเงื่อนไข และเมื่อจำนวนตัวแปรที่สนใจศึกษามีค่าเพิ่มขึ้นความแตกต่างของ Euclidean Distance ระหว่างค่าคาดหวังในกรณีต่างๆ กับค่าตัวแปรที่สนใจศึกษา (y) มีแนวโน้มที่จะเพิ่มมากขึ้นตามไปด้วย

ประเด็นถัดมา เมื่อพิจารณาผลของจำนวนตัวแปรควบคู่ไปกับผลของค่าระดับความสมบูรณ์ข้อมูลอันดับของตัวแปรร่วม (λ) และค่าสัมประสิทธิ์สหสัมพันธ์ (ρ) จะเห็นว่าความแตกต่างของระยะห่างระหว่างค่าคาดหวังในกรณีต่างๆ กับค่าตัวแปรที่สนใจศึกษา (y) จำนวนตัวแปรที่เพิ่มขึ้น นั้นจะยิ่งแตกต่างกันอย่างชัดเจนเมื่อค่าระดับความสมบูรณ์ข้อมูลอันดับของตัวแปรร่วม (λ) มีค่าน้อยและค่าสัมประสิทธิ์สหสัมพันธ์ (ρ) มีค่ามาก

สุดท้ายสังเกตว่าแม้จำนวนตัวแปรที่สนใจศึกษาจะเพิ่มมากขึ้น แต่หากค่าระดับความสมบูรณ์ของข้อมูลอันดับของตัวแปรร่วม (λ) มีค่าสูง ระยะห่างระหว่างค่าจริงกับค่าคาดหวังที่ได้จากทั้ง 3 กรณีจะมีค่าใกล้เคียงกัน โดยเฉพาะกรณีของค่าคาดหวังที่คำนวณบนข้อมูลอันดับของของตัวแปรร่วม (Imperfect Rank) และค่าคาดหวังแบบมีเงื่อนไขโดยนัยบนข้อมูลอันดับของตัวแปรร่วม (Implied

Rank) ที่จากรูปที่ 4.1 คอลัมน์ขวาสุด เมื่อ λ =0.9 เส้นกราฟของ Imperfect Rank และ Implied Rank จะใกล้เคียงกันมาก



รูปที่ 4.1 ระยะห่างระหว่างค่าตัวแปรที่สนใจศึกษากับค่าคาดหวังในกรณีต่างๆ เมื่อจำนวนตัวแปรเปลี่ยนแปลงไป

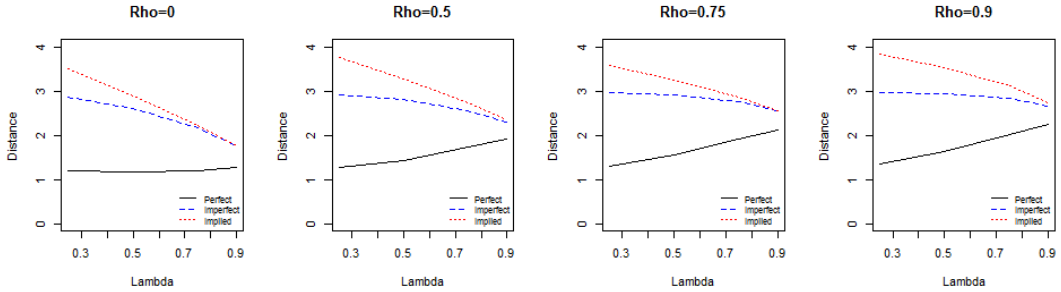
จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

4.1.2 ระดับความสมบูรณ์ของข้อมูลอันดับของตัวแปรร่วม

เมื่อระดับความสมบูรณ์ของข้อมูลอันดับของตัวแปรร่วม (λ) เพิ่มขึ้น ส่งผลโดยตรงให้อันดับของตัวแปรร่วมมีโอกาสสูงขึ้นที่จะถูกต้องตรงกับอันดับของตัวแปรที่สนใจศึกษา สอดคล้องกับผลการจำลองที่พบว่าระยะห่างระหว่างค่าจริงกับค่าคาดหวังในกรณีต่างๆ มีค่าใกล้เคียงกันมากขึ้นด้วย และจะเห็นว่าระยะห่างระหว่างค่าจริงกับค่าคาดหวังที่คำนวณบนข้อมูลอันดับของของตัวแปรร่วม (Imperfect Rank) และค่าคาดหวังแบบมีเงื่อนไขโดยนัยบนข้อมูลอันดับของตัวแปรร่วม (Implied Rank) แตกต่างกันน้อยมาก

ผลการจำลองตามรูปที่ 4.2 แสดงผลในกรณีที่จำนวนข้อมูลเท่ากับ 10 อย่างไรก็ตาม ผลการจำลองเมื่อใช้จำนวนตัวแปร (N) ที่มีค่าต่างกันก็ยังได้ผลของความสัมพันธ์ในลักษณะเดียวกับที่อธิบายไปข้างต้นเช่นเดียวกัน กราฟย่อยในรูปนี้เป็นผลการจำลองเมื่อค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรร่วมแตกต่างกัน ซึ่งจะเห็นได้ว่ารูปแบบความสัมพันธ์ของระยะห่างที่ลดลงเมื่อระดับความ

สมบูรณ์ของข้อมูลอันดับของตัวแปรร่วม (λ) เพิ่มขึ้น เป็นผลลักษณะเดียวกันสำหรับทุกค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรร่วม (ρ)

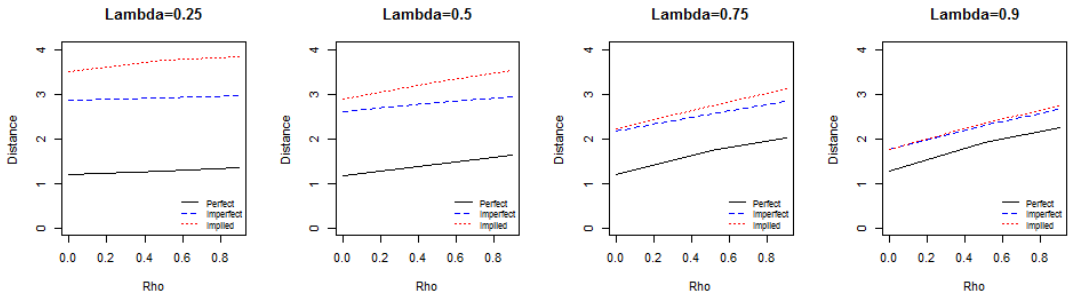


รูปที่ 4.2 ระยะห่างระหว่างค่าตัวแปรที่สนใจศึกษากับค่าคาดหวังในกรณีต่างๆ เมื่อระดับความสมบูรณ์ของข้อมูลอันดับของตัวแปรร่วมเปลี่ยนแปลงไป

4.1.3 ค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรร่วม

รูปที่ 4.3 เป็นผลการจำลองเมื่อใช้ค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรร่วม (ρ) เป็น 0, 0.5 และ 0.9 คือ ตัวแปรร่วมแต่ละตัวไม่มีความสัมพันธ์กัน สัมพันธ์กันปานกลาง และระดับสูง ตามลำดับ ผลการคำนวณพบว่า เมื่อค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรร่วม (ρ) เพิ่มขึ้น ระยะห่างระหว่างค่าจริงกับค่าคาดหวังทั้ง 3 แบบมีแนวโน้มเพิ่มขึ้น โดยความสัมพันธ์ดังกล่าวจะเห็นได้ชัดหากระดับความสมบูรณ์ของข้อมูลอันดับของตัวแปรร่วม (λ) มีค่ามาก เห็นได้จากความชันของกราฟย่อยด้านขวา ที่ λ มีค่าสูง จะเห็นว่าระยะห่างจากระหว่างค่าจริงกับค่าคาดหวังทั้ง 3 รูปแบบจะเพิ่มขึ้นในอัตราสูงกว่ากราฟย่อยด้านซ้ายที่เป็นการจำลองเมื่อ λ มีค่าน้อย

นอกจากนี้ ยังสังเกตเห็นด้วยว่า กรณีที่ระดับความสมบูรณ์ของข้อมูลอันดับของตัวแปรร่วม (λ) มีค่าสูง เส้นกราฟของระยะห่างกรณีที่ใช้ข้อมูลอันดับของของตัวแปรร่วม (Imperfect Rank) และกรณีของเงื่อนไขโดยนัยบนข้อมูลอันดับของตัวแปรร่วม (Implied Rank) อยู่ใกล้กันมาก โดยเฉพาะเมื่อ λ =0.9 เส้นกราฟแทบจะซ้อนทับกัน สอดคล้องกับผลการศึกษาที่ได้อธิบายไปในหัวข้อ 4.1.2

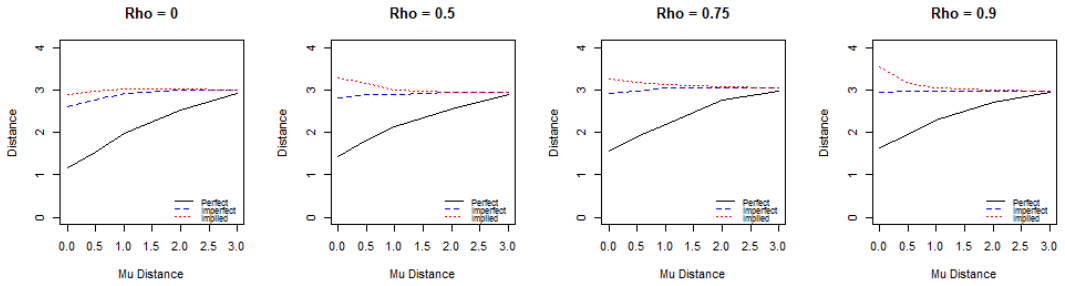


รูปที่ 4.3 ระยะห่างระหว่างค่าตัวแปรที่สนใจศึกษากับค่าคาดหวังในกรณีต่างๆ เมื่อค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรร่วมเปลี่ยนแปลงไป

4.1.4 ระยะห่างระหว่างค่าในเวกเตอร์ค่าเฉลี่ย

การจำลองก่อนหน้านี้กำหนดให้ตัวแปรต่างๆ ถูกปรับให้เป็นค่ามาตรฐาน ที่มีเวกเตอร์ค่าเฉลี่ยเป็นศูนย์ ในหัวข้อนี้จะปรับโมเดลการจำลองเพื่อศึกษาว่า หากตัวแปรแต่ละตัวมีค่าเฉลี่ยแตกต่างกัน ระยะห่างของค่าในเวกเตอร์ค่าเฉลี่ยจะมีผลอย่างไรต่อค่าคาดหวังแบบต่างๆที่คำนวณได้ ในที่นี้เลือกจำลองกรณีที่ไม่ซับซ้อนมากนัก คือ กำหนดให้ค่าเฉลี่ยของแต่ละตัวแปรมีระยะห่างเท่าๆ กัน

หากระยะห่างของค่าในเวคเตอร์ค่าเฉลี่ยของตัวแปรมีค่ามากขึ้นข้อมูลอันดับของตัวแปรร่วมจะมีผลให้ความถูกต้องมากขึ้น จึงพิจารณาผลความแตกต่างของค่าเฉลี่ย กรณีที่ระดับความสมบูรณ์ของข้อมูลอันดับของตัวแปรร่วม มีค่าในระดับปานกลาง คือ เท่ากับ 0.5 แล้วดูว่าหากเพิ่มระยะห่างของค่าเฉลี่ยจาก 0 เป็น 0.5, 1, 2 และ 3 ตามลำดับแล้วจะทำให้ค่าคาดหวังในกรณีต่างๆ เปลี่ยนแปลงไปอย่างไรบ้าง ซึ่งผลที่ได้ตามรูปที่ 4.4 ก็เป็นไปตามสมมติฐานที่คาดไว้ คือ เมื่อระยะห่างของค่าเฉลี่ยเพิ่มขึ้น ระยะห่างของค่าจริงกับค่าคาดหวังทั้ง 3 กรณีจะมีค่าใกล้เคียงกันมากขึ้น โดยเฉพาะอย่างยิ่งกรณีของข้อมูลอันดับของตัวแปรร่วม (Imperfect) กับกรณีเงื่อนไขโดยนัยของตัวแปรร่วม (Implied) ที่ระยะห่างจะใกล้เคียงกันมากอย่างเห็นได้ชัด และลักษณะความสัมพันธ์ดังกล่าวไม่แตกต่างกันมากนักแม้ค่าสัมประสิทธิ์สหสัมพันธ์จะเปลี่ยนแปลงไป



รูปที่ 4.4 ระยะห่างระหว่างค่าของตัวแปรที่สนใจศึกษากับค่าคาดหวังในกรณีต่างๆ เมื่อระยะห่างของค่าในเวคเตอร์ค่าเฉลี่ยเปลี่ยนแปลงไป ณ ระดับความสมบูรณ์ของข้อมูลอันดับของตัวแปรร่วมเท่ากับ 0.5 และจำนวนตัวแปรที่สนใจศึกษามี 10 ตัวแปร

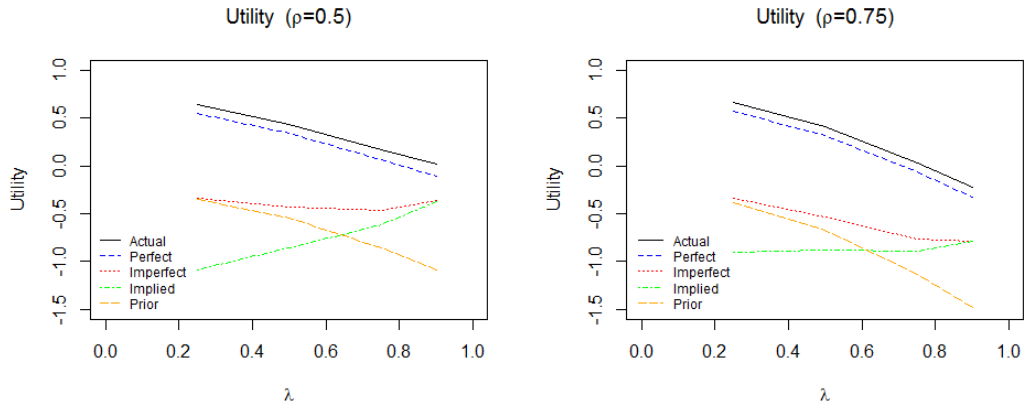
4.2 ผลการจัดพอร์ตการลงทุนจากข้อมูลจำลอง

เมื่อใช้ข้อมูลที่จำลองได้จากการศึกษาในส่วนแรกมาจัดพอร์ตการลงทุนสำหรับ 10 กลุ่มหลักทรัพย์ โดยให้มีการแจกแจงแบบปกติมาตรฐาน ใช้ค่าสัมประสิทธิ์สหสัมพันธ์เป็น 0.5 และ 0.75 เนื่องจากค่าดังกล่าวใกล้เคียงกับระดับความสัมพันธ์จริงของกลุ่มหลักทรัพย์ในตลาด และกำหนดให้นักลงทุนยอมรับความเสี่ยงได้ในระดับปานกลาง คือ ใช้ค่าค่าสัมประสิทธิ์การหลีกเลี่ยงความเสี่ยง (γ) เท่ากับ 4

ผลการจัดพอร์ตการลงทุนตามรูปที่ 4.5 พบว่า การใช้ค่าคาดหวังแบบมีเงื่อนไขบนข้อมูลอันดับที่สมบูรณ์ (Perfect Rank) ให้อรรถประโยชน์เฉลี่ยสูงที่สุด และค่าที่ได้ใกล้เคียงกับการใช้ผลตอบแทนจริงที่ถูกต้อง (Actual Return)

หากใช้ค่าคาดหวังแบบมีเงื่อนไขบนข้อมูลอันดับของตัวแปรร่วม (Imperfect Rank) มาคำนวณสัดส่วนการลงทุนที่เหมาะสมพบว่า ค่าอรรถประโยชน์ที่ได้จะน้อยกว่าการใช้ข้อมูลอันดับที่สมบูรณ์แต่ก็ยังมากกว่าอรรถประโยชน์จากการใช้ข้อมูลของผลตอบแทนในอดีต (Prior)

ส่วนการใช้ค่าคาดหวังแบบมีเงื่อนไข*โดยนัย*ของข้อมูลอันดับของตัวแปรร่วม (Implied Rank) นั้น จะได้ผลดีกว่าการใช้ข้อมูลของผลตอบแทนในอดีต (Prior) ก็ต่อเมื่อระดับความสมบูรณ์ของข้อมูลอันดับของตัวแปรร่วม (λ) มีค่าสูงมากพอ ในรูปที่ 4.5 หาก λ มีค่ามากกว่า 0.6 การใช้ข้อมูลอันดับมาประกอบการตัดสินใจลงทุนก็ยังให้อรรถประโยชน์เพิ่มจากการใช้ข้อมูลในอดีตเพียงอย่างเดียว นอกจากนี้ จะเห็นได้ว่า การใช้ค่าคาดหวังแบบมีเงื่อนไขบนข้อมูลอันดับของตัวแปรร่วม (Imperfect Rank) จะได้ผลดีกว่าการใช้เงื่อนไข*โดยนัย*ของข้อมูลอันดับของตัวแปรร่วม (Implied Rank) เสมอด้วย



รูปที่ 4.5 อรรถประโยชน์ที่ได้รับจากการลงทุนในสัดส่วนที่คำนวณได้จากการใช้ค่าผลตอบแทนแบบต่างๆ

4.3 ผลการจัดพอร์ตการลงทุนจากข้อมูลจริง

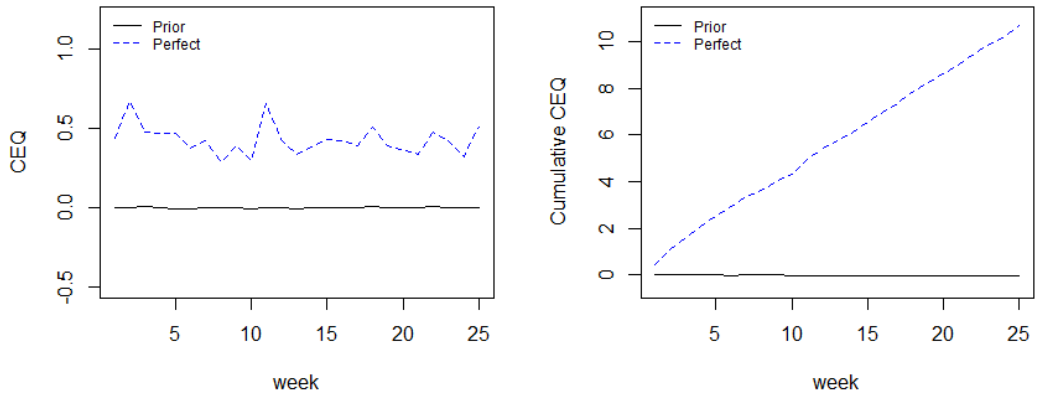
งานศึกษาในส่วนนี้จะใช้ข้อมูลจริงที่รวบรวมได้จากเว็บไซต์ของตลาดหลักทรัพย์แห่งประเทศไทย มาทดลองจัดพอร์ตการลงทุน ในช่วงเวลาระหว่างเดือนสิงหาคม พ.ศ. 2559 ถึงเดือนมกราคม พ.ศ. 2560 (รวม 126 วันทำการ) แล้วประเมินประสิทธิภาพของการจัดพอร์ตการลงทุน โดยพิจารณาจากค่าอรรถประโยชน์ที่ได้รับในแต่ละสัปดาห์ (5 วันทำการ)

4.3.1 ผลเมื่อใช้ข้อมูลอันดับที่ถูกต้อง

เริ่มจากทดลองจัดพอร์ตการลงทุนโดยเปรียบเทียบประสิทธิภาพเมื่อใช้ข้อมูลจากการแจกแจงก่อน (prior) กับเมื่อใช้ข้อมูลอันดับที่ถูกต้อง (Perfect Rank) มาประกอบการคำนวณ โดยข้อมูลอันดับที่ถูกต้องได้จากการจัดอันดับผลตอบแทนจริงของวันทำการถัดไป (t+1) ได้ผลตามรูปที่ 4.6 ซึ่งแสดงให้เห็นว่าการใช้ค่าคาดหวังแบบมีเงื่อนไขของข้อมูลอันดับที่ถูกต้องมาคำนวณหาน้ำหนักการลงทุนที่เหมาะสมของกลุ่มอุตสาหกรรมหลัก 8 กลุ่ม จะได้รับค่าอรรถประโยชน์ (certainty equivalent return) สูงกว่าการใช้เพียงข้อมูลจากการแจกแจงก่อนอย่างเห็นได้ชัดในทุกช่วงเวลา และเมื่อพิจารณากราฟทางขวาซึ่งแสดงค่าอรรถประโยชน์สะสม (cumulative certainty equivalent return) ก็พบว่าเมื่อใช้ข้อมูลอันดับที่ถูกต้อง จะได้ค่าอรรถประโยชน์สะสมสูงกว่ากรณีที่ใช้เฉพาะข้อมูลของการแจกแจงก่อนเพียงอย่างเดียวมาก

ผลที่ได้นี้สอดคล้องกับผลการศึกษาของ Chiarawongse et al. (2012) สุพัตรา (2558) และ ปริยากร (2558) ซึ่งแม้ว่าข้อมูลที่ใช้ศึกษาจะแตกต่างกันในแต่ละงาน แต่ได้ข้อสรุปเช่นเดียวกันว่าการใช้ข้อมูลอันดับที่สมบูรณ์สามารถเพิ่มประสิทธิภาพของการจัดพอร์ตการลงทุนได้

อย่างไรก็ตามในการจำลองข้างต้นข้อมูลอันดับที่ถูกต้องได้จากลำดับของค่าสังเกตของผลตอบแทนในอนาคต (t+1) ซึ่งในทางปฏิบัติ ณ เวลาที่ต้องตัดสินใจลงทุน มักจะไม่ทราบข้อมูลอันดับที่ถูกต้อง



รูปที่ 4.6 อรรถประโยชน์และอรรถประโยชน์สะสมที่ได้รับจากการลงทุนเมื่อใช้ข้อมูลอันดับที่สมบูรณ์เทียบกับการใช้ข้อมูลจากการแจกแจงก่อน

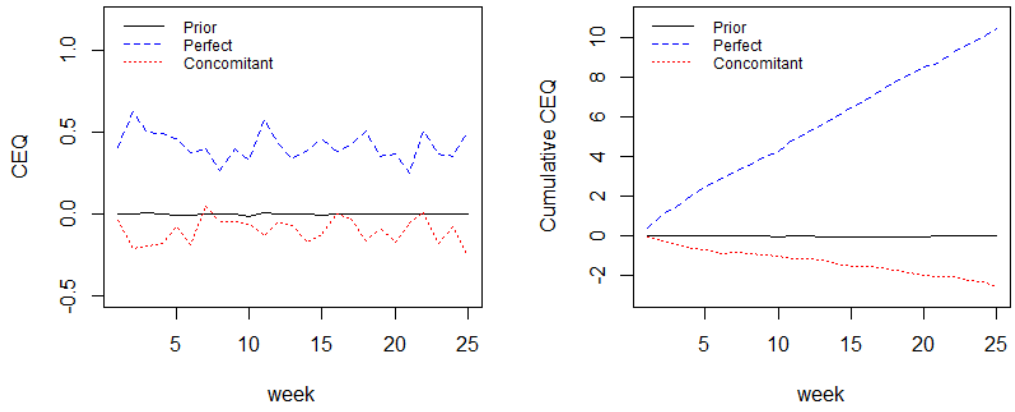
4.3.2 ผลเมื่อใช้ข้อมูลอันดับจากตัวแปรร่วม

เนื่องจากข้อมูลอันดับที่สมบูรณ์หาได้ยากในทางปฏิบัติ ในหัวข้อนี้จึงศึกษาผลการจัดพอร์ตการลงทุนเมื่อใช้ข้อมูลคำแนะนำการซื้อขายหลักทรัพย์จากบทวิเคราะห์ที่เผยแพร่ก่อนเวลาที่ต้องตัดสินใจลงทุน สอดคล้องกับสถานการณ์ที่เป็นไปได้จริง โดยนำข้อมูลที่ได้มาให้คะแนนและถ่วงน้ำหนักแบบต่างๆ เพื่อให้ได้ข้อมูลอันดับที่ใกล้เคียงกับข้อมูลอันดับที่ถูกต้อง จากนั้นจะนำข้อมูลอันดับที่ได้มาจัดพอร์ตการลงทุน แล้วเปรียบเทียบประสิทธิภาพที่จะได้รับจากการใช้ข้อมูลอันดับที่ไม่สมบูรณ์ (Implied Rank) กับกรณีใช้ข้อมูลอันดับที่สมบูรณ์ (Perfect Rank) และกรณีที่ใช้ข้อมูลการแจกแจงก่อน (Prior) เพียงอย่างเดียว

วิธีที่ 1: รวมคะแนนจากรายการคำแนะนำแต่ละรายการ

วิธีนี้จะให้น้ำหนักกับรายการคำแนะนำแต่ละรายการเท่าๆกัน แต่จะให้คะแนนเป็นบวก (1 คะแนน) สำหรับคำแนะนำให้ซื้อและข่าวระบุว่ามีผลดีต่อหุ้นตัวนั้นๆ ส่วนคำแนะนำให้ขายหุ้นหรือข่าวระบุว่ามีผลเสียต่อหุ้น จะให้คะแนนเป็นลบ (-1 คะแนน) จากนั้นรวมคะแนนของแต่ละวันแยกตามกลุ่มอุตสาหกรรม 8 กลุ่ม แล้วเรียงลำดับคะแนนรวมที่ได้ ก็จะได้เป็นข้อมูลอันดับของกลุ่มอุตสาหกรรมแต่ละวัน จากนั้นนำข้อมูลนี้ไปคำนวณหาค่าคาดหวัง (Implied Rank) ก่อนที่จะไปหาน้ำหนักการลงทุนที่เหมาะสมและเปรียบเทียบประสิทธิภาพ

ผลการจัดพอร์ตการลงทุนแสดงได้ดังรูปที่ 4.7 ซึ่งจะเห็นได้ว่าเมื่อใช้ข้อมูลอันดับของตัวแปรร่วม อรรถประโยชน์ที่ได้รับจะมีค่าต่ำกว่าอรรถประโยชน์ที่ได้จากการใช้ข้อมูลของการแจกแจงก่อนแทบจะทุกสัปดาห์ และเมื่อพิจารณาค่าอรรถประโยชน์สะสมก็พบว่าค่าที่ได้ต่ำกว่าที่ได้จากกรณีที่ใช้ข้อมูลของการแจกแจงก่อนมาก สอดคล้องกับค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนกับผลตอบแทนของเวลา t+1 ในช่วงเวลา 6 เดือน ที่มีค่าน้อยมาก (น้อยกว่า 0.2 ทุกกลุ่มอุตสาหกรรม) ตามตารางที่ 4.1



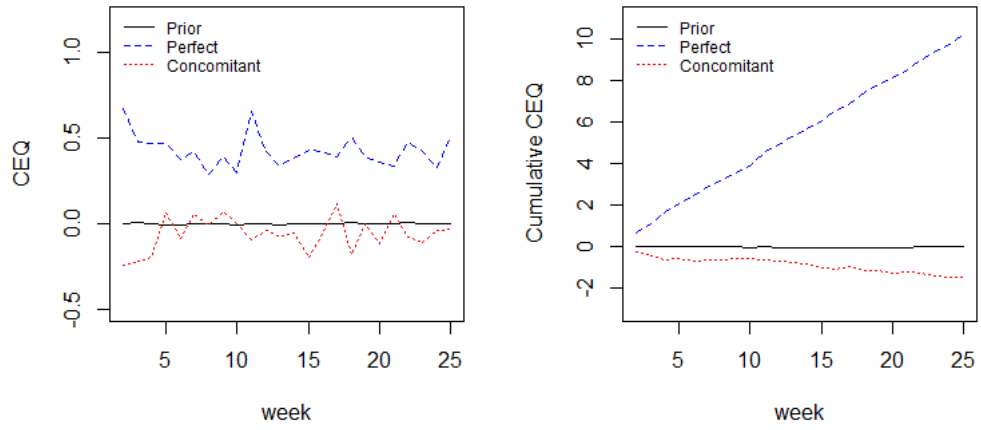
รูปที่ 4.7 อรรถประโยชน์และอรรถประโยชน์สะสมที่ได้รับจากการลงทุนเมื่อใช้ข้อมูลอันดับที่รวบรวมโดยให้น้ำหนักกับคำแนะนำแต่ละรายการเท่าๆกัน โดยประเมินประสิทธิภาพจากผลตอบแทนสิ้นวันทำการถัดไป

นอกจากนี้ได้ทดลองจัดพอร์ตการลงทุน โดยใช้บทวิเคราะห์ที่เผยแพร่ในช่วงเช้าของแต่ละวัน มาเป็นข้อมูลในการจัดพอร์ตการลงทุนและประเมินประสิทธิภาพเทียบกับผลตอบแทน ณ สิ้นวันนั้นๆ (ใช้ราคาปิดของวันนั้นในการคำนวณผลตอบแทน) ส่วนบทวิเคราะห์ช่วงบ่าย จะถือเป็นข้อมูลสำหรับลงทุนในวันทำการถัดไป ซึ่งในความเป็นจริงก็เป็นไปได้ที่อาจมีกรณีที่ข่าวหรือคำแนะนำที่มีผลกับราคาหุ้นในวันนั้นๆ ทันที มากกว่าที่จะไปมีผลในวันทำการถัดไป

สอดคล้องกับค่าสัมประสิทธิ์สหสัมพันธ์ตามตารางที่ 4.1 ที่แสดงให้เห็นว่าข้อมูลคะแนนที่รวบรวมได้มีความสัมพันธ์กับข้อมูลผลตอบที่สิ้นวันของวันนั้นๆ โดยเฉพาะอย่างยิ่งกลุ่มธุรกิจการเงินและกลุ่มสินค้าอุตสาหกรรม ที่ในช่วงเวลาที่สนใจศึกษาได้ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนกับผลตอบแทนสิ้นวันทำการนั้นๆ สูงกว่าระดับความสัมพันธ์กับผลตอบแทนสิ้นวันทำการถัดไปมาก

ด้วยเหตุนี้จึงทดลองคำนวณประสิทธิภาพการจัดพอร์ตการลงทุนเมื่อใช้ข้อมูลอันดับที่ถูกต้องจากข้อมูลผลตอบแทนสิ้นวัน เทียบกับการใช้ข้อมูลอันดับของตัวแปรร่วม ได้ผลตามรูปที่ 4.8 ซึ่งจะเห็นว่าประสิทธิภาพที่ได้จากการจัดพอร์ตการลงทุนเมื่อวัดจากผลตอบแทนสิ้นวันแต่ละวัน ได้ผลดีกว่าการใช้ผลตอบแทนของวันทำการถัดไปเล็กน้อย ในรูปด้านซ้ายจะเห็นว่ามีบางสัปดาห์ที่อรรถประโยชน์ที่ได้รับเมื่อใช้ข้อมูลอันดับของตัวแปรร่วมสูงกว่าการใช้ข้อมูลการแจกแจงก่อนเพียงอย่างเดียว และใน

กราฟด้านขวาอรรถประโยชน์สะสมที่ได้ถึงแม้ว่าจะต่ำกว่าการแจกแจงก่อนอยู่ แต่ถ้าเทียบกับรูปที่ 4.7 ที่ประเมินจากผลตอบแทนของวันทำการถัดไป พบว่าค่าอรรถประโยชน์สะสมจะติดลบน้อยกว่า



รูปที่ 4.8 อรรถประโยชน์และอรรถประโยชน์สะสมที่ได้รับจากการลงทุนเมื่อใช้ข้อมูลอันดับที่รวบรวมโดยให้น้ำหนักกับคำแนะนำแต่ละรายการเท่าๆ กัน โดยประเมินประสิทธิภาพจากผลตอบแทนสิ้นวันทำการนั้นๆ

ตารางที่ 4.1 ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนที่รวบรวมโดยให้น้ำหนักกับคำแนะนำแต่ละรายการเท่าๆ กัน กับผลตอบแทนของวันทำการถัดไป และผลตอบแทนสิ้นวันนั้นๆ

กลุ่มอุตสาหกรรม	ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนกับ	
	ผลตอบแทนของวันทำการถัดไป	ผลตอบแทนสิ้นวันนั้นๆ
เกษตรและอุตสาหกรรมอาหาร (AGRO)	-0.1060	0.0186
สินค้าอุปโภคบริโภค (CONSUMP)	0.0275	-0.0318
ธุรกิจการเงิน (FINCIAL)	-0.1101	0.3894
สินค้าอุตสาหกรรม (INDUS)	0.0154	0.2263
อสังหาริมทรัพย์และก่อสร้าง (PROPCON)	-0.1762	-0.0098
ทรัพยากร (RESOURC)	-0.0516	0.0536
บริการ (SERVICE)	-0.0423	0.1228
เทคโนโลยี (TECH)	0.0490	0.1539

วิธีที่ 2: ถ่วงน้ำหนักคำแนะนำด้วยสัดส่วนมูลค่าของหุ้นในกลุ่มอุตสาหกรรม

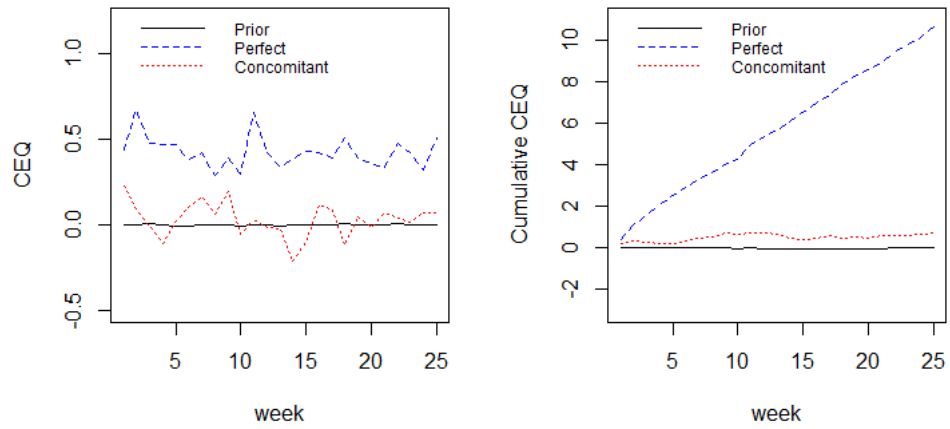
เนื่องจากเราพิจารณาผลตอบแทนและหาน้ำหนักการลงทุนที่เหมาะสมสำหรับกลุ่มอุตสาหกรรม แต่คำแนะนำที่เรารวบรวมได้เป็นของหุ้นแต่ละตัว ดังนั้น หากหุ้นที่ได้รับคำแนะนำให้ซื้อมีมูลค่าไม่มากนักเมื่อเทียบกับมูลค่าของหุ้นทั้งหมดในกลุ่มอุตสาหกรรม การที่ราคาของหุ้นดังกล่าวปรับสูงขึ้น ก็ไม่ได้ทำให้ผลตอบแทนของทั้งกลุ่มอุตสาหกรรมเพิ่มขึ้นมากนัก ดังนั้นในวิธีที่สองนี้ จึงถ่วงน้ำหนักของ

คำแนะนำด้วยสัดส่วนของมูลค่า (market capitalization) หุ้นตัวนั้นๆ เทียบกับมูลค่าหุ้นทั้งกลุ่มอุตสาหกรรม ซึ่งสัดส่วนนี้จะใช้มูลค่าของหุ้นและกลุ่มอุตสาหกรรมของวันก่อนหน้า (ข้อมูล ณ เวลา t-1) มาคำนวณ ดูรายละเอียดข้อมูลที่นำมาคำนวณได้จากรูปที่ 3.6

ค่าสัมประสิทธิ์สหสัมพันธ์ในตารางที่ 4.1 และ 4.2 แสดงให้เห็นว่าคำแนะนำจากบทวิเคราะห์ที่รวบรวมได้ มีผลต่อผลตอบแทนในวันนั้นๆ มากกว่าผลตอบแทนของวันทำการถัดไป ต่อจากนี้ จึงจะขอนำเสนอเฉพาะประสิทธิภาพจากการจัดพอร์ตการลงทุนเมื่อประเมินด้วยผลตอบแทน ณ สิ้นวันทำการแต่ละวัน

โดยเมื่อพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนที่ได้จากวิธีที่ 2 นี้กับผลตอบแทนวันนั้นๆ ในตารางที่ 4.2 เทียบกับค่าที่ได้จากวิธีที่ 1 ตามตารางที่ 4.1 พบว่าคะแนนที่รวบรวมเมื่อถ่วงน้ำหนักด้วยมูลค่าของหุ้นในกลุ่มอุตสาหกรรมมีความสัมพันธ์กับผลตอบแทนมากกว่า โดยเฉพาะกลุ่มเทคโนโลยีและบริการที่ค่าดังกล่าวเพิ่มขึ้นจากตารางที่ 4.1 อย่างเห็นได้ชัด

ผลตามรูปที่ 4.9 เห็นได้ว่าประสิทธิภาพของวิธีที่สองดีขึ้นกว่าผลจากวิธีแรกมาก ค่าอรรถประโยชน์ที่ได้ส่วนมากสูงกว่ากรณีที่ไม่นำข้อมูลอันดับมาประกอบการพิจารณา โดยเฉพาะกราฟด้านขวา ซึ่งแสดงให้เห็นว่าอรรถประโยชน์สะสมสูงกว่าการใช้ข้อมูลของการแจกแจงก่อนเพียงอย่างเดียว



รูปที่ 4.9 อรรถประโยชน์และอรรถประโยชน์สะสมที่ได้รับจากการลงทุนเมื่อใช้ข้อมูลอันดับที่รวบรวมโดยถ่วงน้ำหนักคำแนะนำด้วยสัดส่วนมูลค่าของหุ้นในกลุ่มอุตสาหกรรม

ตารางที่ 4.2 ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนที่รวบรวมโดยถ่วงน้ำหนักคำแนะนำด้วยสัดส่วนมูลค่าของหุ้นในกลุ่มอุตสาหกรรมกับกับผลตอบแทนของวันทำการถัดไป และผลตอบแทนสิ้นวันนั้นๆ

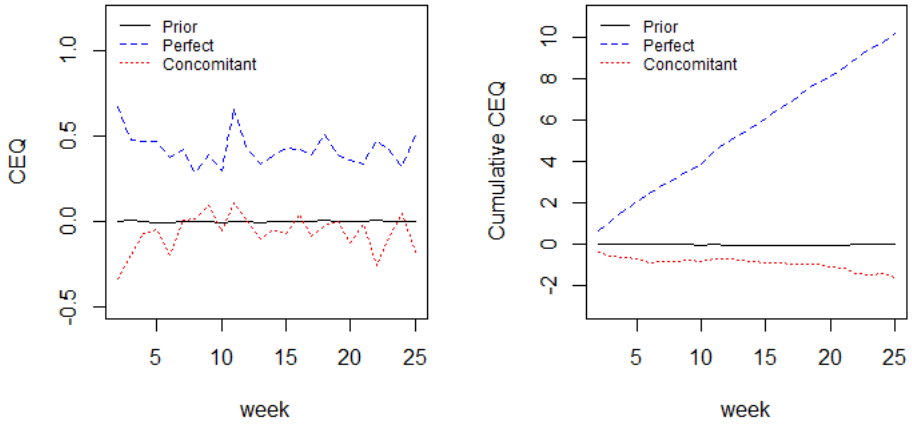
กลุ่มอุตสาหกรรม	ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนกับ	
	ผลตอบแทนของวันทำการถัดไป	ผลตอบแทนสิ้นวันนั้นๆ
เกษตรและอุตสาหกรรมอาหาร (AGRO)	-0.0104	0.1013
สินค้าอุปโภคบริโภค (CONSUMP)	0.1148	0.1828
ธุรกิจการเงิน (FINCIAL)	-0.1305	0.5045
สินค้าอุตสาหกรรม (INDUS)	-0.0363	0.2490
อสังหาริมทรัพย์และก่อสร้าง (PROPCON)	-0.0595	0.2375
ทรัพยากร (RESOURC)	-0.1248	0.1729
บริการ (SERVICE)	0.0464	0.2600
เทคโนโลยี (TECH)	0.1506	0.4541

วิธีที่ 3: ให้คะแนนตามตามระดับความรุนแรงของคำแนะนำ

ในบางบทวิเคราะห์จะมีการแยกประเภทคำแนะนำ เช่น แนะนำให้ซื้ออย่างมาก (Strong Buy) แนะนำให้ถือหรือสะสมหุ้นตัวนั้นๆ ไว้ก่อน หรือแนะนำให้ทยอยขายออก เป็นต้น วิธีที่สามจึงโดยต่อยอดจากวิธีที่ 2 แต่ให้น้ำหนักเพิ่มหรือลดตามสัดส่วนความรุนแรงของคำแนะนำแต่ละแบบ ดังนี้

- คำแนะนำที่มีความรุนแรง เช่น strong buy หรือ strong sell
- คำแนะนำแบบที่ไม่รุนแรงมาก เช่น สะสม หรือทยอยซื้อ
- คำแนะนำแบบกลางๆ เช่น Neutral หรือ Hold

ผลที่ได้ตามรูปที่ 4.10 เห็นได้ว่าอรรถประโยชน์ที่ได้ลดลงกว่าวิธีที่ 2 ส่วนหนึ่งอาจเนื่องมาจากมีเพียงบางบทวิเคราะห์ที่อธิบายระดับความรุนแรงของคำแนะนำ รวมทั้งการตีความหมายของระดับความรุนแรงที่ได้จากข้อมูลแต่ละแหล่งอาจแตกต่างกันไป (subjective) ทำให้การใช้ข้อมูลนี้มาเป็นเกณฑ์ไม่ได้ทำให้ข้อมูลอันดับที่ได้จากตัวแปรร่วมถูกต้องมากขึ้นนัก ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนที่ได้จากวิธีนี้กับผลตอบแทนสิ้นวันนั้นๆ ในตารางที่ 4.3 ก็น้อยกว่าค่าที่ได้จากวิธีที่ 2 มาก



รูปที่ 4.10 อรรถประโยชน์และอรรถประโยชน์สะสมที่ได้รับจากการลงทุนเมื่อใช้ข้อมูลอันดับที่รวบรวมโดยถ่วงน้ำหนักด้วยระดับความรุนแรงของคำแนะนำ

ตารางที่ 4.3 ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนที่รวบรวมโดยถ่วงน้ำหนักด้วยระดับความรุนแรงของคำแนะนำกับผลตอบแทนสิ้นวัน

กลุ่มอุตสาหกรรม	ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนกับผลตอบแทนสิ้นวันนั้นๆ
เกษตรและอุตสาหกรรมอาหาร (AGRO)	0.1157
สินค้าอุปโภคบริโภค (CONSUMP)	0.0351
ธุรกิจการเงิน (FINCIAL)	0.0710
สินค้าอุตสาหกรรม (INDUS)	0.0988
อสังหาริมทรัพย์และก่อสร้าง (PROPCON)	0.2291
ทรัพยากร (RESOURC)	0.0312
บริการ (SERVICE)	0.1869
เทคโนโลยี (TECH)	0.2610

วิธีที่ 4: ให้คะแนนตามตามความแม่นยำของการทำนายในอดีต

เนื่องจากผลของงานวิจัยก่อนหน้าที่พบว่าการนำข้อมูลความแม่นยำของการทำนายในอดีตมาพิจารณาสามารถเพิ่มประสิทธิภาพของผลลัพธ์ที่ได้ให้สูงขึ้น ดังนั้น ในวิธีนี้จึงใช้ข้อมูลในอดีตของแต่ละบริษัทหลักทรัพย์มาประเมินและถ่วงน้ำหนักร่วมด้วย โดยเลือกพัฒนาวิธีการจัดอันดับของกลุ่มหลักทรัพย์ต่อจากวิธีที่ 2

จากแนวคิดที่ว่าบริษัทหลักทรัพย์แต่ละแห่งจะมีความเชี่ยวชาญในการกลุ่มอุตสาหกรรมต่างๆกัน งานวิจัยนี้จึงคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ (Pearson correlation coefficient) ระหว่างคะแนนที่ได้ในวิธีที่ 2 กับผลตอบแทนของวันนั้นๆ โดยพิจารณาในช่วงเวลาหนึ่งๆ ในที่นี้ทดลองประเมินในช่วง 10, 20 และ 30 วันทำการ ตามลำดับ โดยจะคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ของแต่ละ

แหล่งข้อมูล แล้วจึงนำน้ำหนักของคำแนะนำ ณ วันที่ต้องการจัดพอร์ตการลงทุนมาที่คำนวณตามวิธีที่ 2 มาคูณด้วยสัดส่วนความแม่นยำของแหล่งข้อมูลในอดีต (ดูรายละเอียดตามรูปที่ 3.6)

สำหรับเกณฑ์การถ่วงน้ำหนักตามความแม่นยำจะใช้ค่าที่ได้จากค่าสัมบูรณ์ของค่าสัมประสิทธิ์สหสัมพันธ์ เนื่องจาก ระดับความสัมพันธ์ดังกล่าวไม่ว่าจะเป็นความสัมพันธ์ด้านบวกหรือด้านลบ ถ้าสัมพันธ์กับผลตอบแทนในระดับสูง ก็ถือว่าเป็นแหล่งข้อมูลที่มีประโยชน์ ในกรณีที่ค่าสัมประสิทธิ์สหสัมพันธ์เป็นลบ เราจะคูณคำแนะนำจากแหล่งนั้นด้วยค่าลบ เพื่อปรับค่าให้สอดคล้องกับผลตอบแทนที่สนใจ

และจะปรับข้อมูลจากแหล่งที่สัมพันธ์กับผลตอบแทนในระดับต่ำ โดยการคูณค่าที่ใกล้ศูนย์เปรียบเสมือนการขจัดข้อมูลรบกวน (noise) แต่ที่ไม่คูณด้วยศูนย์ เนื่องจากข้อมูลที่ใช้มีปริมาณไม่มาก เมื่อลองตัดข้อมูลรบกวนให้เป็นศูนย์ ทำให้มีหลายอุตสาหกรรมที่มีคะแนนเป็นศูนย์ แล้วทำให้การจัดอันดับผิดเพี้ยนไปมาก สำหรับค่าที่นำมาถ่วงน้ำหนักกำหนดตามหลักเกณฑ์ต่อไปนี้

หากในช่วงเวลาที่ประเมินเกิดมีแหล่งข้อมูลที่ไม่เคยให้คำแนะนำให้ซื้อขายหุ้นในกลุ่มอุตสาหกรรมที่กำลังพิจารณาอยู่เลย จะกำหนดให้น้ำหนักที่จะนำไปคูณกับสัดส่วนมูลค่าของหุ้นในกลุ่มอุตสาหกรรม

นอกจากนี้พบว่าค่าสัมประสิทธิ์สหสัมพันธ์ที่ใช้ประเมินความแม่นยำจะมักมีค่าสูงเกินกว่าที่ควรจะเป็นเมื่อจำนวนข้อมูลที่ใช้คำนวณมีน้อย ซึ่งจะเกิดกับบริษัทหลักทรัพย์เคยมีคำแนะนำสำหรับหุ้นในกลุ่มอุตสาหกรรมนั้นๆ ในช่วงเวลาที่พิจารณาน้อย ในที่นี้จึงปรับลดน้ำหนักของคำแนะนำกรณีมีคำแนะนำในอุตสาหกรรมนั้นๆ น้อยกว่า 7 ครั้ง ในช่วงเวลาที่พิจารณา

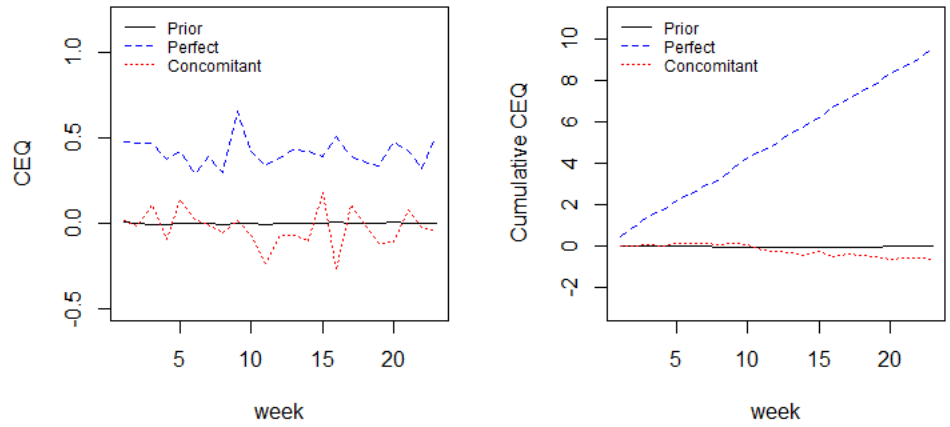
อย่างไรก็ตาม เกณฑ์ข้างต้นนี้ได้จากการทดลองเลือกค่าหลายๆแบบ (trial and error) จนสุดท้ายพบว่าการกำหนดค่าตามเกณฑ์ลักษณะนี้ได้ผลดีกว่าแบบอื่นๆ ที่ได้ทดลองมา ในงานวิจัยต่อๆไปอาจศึกษาหาวิธีการรวบรวมและถ่วงน้ำหนักแบบอื่นๆ ที่ทำให้ได้ข้อมูลอันดับที่ใกล้เคียงกับข้อมูลอันดับที่สมบูรณ์มากขึ้น (จะกล่าวถึงรายละเอียดเรื่องนี้อีกครั้งในบทที่ 5)

- *ผลการจัดพอร์ตการลงทุนเมื่อประเมินความแม่นยำในอดีตในช่วง 10 วันทำการย้อนหลัง*

ผลการจัดพอร์ตการลงทุนเมื่อประเมินความแม่นยำของแหล่งข้อมูลโดยใช้ข้อมูลย้อนหลังไป 10 วันทำการ ปรากฏว่ามีหลายช่วงที่อรรถประโยชน์ที่ได้รับน้อยกว่ากรณีที่ใช้ข้อมูลของการแจกแจงก่อน และอรรถประโยชน์สะสมก็ต่ำกว่ากรณีใช้ข้อมูลของการแจกแจงก่อน เช่นเดียวกับค่าสัมประสิทธิ์สหสัมพันธ์ในตารางที่ 4.4 ที่แสดงให้เห็นว่าข้อมูลคะแนนที่ได้จากวิธีนี้สัมพันธ์กับข้อมูลผลตอบแทนจริงน้อยกว่าการรวมคะแนนด้วยวิธีที่ 2

สาเหตุหลักที่ทำให้เกิดเหตุการณ์นี้ คือ การที่ขาดข้อมูลคำแนะนำในอดีตที่จะนำมาใช้ในการประเมิน เนื่องจาก เราคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แยกตามกลุ่มอุตสาหกรรมและแยกตาม

แหล่งข้อมูลแต่ละแหล่ง ทำให้เกิดกรณีที่กลุ่มอุตสาหกรรมบางกลุ่มไม่ค่อยได้รับเลือกมาวิเคราะห์ หรืออาจเกิดจากแหล่งข้อมูลเองที่เลือกให้คำแนะนำเฉพาะกลุ่มอุตสาหกรรมที่ตนเองมีความเชี่ยวชาญ ส่งผลให้การถ่วงน้ำหนักโดยใช้การประเมินความถูกต้องในช่วงเวลา 10 วันทำการย้อนหลังมีประสิทธิภาพที่ได้ต่ำกว่าการใช้ข้อมูลของการแจกแจงก่อน จึงทดลองขยายระยะเวลาการประเมินเป็น 20 และ 30 วันตามลำดับ



รูปที่ 4.11 อรรถประโยชน์และอรรถประโยชน์สะสมที่ได้รับจากการลงทุนเมื่อใช้ข้อมูลอันดับที่รวบรวมโดยประเมินความแม่นยำของแหล่งข้อมูลย้อนหลัง 10 วันทำการ

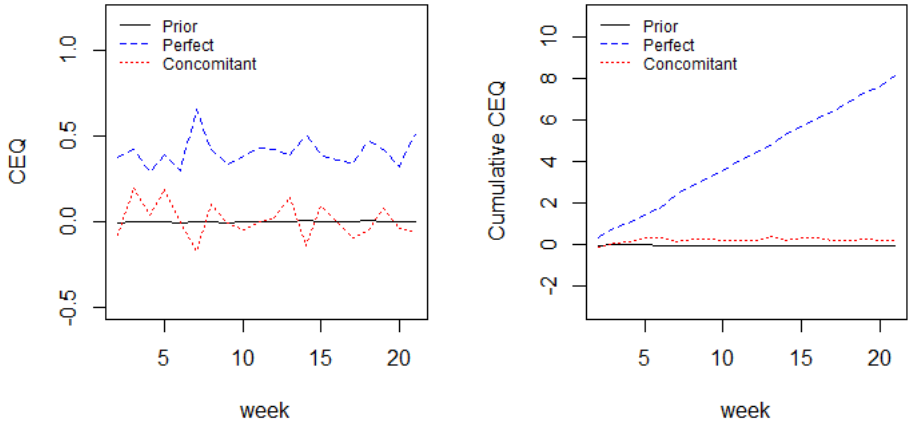
ตารางที่ 4.4 ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนที่รวบรวมโดยประเมินความแม่นยำของแหล่งข้อมูลย้อนหลัง 10 วันกับผลตอบแทนสิ้นวัน

กลุ่มอุตสาหกรรม	ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนกับผลตอบแทนสิ้นวันนั้นๆ
เกษตรและอุตสาหกรรมอาหาร (AGRO)	0.0543
สินค้าอุปโภคบริโภค (CONSUMP)	0.0451
ธุรกิจการเงิน (FINCIAL)	0.5607
สินค้าอุตสาหกรรม (INDUS)	0.2413
อสังหาริมทรัพย์และก่อสร้าง (PROPCON)	0.2461
ทรัพยากร (RESOURC)	0.0968
บริการ (SERVICE)	0.1910
เทคโนโลยี (TECH)	0.3774

- *ผลการจัดพอร์ตการลงทุนเมื่อประเมินความแม่นยำในอดีตในช่วง 20 วันทำการย้อนหลัง*

เมื่อทดลองประเมินความแม่นยำของคำแนะนำในอดีตย้อนหลังไป 20 วันทำการ พบว่า ค่าอรรถประโยชน์ที่ได้รับดีขึ้นกว่าตอนใช้ข้อมูลย้อนหลัง 10 วันทำการ ดังแสดงในรูปที่ 4.12 โดย

อรรถประโยชน์สะสมสูงกว่าการใช้ข้อมูลการแจกแจงก่อน และเมื่อพิจารณาระดับความสัมพันธ์ของคะแนนที่คำนวณได้กับผลตอบแทนจริงในตารางที่ 4.5 สูงกว่าเมื่อประเมินย้อนหลัง 10 วัน แต่ก็ยังไม่สูงมากนัก จึงทดลองต่อโดยขยายระยะเวลาประเมินออกเป็น 30 วันทำการ



รูปที่ 4.12 อรรถประโยชน์และอรรถประโยชน์สะสมที่ได้รับจากการลงทุนเมื่อใช้ข้อมูลอันดับที่รวบรวมโดยประเมินความแม่นยำของแหล่งข้อมูลย้อนหลัง 20 วันทำการ

ตารางที่ 4.5 ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนที่รวบรวมโดยประเมินความแม่นยำของแหล่งข้อมูลย้อนหลัง 20 วันกับผลตอบแทนสิ้นวัน

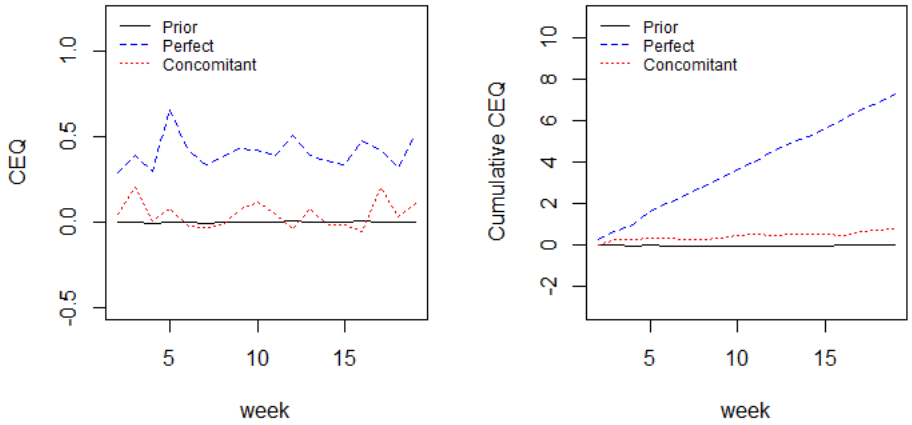
กลุ่มอุตสาหกรรม	ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนกับผลตอบแทนสิ้นวันนั้นๆ
เกษตรและอุตสาหกรรมอาหาร (AGRO)	0.1731
สินค้าอุปโภคบริโภค (CONSUMP)	0.0849
ธุรกิจการเงิน (FINCIAL)	0.6159
สินค้าอุตสาหกรรม (INDUS)	0.1940
อสังหาริมทรัพย์และก่อสร้าง (PROPCON)	0.3023
ทรัพยากร (RESOURC)	0.2630
บริการ (SERVICE)	0.3513
เทคโนโลยี (TECH)	0.4275

- *ผลการจัดพอร์ตการลงทุนเมื่อประเมินความแม่นยำในอดีตในช่วง 30 วันย้อนหลัง*

เมื่อประเมินความแม่นยำของแหล่งข้อมูลโดยใช้ข้อมูลย้อนหลังไป 30 วันทำการ ผลปรากฏว่า ค่าอรรถประโยชน์ที่ได้รับสูงกว่าการใช้ข้อมูลจากการแจกแจงก่อนเกือบทุกสัปดาห์ โดยอรรถประโยชน์สะสมตามรูปที่ 4.13 ก็สูงกว่าการแจกแจงก่อน และยังสูงกว่าวิธีการถ่วงน้ำหนักด้วยวิธีอื่นๆที่ได้ทดลองผ่านมาทั้งหมดด้วย (เปรียบเทียบในช่วง 19 สัปดาห์เท่ากัน) นอกจากนี้ เมื่อพิจารณาค่า

สัมประสิทธิ์สหสัมพันธ์ที่แสดงไว้ในตารางที่ 4.6 ก็พบว่าระดับความสัมพันธ์ของคะแนนที่ได้จากวิธีนี้กับผลตอบแทนจริงมีค่าสูงกว่ากรณีที่ประเมินย้อนหลังไป 10 และ 20 วันทำการ ซึ่งสาเหตุหลักน่าจะมาจากการที่มีข้อมูลในการประเมินความแม่นยำเพิ่มขึ้นนั่นเอง

ข้อสังเกตอีกประการหนึ่งของการถ่วงน้ำหนักด้วยความแม่นยำในอดีต คือ ถ้าแหล่งข้อมูลที่มีความสัมพันธ์กับผลตอบแทนมากเกิดให้คำแนะนำไม่ถูกต้อง จะทำให้ผลตอบแทนที่ได้รับต่ำกว่าที่ควร สังเกตได้จากผลในสัปดาห์ที่ 16 ที่ค่าอรรถประโยชน์ลดลงต่ำกว่ากรณีใช้ข้อมูลผลตอบแทนก่อน



รูปที่ 4.13 อรรถประโยชน์และอรรถประโยชน์สะสมที่ได้รับจากการลงทุนเมื่อใช้ข้อมูลอันดับที่รวบรวมโดยประเมินความแม่นยำของแหล่งข้อมูลย้อนหลัง 30 วันทำการ

ตารางที่ 4.6 ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนที่รวบรวมโดยประเมินความแม่นยำของแหล่งข้อมูลย้อนหลัง 30 วันกับผลตอบแทนสิ้นวัน

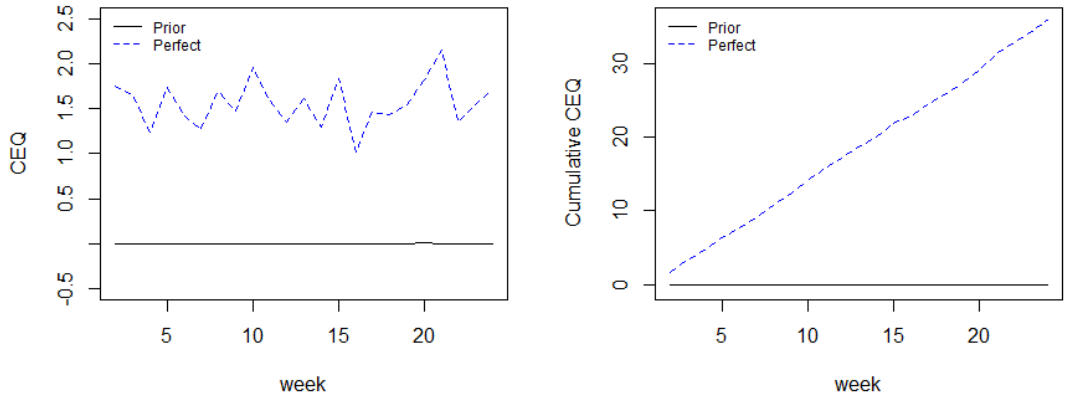
กลุ่มอุตสาหกรรม	ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนกับผลตอบแทนสิ้นวันนั้นๆ
เกษตรและอุตสาหกรรมอาหาร (AGRO)	0.2511
สินค้าอุปโภคบริโภค (CONSUMP)	0.2843
ธุรกิจการเงิน (FINCIAL)	0.6178
สินค้าอุตสาหกรรม (INDUS)	0.4252
อสังหาริมทรัพย์และก่อสร้าง (PROPCON)	0.4283
ทรัพยากร (RESOURC)	0.3271
บริการ (SERVICE)	0.3454
เทคโนโลยี (TECH)	0.4863

4.3.3 ผลการจัดพอร์ตการลงทุนสำหรับกลุ่มอุตสาหกรรมย่อย

ผลที่นำเสนอก่อนหน้านี้ เป็นการหาน้ำหนักการลงทุนที่เหมาะสมสำหรับกลุ่มอุตสาหกรรมหลัก 8 กลุ่มอุตสาหกกรม ในหัวข้อสุดท้ายนี้ จะทดลองจัดพอร์ตการลงทุนเมื่อจำนวนหลักทรัพย์ที่สนใจมีจำนวนมากขึ้น โดยใช้ข้อมูลชุดเดิมแต่แบ่งกลุ่มของหุ้นออกเป็นกลุ่มอุตสาหกรรมย่อย 28 กลุ่มแทน เพื่อศึกษาว่าเมื่อจำนวนตัวแปรที่สนใจศึกษาเพิ่มมากขึ้นแล้ว ประสิทธิภาพของการจัดพอร์ตการลงทุนจะเปลี่ยนแปลงไปอย่างไรบ้าง สำหรับพารามิเตอร์ต่างๆ เช่น ค่าสัมประสิทธิ์การหลีกเลี่ยงความเสี่ยง หรือค่า τ ที่ใช้ปรับค่าในเมทริกซ์ความแปรปรวนร่วม จะใช้ค่าเท่ากับที่ใช้ในการจัดพอร์ตการลงทุนของกลุ่มอุตสาหกรรมหลัก

- **กรณีใช้ข้อมูลอันดับที่ถูกต้อง**

ผลทดลองจัดพอร์ตการลงทุนของกลุ่มอุตสาหกรรมย่อยตามรูปที่ 4.14 พบว่า อรรถประโยชน์ที่ได้รับเมื่อนำข้อมูลอันดับที่สมบูรณ์มาประกอบการคำนวณสูงกว่าเมื่อใช้เพียงข้อมูลจากการแจกแจงก่อนเพียงอย่างเดียว สอดคล้องกับกรณีที่พิจารณากลุ่มอุตสาหกรรมหลัก และสอดคล้องกับงานวิจัยอื่นๆ ด้วย



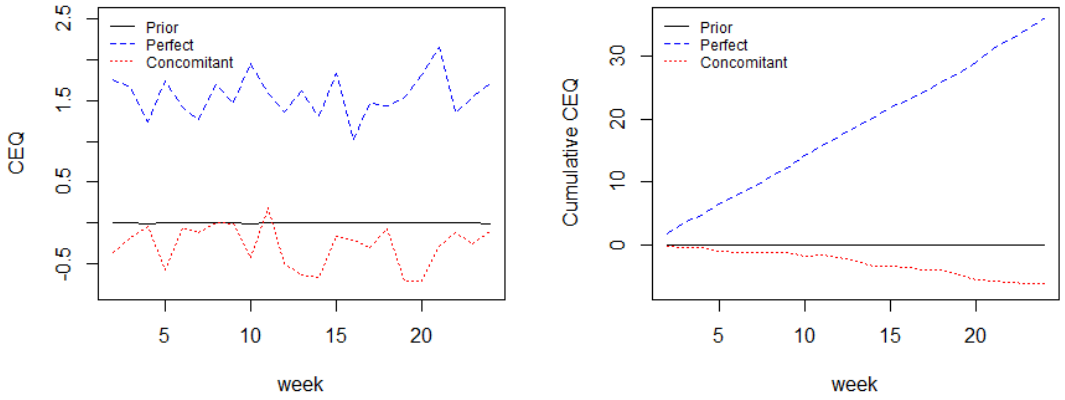
รูปที่ 4.14 อรรถประโยชน์และอรรถประโยชน์สะสมที่ได้รับจากการลงทุนในกลุ่มอุตสาหกรรมย่อย เมื่อใช้ข้อมูลอันดับที่สมบูรณ์เทียบกับการใช้ข้อมูลจากการแจกแจงก่อน

- **กรณีใช้ข้อมูลอันดับของตัวแปรร่วม**

เริ่มพิจารณาจากวิธีการรวมข้อมูลและจัดอันดับโดยสัดส่วนมูลค่าของหุ้นในกลุ่มอุตสาหกรรมย่อยก่อน เนื่องจากวิธีนี้เคยได้ผลดี กล่าวคือ เมื่อใช้กับกลุ่มอุตสาหกรรมหลักแล้ว อรรถประโยชน์สะสมที่ได้สูงกว่ากรณีที่ใช้ข้อมูลการแจกแจงก่อน

อย่างไรก็ตาม ผลการจัดพอร์ตการลงทุนในรูปที่ 4.15 กลับพบว่าอรรถประโยชน์ที่ได้รับต่ำกว่าที่ได้จากการใช้ข้อมูลการแจกแจงก่อนแทบจะทุกสัปดาห์ และค่าอรรถประโยชน์สะสมก็ต่ำกว่าด้วย ซึ่งเมื่อพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ในตารางที่ 4.7 ประกอบด้วยแล้ว ยิ่งเห็นชัดว่าคะแนนที่รวบรวมได้ด้วยวิธีนี้ ส่วนใหญ่แล้วจะสัมพันธ์กับผลตอบแทนจริงน้อยมาก

สาเหตุหลักที่ทำให้ผลการจัดพอร์ตการลงทุนออกมาไม่ดี เนื่องจาก เมื่อเราพิจารณากลุ่มอุตสาหกรรม เพิ่มจาก 8 กลุ่มหลัก มาเป็น 28 กลุ่มย่อย ทำให้โอกาสที่กลุ่มอุตสาหกรรมย่อยจะไม่มีหุ้นที่ได้รับคำแนะนำก็จะมากตามไปด้วย และเมื่อมาดูข้อมูลที่ใช้พบว่าจากคำแนะนำ 20,924 รายการที่รวบรวมได้ พอนำมาจำแนกตามกลุ่มอุตสาหกรรมย่อยตามตารางที่ 4.8 แล้ว พบว่ามีบางกลุ่มที่แทบไม่มีคำแนะนำเลย เช่น กลุ่มเหมืองแร่ ที่มีเพียง 10 คำแนะนำ หรือกลุ่มของใช้ส่วนตัวและเวชภัณฑ์ และกลุ่มแฟชั่น ที่มีประมาณ 40 คำแนะนำ ในช่วง 6 เดือนที่พิจารณา ด้วยเหตุนี้เอง พอนำคะแนนที่รวบรวมได้ไปจัดอันดับจึงมีหลายกลุ่มที่ได้คะแนนเป็นศูนย์ ทำให้เมื่อแปลงเป็นข้อมูลอันดับเพื่อนำไปคำนวณค่าคาดหวังจึงได้ค่าที่ผิดไปจากที่ควรจะเป็นมาก



รูปที่ 4.15 อรรถประโยชน์และอรรถประโยชน์สะสมที่ได้รับจากการลงทุนเมื่อใช้ข้อมูลอันดับที่รวบรวมโดยถ่วงน้ำหนักคำแนะนำด้วยสัดส่วนมูลค่าของหุ้นในกลุ่มอุตสาหกรรมย่อย

ตารางที่ 4.7 ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนที่รวบรวมโดยถ่วงน้ำหนักคำแนะนำด้วยสัดส่วนมูลค่าของหุ้นในกลุ่มอุตสาหกรรมย่อยกับผลตอบแทนของวันทำการถัดไปและผลตอบแทนสิ้นวันนั้นๆ

กลุ่มอุตสาหกรรมย่อย	ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนกับ	
	ผลตอบแทนของวันทำการถัดไป	ผลตอบแทนสิ้นวันนั้นๆ
ธุรกิจการเกษตร (AGRI)	0.0376	0.1390
อาหารและเครื่องดื่ม (FOOD)	0.0456	0.0322
แฟชั่น (FASHION)	-0.0079	0.1465
ของใช้ในครัวเรือนและสำนักงาน (HOME)	-0.0046	0.1681
ของใช้ส่วนตัวและเวชภัณฑ์ (PERSON)	0.0882	0.1381
ธนาคาร (BANK)	-0.1021	0.0866
เงินทุนและหลักทรัพย์ (FIN)	-0.1635	0.0752
ประกันภัยและประกันชีวิต (INSUR)	0.1510	0.0116
ยานยนต์ (AUTO)	-0.0387	-0.0638
วัสดุอุตสาหกรรมและเครื่องจักร (IMM)	-0.0790	0.0496
กระดาษและวัสดุการพิมพ์ (PAPER)	0.0168	0.1772
ปิโตรเคมีและเคมีภัณฑ์ (PETRO)	0.0684	-0.0643
บรรจุภัณฑ์ (PKG)	0.0461	0.2030
เหล็ก (STEEL)	0.1236	0.3026
วัสดุก่อสร้าง (CONMAT)	0.0513	0.0483
พัฒนาอสังหาริมทรัพย์ (PROP)	-0.1223	0.0639
กองทุนรวมอสังหาริมทรัพย์และกองทรัสต์เพื่อการลงทุนในอสังหาริมทรัพย์ (PF&REIT)	-0.0386	-0.0584
บริการรับเหมาก่อสร้าง (CONS)	0.1215	0.0612
พลังงานและสาธารณูปโภค (ENERG)	-0.1038	-0.0351
เหมืองแร่ (MINE)	-0.0918	-0.1257
พาณิชย์ (COMM)	-0.1509	0.2174
การแพทย์ (HELTH)	-0.0623	-0.0481
สื่อและสิ่งพิมพ์ (MEDIA)	-0.0139	0.0132
บริการเฉพาะกิจ (PROF)	0.0492	0.0043
การท่องเที่ยวและสันทนาการ (TOURISM)	-0.1464	-0.0014
ขนส่งและโลจิสติกส์ (TRANS)	0.0241	0.0422
ชิ้นส่วนอิเล็กทรอนิกส์ (ETRON)	0.0192	-0.0894
เทคโนโลยีสารสนเทศและการสื่อสาร (ICT)	0.0337	0.1989

ตารางที่ 4.8 จำนวนและสัดส่วนของคำแนะนำจำแนกตามกลุ่มอุตสาหกรรมย่อย

กลุ่มอุตสาหกรรมย่อย	จำนวนคำแนะนำ	ร้อยละของคำแนะนำทั้งหมด
ธุรกิจการเกษตร (AGRI)	337	1.61%
อาหารและเครื่องดื่ม (FOOD)	1,661	7.94%
แฟชั่น (FASHION)	41	0.20%
ของใช้ในครัวเรือนและสำนักงาน (HOME)	130	0.62%
ของใช้ส่วนตัวและเวชภัณฑ์ (PERSON)	40	0.19%
ธนาคาร (BANK)	1,543	7.37%
เงินทุนและหลักทรัพย์ (FIN)	712	3.40%
ประกันภัยและประกันชีวิต (INSUR)	137	0.65%
ยานยนต์ (AUTO)	264	1.26%
วัสดุอุตสาหกรรมและเครื่องจักร (IMM)	166	0.79%
กระดาษและวัสดุการพิมพ์ (PAPER)	48	0.23%
ปิโตรเคมีและเคมีภัณฑ์ (PETRO)	596	2.85%
บรรจุภัณฑ์ (PKG)	217	1.04%
เหล็ก (STEEL)	329	1.57%
วัสดุก่อสร้าง (CONMAT)	882	4.22%
พัฒนาอสังหาริมทรัพย์ (PROP)	2,013	9.62%
กองทุนรวมอสังหาริมทรัพย์และกองทรัสต์เพื่อการลงทุนในอสังหาริมทรัพย์ (PF&REIT)	53	0.25%
บริการรับเหมาก่อสร้าง (CONS)	1,308	6.25%
พลังงานและสาธารณูปโภค (ENERG)	3,210	15.34%
เหมืองแร่ (MINE)	10	0.05%
พาณิชย์ (COMM)	2,048	9.79%
การแพทย์ (HELTH)	792	3.79%
สื่อและสิ่งพิมพ์ (MEDIA)	653	3.12%
บริการเฉพาะกิจ (PROF)	98	0.47%
การท่องเที่ยวและสันทนาการ (TOURISM)	215	1.03%
ขนส่งและโลจิสติกส์ (TRANS)	1,570	7.50%
ชิ้นส่วนอิเล็กทรอนิกส์ (ETRON)	513	2.45%
เทคโนโลยีสารสนเทศและการสื่อสาร (ICT)	1,338	6.39%
รวม	**20,924**	**100%**

บทที่ 5
บทสรุปและข้อเสนอแนะ

5.1 อภิปรายและสรุปผล

จากการศึกษาความสัมพันธ์ของค่าจริงกับค่าคาดหวังที่คำนวณได้ในกรณีต่างๆ พบว่า ค่าคาดหวังแบบมีเงื่อนไขบนข้อมูลอันดับที่สมบูรณ์ของผลตอบแทน (Perfect Rank) จะมีระยะห่างจากค่าจริงน้อยที่สุด รองลงมาเป็นค่าคาดหวังแบบมีเงื่อนไขบนข้อมูลอันดับของตัวแปรร่วม (Imperfect Rank) และค่าคาดหวังแบบมีเงื่อนไข*โดยนัย*ของข้อมูลอันดับของตัวแปรร่วม (Implied Rank) ตามลำดับ โดยผลที่ได้จะเป็นเช่นนี้ไม่ว่าจำนวนข้อมูล ค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรร่วม (ρ) หรือระดับความสมบูรณ์ของข้อมูลอันดับของตัวแปรร่วม (λ) จะเปลี่ยนแปลงไปอย่างไร

เมื่อค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรร่วม (ρ) มีค่าน้อย และระดับความสมบูรณ์ของข้อมูลอันดับของตัวแปรร่วม (λ) มีค่ามาก จะทำให้ค่าคาดหวังทั้งสามกรณีมีค่าใกล้เคียงกันมากขึ้น โดยเฉพาะค่าคาดหวังแบบมีเงื่อนไขบนข้อมูลอันดับของตัวแปรร่วม (Imperfect Rank) และค่าคาดหวังแบบมีเงื่อนไข*โดยนัย*ของข้อมูลอันดับของตัวแปรร่วม (Implied Rank) ที่จะมีระยะห่างจากค่าจริงใกล้เคียงกันมากเมื่อระดับความสมบูรณ์ของข้อมูลอันดับของตัวแปรร่วม (λ) มีค่าสูง

นอกจากนี้ระยะห่างระหว่างค่าในเวกเตอร์ค่าเฉลี่ยก็เป็นอีกปัจจัยหนึ่งที่มีผลต่อค่าคาดหวังแบบมีเงื่อนไข โดยเมื่อพิจารณากรณีที่ระดับความสมบูรณ์ของข้อมูลอันดับของตัวแปรร่วมอยู่ในระดับปานกลาง และเปรียบเทียบผลเมื่อความแปรปรวนของตัวแปรเท่ากัน หากระยะห่างระหว่างค่าในเวกเตอร์ค่าเฉลี่ยเพิ่มขึ้น ค่าคาดหวังทั้งสามกรณีมีค่าแตกต่างจากค่าของตัวแปรที่สนใจศึกษาใกล้เคียงกัน

ข้อสรุปที่ได้จากศึกษาในส่วนนี้สามารถนำไปประยุกต์ใช้ในสถานการณ์จริงได้ว่า หากต้องทำการอนุมานเชิงสถิติโดยใช้ค่าคาดหวังแบบมีเงื่อนไขของข้อมูลเชิงอันดับ แต่ไม่สามารถหาข้อมูลอันดับที่สมบูรณ์ของตัวแปรที่สนใจศึกษาได้ ก็สามารถนำข้อมูลอันดับของตัวแปรร่วมมาใช้อนุมานได้ โดยผลที่ได้จะใกล้เคียงกับการใช้ข้อมูลอันดับที่สมบูรณ์ เมื่อจำนวนตัวแปรที่สนใจ (N) ไม่มากนัก ค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรร่วม (ρ) มีค่าน้อย ระดับความสมบูรณ์ของข้อมูลอันดับของตัวแปรร่วม (λ) มีค่ามาก และค่าในเวกเตอร์ค่าเฉลี่ยมีระยะห่างจากกันมาก โดยหากทราบลักษณะการแจกแจงของตัวแปรร่วมก็ควรใช้วิธีหาค่าคาดหวังแบบมีเงื่อนไขบนข้อมูลอันดับของตัวแปรร่วม (Imperfect Rank) เนื่องจากค่าที่ได้จะใกล้เคียงกับข้อมูลที่ถูกต้องมากกว่า แต่ในบางกรณีที่ทราบเพียงข้อมูลอันดับของตัวแปรร่วม โดยไม่ทราบลักษณะการแจกแจง การหาค่าคาดหวังแบบมีเงื่อนไข*โดยนัย*ของข้อมูลอันดับของตัวแปรร่วม (Implied Rank) ยังพอใช้เป็นทางเลือกในการทำการอนุมาน

ได้ ทั้งนี้ผู้ทำการตัดสินใจควรคำนึงถึงความสูญเสียที่เกิดขึ้นจากความผิดพลาดของการอนุมานในกรณีต่างๆ ด้วย

ผลการจำลองข้อมูลเพื่อศึกษาการจัดพอร์ตการลงทุน ได้ผลสอดคล้องกับข้อสรุปข้างต้น โดยการใช้ค่าคาดหวังของข้อมูลอันดับที่สมบูรณ์ของผลตอบแทน (Perfect Rank) ให้อรรถประโยชน์เฉลี่ยใกล้เคียงกับการใช้ข้อมูลผลตอบแทนจริงมากที่สุด ส่วนการใช้ตัวแปรร่วม ผลการศึกษาพบว่าการใช้ข้อมูลลักษณะการแจกแจงของตัวแปรดังกล่าวร่วมกับข้อมูลอันดับแล้วมาคำนวณค่าคาดหวังแบบมีเงื่อนไขบนข้อมูลอันดับของตัวแปรร่วม (Imperfect Rank) จะให้ค่าอรรถประโยชน์เฉลี่ยสูงกว่าการใช้ข้อมูลผลตอบแทนในอดีต (Prior) ส่วนในกรณีที่มีเพียงข้อมูลอันดับที่ได้จากความคิดเห็นของนักลงทุน หากข้อมูลนั้นมีระดับความสมบูรณ์ของข้อมูลอันดับของตัวแปรร่วม (λ) มีค่ามากพอในระดับหนึ่ง ผลการจัดพอร์ตการลงทุนที่ใช้ค่าคาดหวังแบบมีเงื่อนไข*โดยนัย*ของข้อมูลอันดับของตัวแปรร่วม (Implied Rank) มาคำนวณก็ยังทำให้ได้ประสิทธิภาพสูงกว่าการใช้ข้อมูลผลตอบแทนในอดีต (Prior) เพียงอย่างเดียว

งานวิจัยส่วนสุดท้ายที่ศึกษาผลการจัดพอร์ตการลงทุนเมื่อใช้ข้อมูลจริง โดยใช้ข้อมูลผลตอบแทนในอดีตร่วมกับข้อมูลอันดับที่รวบรวมจากบทวิเคราะห์ของบริษัทหลักทรัพย์ต่างๆ พบว่า ยิ่งเราสามารถหาวิธีการรวมรวมและสร้างข้อมูลอันดับของตัวแปรร่วมที่มีระดับความสัมพันธ์กับข้อมูลผลตอบแทนได้มากเท่าไหร่ ประสิทธิภาพของการจัดพอร์ตการลงทุนก็จะเพิ่มขึ้นตามไปด้วย โดยสำหรับข้อมูลกลุ่มอุตสาหกรรมหลักในช่วงเวลาที่นำมาวิเคราะห์ วิธีการรวบรวมข้อมูลที่ได้ผลดีที่สุด คือ การถ่วงน้ำหนักด้วยสัดส่วนของมูลค่าของหุ้นที่ได้รับคำแนะนำเทียบกับมูลค่าของกลุ่มอุตสาหกรรมนั้นๆ และความแม่นยำของคำแนะนำในอดีตโดยประเมินย้อนหลังไป 30 วันทำการ จะทำให้ได้อรรถประโยชน์สะสมสูงกว่าการใช้ข้อมูลจากการแจกแจงก่อนเพียงอย่างเดียว

อย่างไรก็ตาม เมื่อทดลองใช้วิธีดังกล่าวกับกลุ่มอุตสาหกรรมย่อย กลับพบว่าผลการจัดพอร์ตการลงทุนเมื่อใช้ตัวแปรร่วมได้รับอรรถประโยชน์ต่ำกว่าการแจกแจงก่อน แสดงให้เห็นว่า การเลือกและปรับใช้วิธีการให้เหมาะสมกับลักษณะของข้อมูลแต่ละชุด ส่งผลต่อประสิทธิภาพของการจัดพอร์ตการลงทุนเป็นอย่างมาก

5.2 ข้อเสนอแนะ

งานวิจัยส่วนแรกที่ทำการจำลองข้อมูลเพื่อศึกษาปัจจัยที่มีผลต่อค่าคาดหวังแบบต่างๆ แม้ว่าจะได้เห็นลักษณะความสัมพันธ์ของปัจจัยหลักๆ ที่สำคัญแล้ว แต่ก็ยังมีรายละเอียดเพิ่มเติมในประเด็นอื่นๆ ให้ศึกษาต่อไปอีกได้ เช่น

ในส่วนของการอนุมานทางสถิติโดยใช้ค่าคาดหวังของข้อมูลอันดับ

- การกำหนดรูปแบบความสัมพันธ์ของตัวแปร ในที่นี้กำหนดให้ตัวแปรต่างๆ มีการแจกแจงแบบปกติหลายตัวแปร และกำหนดให้ตัวแปรร่วมและค่าความผิดพลาดเป็นอิสระกัน อาจลองพิจารณาการแจกแจงในรูปแบบอื่น หรืออาจศึกษากรณีที่ค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรร่วมแต่ละคู่มีค่าแตกต่างกัน
- การวัดค่าความแตกต่างของค่าคาดหวังกับค่าของตัวแปรที่สนใจศึกษา โดยงานวิจัยนี้เลือกใช้ Euclidean distance เนื่องจากเป็นค่าที่เข้าใจและคำนวณได้ง่าย และยังสะท้อนให้เห็นถึงขนาดของความผิดพลาดที่เพิ่มขึ้นได้ชัดเจน อย่างไรก็ตาม ยังมีค่าสถิติอื่นๆ ที่ใช้วัดความสัมพันธ์ของค่าคาดหวังกับค่าของตัวแปรได้อีกหลายแบบให้เลือกใช้ตามความเหมาะสมกับลักษณะของข้อมูลและวัตถุประสงค์การนำไปใช้งาน

ในส่วนของการจัดพอร์ตการลงทุน

- สำหรับข้อมูลผลตอบแทนในอดีต งานวิจัยนี้เสนอผลศึกษาโดยใช้ข้อมูลผลตอบแทนรายวันของกลุ่มหลักทรัพย์ในตลาดหลักทรัพย์แห่งประเทศไทย ในช่วง 6 เดือน ซึ่งอาจทดลองเปลี่ยนหลักทรัพย์ที่นำมาพิจารณา เป็นหลักทรัพย์ในตลาดต่างประเทศ หรืออาจเป็นสินทรัพย์ประเภทอื่นๆ รวมทั้งอาจลองพิจารณาผลตอบแทนในช่วงเวลาที่ยาวขึ้น และอาจปรับความถี่เป็นดูผลตอบแทนรายเดือน หรือหากมีข้อมูลการเปลี่ยนแปลงของราคาเป็นรายนาทีหรือรายชั่วโมง ก็จะทำให้สามารถศึกษาความเร็วในการการตอบสนองของราคาที่มีหลังมีการเผยแพร่รายงานหรือบทวิเคราะห์แต่ละฉบับได้ชัดเจนขึ้น
- สำหรับข้อมูลจากบทวิเคราะห์ อาจทดลองใช้คำแนะนำการซื้อขายหุ้นจากแหล่งอื่นๆ เช่น การดึงข้อมูลความคิดเห็นของนักลงทุนจากอินเตอร์เน็ต หรือ social network ต่างๆ รวมทั้งข้อมูลจากบทวิเคราะห์ที่งานวิจัยนี้เลือกเก็บข้อมูลเฉพาะคำแนะนำให้ซื้อขายในระยะสั้น งานวิจัยต่อไปอาจลองศึกษาผลของคำแนะนำระยะยาวต่อประสิทธิภาพจัดพอร์ตการลงทุน
- หากสามารถเก็บข้อมูลได้จำนวนมากพอสมควร น่าจะทดลองประมาณการแจกแจงของตัวแปรร่วมและหาค่าคาดหวังแบบมีเงื่อนไขบนข้อมูลอันดับของตัวแปรร่วม (Imperfect Rank) ซึ่งผลจากการจำลองข้อมูลพบว่ามีประสิทธิภาพเหนือกว่าข้อมูลจากการแจกแจงก่อนเสมอ แต่ยังไม่ได้ทดลองผลส่วนนี้กับข้อมูลจริง

สุดท้ายการทดลองจัดอันดับหลักทรัพย์จากข้อมูลบทวิเคราะห์ทำให้เห็นปัญหาของข้อมูลที่เกิดขึ้น รวมทั้งประเด็นที่สามารถพัฒนาเพิ่มความถูกต้องของข้อมูลอันดับที่ได้จากการรวบรวมความเห็นจากแหล่งต่างๆ ดังต่อไปนี้

- ปัญหาความไม่ครบถ้วนของข้อมูล เช่น บางกลุ่มอุตสาหกรรมที่ไม่ค่อยมีคำแนะนำให้ซื้อหรือขาย ทำให้เวลาจัดอันดับเกิดข้อผิดพลาดได้ง่าย รวมทั้งการประเมินความแม่นยำในอดีตก็ทำได้ยากอีกด้วย วิธีการหนึ่งที่อาจลดปัญหานี้ได้ คือ การพิจารณาข้อมูลอันดับที่ไม่สมบูรณ์ที่เรียกว่า Partial Rank หรือพิจารณากรณีที่เกิด tied rank ซึ่งในการคำนวณและรวบรวมข้อมูลจำเป็นต้องพัฒนาและปรับวิธีการต่างๆ ให้รองรับข้อมูลประเภทนี้ด้วย
- การกำหนดค่าพารามิเตอร์ในการถ่วงน้ำหนัก งานศึกษานี้ในแต่ละวิธีจะถ่วงน้ำหนักด้วยค่าคงที่ตลอดช่วงเวลาที่พิจารณาและใช้การลองผิดลองถูกในการหาค่าที่เหมาะสม เพื่อแสดงให้เห็นผลเมื่อระดับความสัมพันธ์ของตัวแปรร่วมและตัวแปรที่สนใจศึกษาแตกต่างกันไป หากลองใช้เทคนิคการ learning และการหาค่าที่ดีที่สุด (optimal) มาช่วยในการกำหนดค่าที่ใช้ถ่วงน้ำหนัก น่าจะทำให้ข้อมูลอันดับของตัวแปรร่วมใกล้เคียงกับข้อมูลอันดับที่ถูกต้องมากขึ้น ซึ่งจะทำให้ประสิทธิภาพของการจัดพอร์ตการลงทุนสูงขึ้นด้วย รวมทั้งการกำหนดช่วงเวลาของข้อมูลที่นำมาประเมินความแม่นยำของการทำนายในอดีตที่เหมาะสมก็เป็นเรื่องสำคัญ เนื่องจากหากกำหนดช่วงสั้นเกินไปก็มักมีปัญหาเรื่องความไม่ครบถ้วนของข้อมูล แต่หากกำหนดช่วงยาวเกินไป ก็อาจไม่สะท้อนถึงความถูกต้องในช่วงเวลาที่สนใจศึกษา
- วิธีประเมินความแม่นยำของคำแนะนำในอดีต ในที่นี้ใช้ค่าสัมประสิทธิ์สหสัมพันธ์ในการประเมิน ซึ่งมีข้อดีที่สามารถวัดระดับความสัมพันธ์และบอกทิศทางได้ด้วย แต่วิธีนี้ก็มีข้อจำกัดดังที่เห็นในผลการศึกษาข้างต้นว่า หากข้อมูลที่ใช้มีรายการข้อมูลที่ไม่มากและไม่ได้กระจายไปครบทุกกลุ่มอุตสาหกรรมที่สนใจ จะทำให้ผลการประเมินได้ผลไม่ดีนัก รวมทั้งการประเมินทีละกลุ่มอุตสาหกรรมบางครั้งอาจทำให้ผลคะแนนในภาพรวมเมื่อมองเปรียบเทียบระหว่างกลุ่มผิดพลาดไปได้

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

รายการอ้างอิง

Ang, A. (2014). *Asset Management: A Systematic Approach to Factor Investing*: Oxford University Press.

Antunovich, P., & Sarkar, A. (2006). Fifteen Minutes of Fame? The Market Impact of Internet Stock Picks. *Journal of Business, 79*(6), 3209-3251.

Black, F., & Litterman, R. B. (1991). Asset Allocation: Combining Investor Views with Market Equilibrium. *The Journal of Fixed Income, 1*(2), 7-18

Bohn, L. L., & Wolfe, D. A. (1994). The Effect of Imperfect Judgment Rankings on Properties of Procedures Based on the Ranked-Set Samples Analog of the Mann-Whitney-Wilcoxon Statistic. *Journal of the American Statistical Association, 89*(425), 168-176.

Chen, H., Stasny, E. A., & Wolfe, D. A. (2006). An empirical assessment of ranking accuracy in ranked set sampling. *Computational Statistics & Data Analysis, 51*(2), 1411-1419.

Chen, Z., Bai, Z., & Sinha, B. K. (2004). *Ranked Set Sampling: Theory and Applications*: New York, NY : Springer New York : Imprint: Springer.

Chen, Z., & Shen, L. (2003). Two-layer ranked set sampling with concomitant variables. *Journal of Statistical Planning and Inference, 115*, 45-57.

Chiarawongse, A., Kiatsupaibul, S., Tirapat, S., & Roy, B. V. (2012). Portfolio selection with qualitative input. *Journal of Banking and Finance, 36*, 489-496.

David, H. A., & Levine, D. N. (1972). Ranked set sampling in the presence of judgment error. *Biometrics, 28*, 553-555.

Davis-Stober, C. P., Budescu, D. V., Dana, J., & Broomell, S. B. (2014). When is a crowd wise? *Decision, 1*(2), 79-101.

Dell, T. R., & Clutter, J. L. (1972). Ranked Set Sampling Theory with Order Statistics Background. *Biometrics, 28*(2), 545-553.

Hayter, A. J. (2006). Recursive integration methodologies with statistical applications. *Journal of Statistical Planning and Inference, 136*(7), 2284-2296.

Hill, S., & Ready-Campbell, N. (2011). Expert Stock Picker: The Wisdom of (Experts in) Crowds. *International Journal of Electronic Commerce, 15*(3), 73-102.

Idzorek, T. (2007). A step-by-step guide to the Black-Litterman model *Forecasting Expected Returns in the Financial Markets* (pp. 17-38). Oxford: Academic Press.

Kiatsupaibul, S., J. Hayter, A., & Liu, W. (2017). Rank constrained distribution and moment computations. *Computational Statistics and Data Analysis, 105*, 229-242.

Lee, M. D., Steyvers, M., & Miller, B. (2014). A Cognitive Model for Aggregating People's Rankings. *PLoS ONE, 9*(5), e96431.

Levy, H., & Levy, A. (1991). Arrow-Pratt Measures of Risk Aversion: The Multivariate Case. *International Economic Review, 32*(4), 891-898.

Markowitz, H. (1952). Portfolio Selection. *Journal of Finance, 7*(1), 77-91.

McIntyre, G. (1952). A method for unbiased selective sampling, using ranked sets. *Australian Journal of Agricultural Research, 3*(4), 385-390.

Menezes, L. M., W. Bunn, D., & Taylor, J. W. (2000). Review of guidelines for the use of combined forecasts. *European Journal of Operational Research, 120*(1), 190-204.

Patil, G. P., Surucu, B., & Egemen, D. (2006). Ranked set sampling *Encyclopedia of Environmetrics*: John Wiley & Sons, Ltd.

Pratt, J. W. (1964). Risk Aversion in the Small and in the Large. *Econometrica, 32*(1/2), 122-136.

Satchell, S., & Scowcroft, A. (2003). *Advances in Portfolio Construction and Implementation* (1st ed.). Amsterdam: Butterworth-Heinemann.

Savage, I. R. (1957). Contributions to the Theory of Rank Order Statistics--The "Trend" Case. *The Annals of Mathematical Statistics*, 968-977.

Schalekamp, F., & Zuylen, A. (2009). Rank Aggregation: Together We're Strong *2009 Proceedings of the Eleventh Workshop on Algorithm Engineering and Experiments (ALENEX)* (pp. 38-51): Society for Industrial and Applied Mathematics.

Stokes, S. L. (1977). Ranked set sampling with concomitant variables. *Communications in Statistics - Theory and Methods, 6*(12), 1207-1211.

Tong, Y. L. (1990). *The Multivariate Normal Distribution. [electronic resource]*: New York, NY : Springer New York, 1990.

Wang, G., Wang, T., Wang, B., Sambasivan, D., Zhang, Z., Zheng, H., & Zhao, B. Y. (2015). *Crowds on Wall Street: Extracting Value from Collaborative Investing Platforms*. Paper presented at the Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, Vancouver, BC, Canada.

ปริยากร มณีกูล. (2015). ผลกระทบของค่าสัมประสิทธิ์สหสัมพันธ์ต่อการอนุมานเชิงสถิติแบบมีข้อมูลประกอบเชิงอันดับในการจัดพอร์ตการลงทุน: จุฬาลงกรณ์มหาวิทยาลัย.

สุพัตรา เพ็ชรน้ำขาว. (2015). การอนุมานเชิงสถิติแบบมีข้อจำกัดเชิงอันดับสำหรับการจัดพอร์ตลงทุน: จุฬาลงกรณ์มหาวิทยาลัย.

เสกสรร เกียรติสุไพบูลย์. (2012). การจำลอง: กรุงเทพฯ : สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย, 2555. พิมพ์ครั้งที่ 1.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวชุติมน สินธุประมา เกิดเมื่อวันที่ 2 พฤศจิกายน พ.ศ. 2527 สำเร็จการศึกษาปริญญาเศรษฐศาสตร์บัณฑิต (เกียรตินิยมอันดับ 1) วิชาเอกเศรษฐศาสตร์การเงินและเศรษฐมิติ จากมหาวิทยาลัยธรรมศาสตร์ และ Master of Science in Information จาก University of Michigan เข้าศึกษาต่อในหลักสูตรวิทยาศาสตร์มหาบัณฑิต สาขาวิชาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2558

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University