การเปรียบเทียบวิธีการการสกัดปัจจัยเพื่อการพยากรณ์ราคาทองคำ

บุญมี พันธวงศ์ และ กฤษณะ ชินสาร

คณะวิทยาการสารสนเทศ มหาวิทยาลัยบูรพา อำเภอเมือง จังหวัดชลบุรี 20131

E-mail: bounmychup@yahoo.com, ckisana@gmail.com

บทคัดย่อ

งานวิจัยนี้เป็นงานวิจัยเชิงเปรียบเทียบวิธีการสกัดคุณลักษณะ (ปัจจัย) ที่มีความสำคัญต่อการพยากรณ์ราคาทองคำ โดยผู้วิจัยนำเสนอวิธีการสกัดปัจจัย 5 วิธี ประกอบด้วย(1) วิธีการ Partial Least Square Regression (PLSR), (2) การวิเคราะห์องค์ประกอบหลัก (PCA), (3) วิธีผสมระหว่าง PCA กับ Stepwise Selection, (4) วิธีการวิเคราะห์การถดถอยหลายตัวแปรแบบ *Enter Selection* และ (5) วิธีการวิเคราะห์การถดถอยหลายตัวแปรแบบ Stepwise *Selection* จากนั้นทำการเปรียบเทียบประสิทธิภาพของการสกัดปัจจัยโดยนำปัจจัยที่ได้เข้าสู่การพยากรณ์ด้วยโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ (*BPNN*) และ แบบฟังก์ชันรัศมีฐาน (*RBF*NN) โดยการศึกษาจะพิจารณาจากผลของการพยากรณ์ กล่าวคือ วิธีการสกัดปัจจัยที่ให้ผลการพยากรณ์ที่สูงกว่าจะถือว่าวิธีการสกัดปัจจัยนั้นมีประสิทธิภาพมากกว่าวิธีการสกัดปัจจัยแบบอื่น จากผลการทดลอง พบว่า เมื่อนำปัจจัยราคาทองคำ 18 ปัจจัย (มิติ) ผ่านกระบวนการสกัดคุณลักษณะที่กล่าวมาข้างต้น ได้ปัจจัยที่เหมาะสมกับการรพยากรณ์ คือ 6 มิติ, 9 มิติ, 6 มิติ, 5 มิติ และ 9 มิติ ตามลำดับ นอกจากนี้ ผู้วิจัยพบว่า วิธีการ PLSR เป็นวิธีการสกัดปัจจัยที่ให้ผลการพยากรณ์สูงสุด โดยให้ค่าร้อยละของค่าความคลาดเคลื่อนสมบูรณ์ (MAPE) ร้อยละ 0.686 และ 0.384 สำหรับการพยากรณ์ด้วย BPNN และ RBFNN ตามลำดับ

Key Words : PLSR, PCA, PCA-Stepwise, MLR-Enter Selection, and Stepwise Selection.

1. บทนำ

การพยากรณ์การเคลื่อนไหวของราคาทองคำที่มีประสิทธิภาพความแม่นยำสูงนั้นก็เนื่องมาจากปัจจัยที่ส่งผลกระทบต่อการเคลื่อนไหวของราคาทองคำเมื่อได้ปัจจัยที่มีความสัมพันธ์สูงกับราคาทองก็จะทำให้ได้ผลการพยากรณ์ที่ถูกต้องแม่นยำสูง เพราะฉะนั้นจึงมีการศึกษาวิจัยและพัฒนาวิธีการสกัดปัจจัยหรือลดมิติข้อมูลด้วยหลายวิธีที่แตกต่างกันโดยมีจุดประสงค์เพื่อการพยากรณ์ที่มีประสิทธิภาพความแม่นยำสูง

Ying Sai , Shiwei Zhu และ Tao Zhang [8] นำเสนอวิธีการพยากรณ์ปัญหาทางการเงินโดยใช้วิธีแบบผสมระหว่าง PCA และ RBFNN จากผลการศึกษา พบว่าวิธีการที่นำเสนอให้ค่า MAPE เท่ากับ 89.16% แต่ในการใช้วิธีการของ PCA ในการสกัดคุณลักษณะก็จะได้คุณลักษณะตัวใหม่ที่ไม่มีความสัมพันธ์กับตัวแปรเกณฑ์ซึ่งจะทำให้เกิดความคลาดเคลื่อนในการพยากรณ์ตัวแปรเกณฑ์

Z. Ismail, A. Yahya and A. Shabri. [9] นำเสนอวิธีการพยากรณ์ราคาทองคำโดยใช้ Multi Linear Regression (MLR) ปัจจัยที่ใช้ในการพยากรณ์มีทั้งหมด 8 ปัจจัย สร้างตัวแบบด้วยวิธี MLR และวิธี A Naïve model ตัวแบบแรก ได้ 3 ปัจจัย ตัวแบบที่สอง ได้ 4 ปัจจัยที่พร้อมจะพยากรณ์ เมื่อวัดประสิทธิ์ภาพของตัวแบบพบว่าตัวแบบแรกและตัวแบบที่สองได้ค่าเฉลี่ยของค่าความผิดพลาดกำลังสอง (MSE) เท่ากับ 96.923% และ 221.88% ตามลำดับ การสร้างสมการพยากรณ์ด้วยวิธีดังกล่าวเมื่อมีปัจจัยมากก็จะทำให้เกิดปัญหาปัจจัยสัมพันธ์กันเอง (Multicollinearity) ก็จะทำให้การพยากรณ์ตัวแปรเกณฑ์มีความคลาดเคลื่อนมาก

วิภาดา เวทย์ประสิทธิ์ และ พรพิมล ณ นคร [5] นำเสนอวิธีการสกัดปัจจัยสำหรับการพยากรณ์น้ำฝนโดยใช้โครงข่ายประสาทเทียมโดยตัวแบบ BPNN และตัวแบบ RBFNN เป็น 4:4:1โดยมีปัจจัยนำเข้าที่เหมาะสม 4 ปัจจัย ผลการทดลองพบว่าวิธีการที่นำเสนอ ให้ค่าความถูกต้อง 88.00% และ 84.91% ตามลำดับ

Alexandra Debiolles และคณะ [6] ได้นำเสนอวิธีการรวมในการวินิจฉัยโรคด้วยวิธี PLSR และโครงข่ายประสาทเทียม ผลการทดลองพบว่าค่าความคลาดเคลื่อนมาตรฐานของการพยากรณ์ (MSE) มีค่าต่ำสุดเท่ากับ 0.7%

จากวิธีการและปัญหาในการสกัดปัจจัยที่แตกต่างกันในแต่ละงานวิจัยและการพยากรณ์ที่ให้ประสิทธิ์ภาพความแม่นยำสูงของBPNN และ RBFNN งานวิจัยนี้ได้นำ เสนอการเปรียบเทียบวิธีการสกัดปัจจัยเพื่อการพยากรณ์ราคาทองคำด้วยการนำเทคนิคการสกัดปัจจัยด้วยวิธี PLSR, PCA, PCA-Stepwise, Enter Selection และ Stepwise Selection แล้วนำผลที่ได้ไปพยากรณ์ด้วยโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับและแบบฟังก์ชันรัศมีฐาน เพื่อวัดประสิทธิภาพของวิธีการสกัดปัจจัย

2. ทฤษฎีที่เกี่ยวข้อง

2.1 การวิเคราะห์องค์ประกอบหลัก PCA

การวิเคราะห์องค์ประกอบหลักด้วย (PCA) [8] เป็นเทคนิคที่ใช้การในการลดมิติของข้อมูลที่มีจำนวนมากโดยการสร้างตัวแปรใหม่ซึ่งเป็นฟังก์ชันเชิงเส้นของตัวแปรเดิมโดยดึงรายละเอียดหรือค่าแปรปรวนจากตัวแปรเดิมมาไว้ในตัวแปรใหม่ให้มากที่สุดโดยการฉาย (project) เวกเตอร์ไปบนแกนใหม่ที่เรียกว่าแกนองค์ประกอบหลัก (Principal Component) ซึ่งแกนเหล่านี้มีความสำคัญแตกต่างกันลงไปตามความแปรปรวน(variance) บนแต่ละแกน กำหนดให้ข้อมูลที่ใช้สอน คือ $X_k \in R^D, k = 1,2,...,M$ โดย

D คือ มิติของ X_k และ M คือจำนวนข้อมูล และมีขั้นตอนการหาองค์ประกอบหลักดังนี้

ขั้นตอนที่ 1 หาค่าเมตริกซ์ความแปรปรวนร่วม (Covariance matrix, C) หาค่าเฉลี่ยแต่ละคอลัมน์

$$\mu = \frac{1}{M}\sum_{i=1}^{M} X_i \quad (1)$$

หาค่าความแตกต่างระหว่างตัวแปรพยากรณ์กับค่าเฉลี่ย

$$\phi_j = X_j - \mu \quad (2)$$

สร้างเมตริกซ์ของความแปรปรวนร่วม

$$C = \frac{1}{M}\sum_{j=1}^{M} \phi_j \phi_j^T \quad (3)$$

ขั้นตอนที่ 2

หาค่าเจาะจง (eigenvalue , λ)

$$\lambda = CC^T \quad (4)$$

ค่าเวกเตอร์เจาะจง (eigenvector, v)

$$v = \frac{\phi C}{\sqrt{\lambda}} \quad (5)$$

ค่าเวกเตอร์เจาะจง (eigenvector, v) หรือแกนองค์ประกอบหลัก จาก C กำหนดให้ $\lambda_1,...,\lambda_N (\lambda_i > 0)$ เป็นค่าเจาะจงจำนวน N ค่าแรกที่เรียงจากมากไปน้อยและ $v_1,...,v_N$ เป็นเวกเตอร์เจาะจงที่ตรงกับค่าเจาะจงซึ่งค่าของ N จะมีค่าน้อยกว่าค่าของ D และค่าเจาะจง λ และเวกเตอร์เจาะจง v ข้อมูลทั้งสองมีความสมนัย (Correspondence) ซึ่งกันและกัน

2.2 ขั้นตอนวิธี Partial Least Square Regression

เมื่อเราพิจารณาตัวแปรพยากรณ์ X ที่มีจำนวน n แถว และ มีจำนวน P สดมภ์ (มิติ) และ กำหนดให้ตัวแปรเกณฑ์ Y เป็นคำตอบของสมการ วิธีการทำงานของ PLSR [6] จะมีขั้นตอนในการทำงานหลักดังนี้

ขั้นตอนที่ 1 ดำเนินการให้ข้อมูลแต่ละมิติให้เป็นบรรทัดฐาน (Normalization) โดยค่าข้อมูลที่ได้จากการทำให้เป็นบรรทัดฐานจะเป็นค่า Z-score ของแต่ละมิติ ซึ่งจะทำให้ ค่าเฉลี่ยของข้อมูลเป็นศูนย์ และ ความแปรปรวนเป็นหนึ่ง

ขั้นตอนที่ 2 หาค่าองค์ประกอบแรก (T_1) ที่ได้จากสมการ (6)

$$T_1 = XW_1 \quad (6)$$

โดยที่ $W_1 = [w_{j1}]_{1\le j\le P}$

$$w_{j1} = \frac{Cov(X_j, Y)}{\sqrt{\sum_{i=1}^{P} Cov(X_i, Y)^2}} \quad (7)$$

ขั้นตอนที่ 3 สร้างสมการความถดถอยจากตัวแปรอิสระ P ตัว กับองค์ประกอบ T_1 ดังนี้

$$Y = C_1 T_1 + Y^{(1)} \quad (8)$$

$$X = T_1 D_1^T + X^{(1)} \quad (9)$$

ซึ่ง C_1 และ D_1 เป็นสัมประสิทธิ์ของความสัมพันธ์และเวคเตอร์ถดถอย (regression vector)

$Y^{(1)}$ คือเวคเตอร์เอาต์พุตที่เหลือ

$X^{(1)}$ คือเมทริกของตัวแปรอิสระที่เหลือ

ขั้นตอนที่ 4 ดำเนินการซ้ำขั้นตอนที่ 1 ถึง ขั้นตอนที่ 3 เพื่อคำนวณหาค่า $T_2,\ldots,T_p$ โดยพิจารณาจากค่าของ $X^{(1)}$ ในสมการที่ (9) ดังตัวอย่างต่อไปนี้ จะเป็นการแสดงการคำนวณค่าของ T_2

ถ้าเวคเตอร์ $Y^{(1)}$ ที่เหลือยังมีนัยสำคัญก็จะมีการคำนวณองค์ประกอบ T_2 แต่เนื่องจาก T_2 ไม่มีความสัมพันธ์กับ T_1 หรือ ทั้งสองตั้งฉากกัน เมื่อคำนวณความแปรปรวนรวมระหว่าง T_2 กับ T_2 คือ $Cov(T_1, T_2) = 0$ และ T_2 มีความสัมพันธ์กับเวคเตอร์ $Y^{(1)}$ ที่เหลือ

$$T_2 = X^{(1)} W_2 \quad (10)$$

$$W_2 = [w_{j2}]_{1\le j\le P}$$

โดยที่

$$w_{j2} = \frac{Cov(X_j^{(1)}, Y^{(1)})}{\sqrt{\sum_{i=1}^{P} Cov(X_i^{1}, Y^{(1)})^2}} \quad (11)$$

สมการถดถอยของทั้งสององค์ประกอบสามารถเขียนได้ ดังนี้

$$Y = C_1 T_1 + Y^{(1)} = C_1 T_1 + C_2 T_2 + Y^{(2)} \quad (12)$$

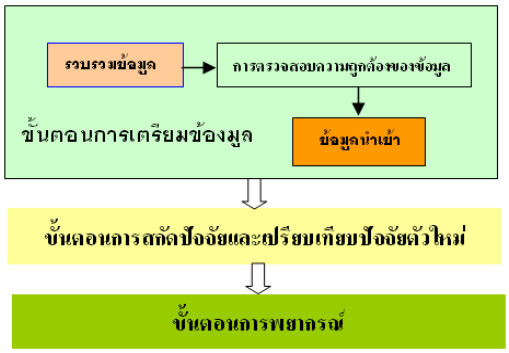
ทำตามขั้นตอนนี้ไปเรื่อย ๆ โดยใช้เทอมที่เหลือของความถดถอย (regression residual terms) ของทั้งตัวแปรพยากรณ์ และ ตัวแปรเกรซ์มาพิจารณาและทำไปจนถึง P_r

$$Y = \hat{Y} + \psi = TC + \psi \quad (13)$$

$$T = [T_1 \quad T_2 ... T_{P_r}] \quad C = [C_1 \quad C_2 ... C_{P_r}]^T$$

3. การดำเนินงานวิจัย

ในงานวิจัยนี้ได้นำเสนอวิธีการเปรียบเทียบการสกัดปัจจัยเพื่อการพยากรณ์ราคาทองคำ ขั้นตอนแรกเป็นขั้นตอนในการเตรียมข้อมูลได้แบ่งออกเป็น 3 ขั้นตอนย่อย ขั้นตอนการการสกัดปัจจัยแบ่งออกเป็น 2 ขั้นตอน คือ ขั้นตอนการสกัดและการเปรียบเทียบและขั้นตอนสุดท้ายคือขั้นตอนพยากรณ์ด้วยโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับและแบบฟังก์ชันรัศมีฐาน ดังที่ได้แสดงภาพรวมการทำงานในรูปที่ 1



รูปที่ 1 ขั้นตอนวิธีการเปรียบเทียบการสกัดปัจจัยเพื่อการพยากรณ์ราคาทองคำ

3.1 ขั้นตอนการเตรียมข้อมูล

3.1.1 การรวบรวมข้อมูล

ในงานวิจัยนี้ได้รวบรวมข้อมูลที่เป็นราคาทองคำย้อนหลัง โดยสมาคมค้าทองคำแห่งประเทศไทย [2] จำนวน 60 เดือนระหว่าง เดือนพฤศจิกายน 2546 ถึงเดือน ตุลาคม 2551 นอกจากนี้ปัจจัยภายนอกที่คิดว่าจะมีความสัมพันธ์กับราคาทองคำในประเทศไทยได้พิจารณาคัดเลือกปัจจัยภายในที่เป็นราคาทองคำและปัจจัยภายนอกที่ได้รวบรวมมาจากงานวิจัยของ ปรีดา อนุสรณ์ธีรกุลและ พุธษดี ศิริแสงตระกูล [3]

3.1.2 การตรวจสอบความถูกต้องของข้อมูล

ข้อมูลที่ได้คัดเลือกเพื่อนำมาใช้ในงานวิจัยนี้จากที่ได้กลาวมาในข้อ 3.1.1 ส่วนมากแล้วเป็นข้อมูลของตลาดหลักทรัพย์ซึ่งมีความแตกต่างกันในเรื่องของวันและเวลาเปิดทำการ ดังนั้นก่อนการนำข้อมูลเหล่านี้มาใช้งานจึงจำเป็นต้องมีการตรวจสอบและรวบรวมข้อมูลเฉพาะวันที่เปิดทำการในวันเวลาเดียวกันเพื่อให้เกิดความผิดพลาดน้อยที่สุด

3.1.3 ข้อมูลนำเข้า

ข้อมูลที่ได้รวบรวมมาจากหลายแหล่งที่แตกต่างกันเมื่อตรวจสอบความถูกต้องและทำให้มีสภาวะแวดล้อมที่เหมือนกันแล้วก่อนนำเข้าเราต้องทำการจัดรูปแบบข้อมูลทั้งหมดให้ตรงตามรูปแบบการประมวลผลของตัวโปรแกรมที่ผู้วิจัยได้ทำการเขียนขึ้นมา

3.2 ขั้นตอนการสกัดปัจจัยและเปรียบเทียบปัจจัยตัวใหม่

เมื่อนำปัจจัยทั้งหมดที่ผ่านกระบวนการเตรียมข้อมูลก็จะได้ปัจจัยที่พร้อมจะนำไปผ่านกระบวนการสกัดปัจจัยซึ่งประกอบด้วย 5 วิธีการ ดังนี้

(1) วิธีการ *Partial Least Square Regression (PLSR)* [6] ซึ่งมีขั้นตอนการคำนวณตามข้อ 2.2 , (2) การวิเคราะห์องค์ประกอบหลัก (PCA) [9] ขั้นตอนการคำนวณตามข้อ 2.1 , (3) วิธีผสมระหว่าง *PCA* กับ *Stepwise Selection*] ซึ่งเป็นการนำปัจจัยทั้งหมดไปผ่านการสกัดปัจจัยด้วย PCA นำผลที่ได้ไปผ่าน MLR ด้วยวิธี Stepwise Selection เพื่อหาปัจจัยที่มีความสัมพันธ์กับราคาทองคำที่เราจะพยากรณ์, (4) วิธีการวิเคราะห์การถดถอยหลายตัวแปรแบบ Enter Selection [1,9] ขั้นตอนนี้เป็นการหาความสัมพันธ์ระหว่างตัวแปรเกณฑ์กับปัจจัยทั้งหมดด้วยค่าทดสอบนัยสำคัญของสัมประสิทธิ์สหสัมพันธ์ และ (5) วิธีการวิเคราะห์การถดถอยหลายตัวแปรแบบ Stepwise Selection [1,4,9]

3.3 ขั้นตอนการพยากรณ์

นำปัจจัยตัวใหม่ที่ได้จากแต่ละวิธีการในข้อ 3.2 มาทำการสุ่มเลือกให้ได้เป็น 4 กลุ่มเพื่อใช้ในการฝึกสอนและทดสอบดังตารางที่ 1

ข้อมูล	กลุ่มที่1	กลุ่มที่ 2	กลุ่มที่ 3	กลุ่มที่ 4
ข้อมูลสำหรับฝึกสอน	60%	70%	80%	90%
ข้อมูลสำหรับทดสอบ	40%	30%	20%	10%

ตารางที่ 1 การแบ่งกลุ่มข้อมูลสำหรับการฝึกสอนและทดสอบ

ในการสร้างตัวแบบของโครงข่ายประสาทเทียม [7] ทั้งสองโครงข่ายในงาน วิจัยนี้ได้กำหนดดังนี้

ชั้นอินพุต การกำหนดจำนวนโหนดในชั้นนี้จะกำหนดตามจำนวนปัจจัยที่สกัดได้จากแต่ละวิธี แต่โครงข่ายประสาทเทียมแบบฟังก์ชันรัศมีฐานไม่มีการกำหนดจำนวนโหนดเพราะไม่เกิดการคำนวณในชั้นนี้

ชั้นซ่อน ในงานวิจัยนี้ให้ความสำคัญกับการกำหนดจำนวนนิวรอลในชั้นซ่อนมากกว่าการกำหนดจำนวนชั้นซ่อนของตัวแบบBPNN จำนวนโหนดได้กำหนดด้วยการทดลองซึ่งกำหนดจำนวนโหนดตั้งแต่ 2 โหนดถึง 6 โหนด แต่ RBFNN ได้กำหนดจำนวนนิวรอนในชั้นซ่อนเท่ากับจำนวนของข้อมูลที่ได้จากชั้นอินพุตเพื่อให้ครอบคลุมข้อมูลอินพุตทั้งหมด

ชั้นเอาต์พุต สำหรับชั้นนี้จะเป็นชั้นแสดงผลของการพยากรณ์ซึ่งในโครงข่ายประสาทเทียมของงานวิจัยนี้มีคำตอบที่ออกมาเป็นค่าเดียวคือค่าพยากรณ์ ดังนั้นชั้นเอาต์พุตจึงมีเพียงโหนดเดียวเท่านั้น

4. การทดลองและผลการทดลอง

4.1 ผลที่ได้จากการเตรียมข้อมูล

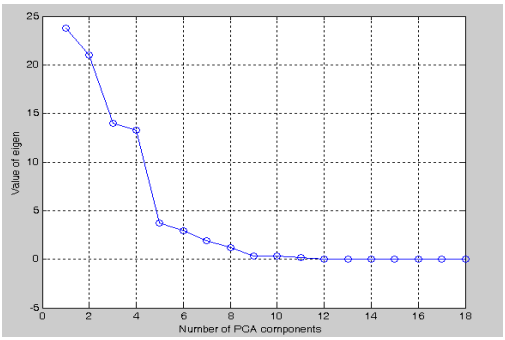
ปัจจัยทั้งหมดที่ได้รวบรวมและได้ตรวจสอบความถูกต้องและทำให้มีสภาวะแวดล้อมที่เหมือนกันได้ทั้งหมด 18 ปัจจัย [3] เช่น ดัชนีราคาหลักทรัพย์ที่เป็นมาตรฐาน (S&P500 index) ดัชนีราคาหุ้นกู้อายุ 30 ปี (30 year bond yields) ดัชนีราคาหุ้นกู้อายุ 10 ปี (10 year bond index) ดัชนีการทำเหมือง (Gold bugs index) ดัชนีราคาน้ำมัน (Oil index) อัตราแลกเปลี่ยนเงินบาทของไทย ราคาโลหะเงิน แพลทินัม ราคาทองคำสูงสุดของวันนี้

ราคาทองคำต่ำสุดของวันนี้ ราคาทองคำสูงสุดก่อนวันนี้ ราคาทองคำต่ำสุดก่อนวันนี้ ราคาทองคำสูงสุดของสองวันที่แล้ว ราคาทองคำต่ำสุดของสองวันที่แล้ว ราคาทองคำเปิดตลาดของวันนี้ ราคาทองคำปิดตลาดของวันนี้ ราคาทองคำเปิดตลาดของก่อนวันนี้ ราคาทองคำปิดตลาดของก่อนวันนี้

4.2 ผลที่ได้จากสกัดปัจจัยในแต่ละวิธี

4.2.1 วิธีการวิเคราะห์องค์ประกอบหลัก (PCA)

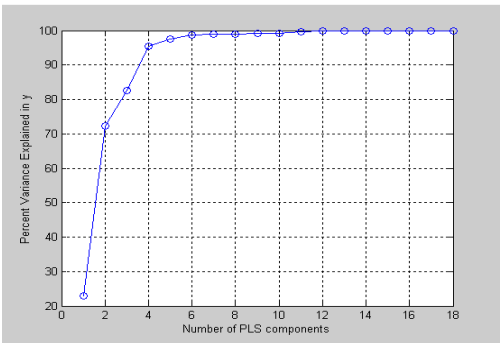
ผลลัพธ์ที่ได้จากการสกัดปัจจัยด้วย PCA จะได้ปัจจัยใหม่ซึ่งสามารถกำหนดจำนวนปัจจัยใหม่ได้จากค่าเจาะจง ซึ่งค่าเจาะจงแรกจะมากที่สุดและรองลงมาก็จะน้อยกว่าเรื่อยๆ จนถึงค่าที่เราสามารถยอมรับ ดังรูปที่ 2 ค่าเจาะจงลดลงมาถึงค่าเจาะจงที่ 9 ส่วนค่าเจาะจงที่ 10 ถึง 18 ไม่ทำให้ค่าเจาะจงเปลี่ยนแปลง เพราะฉะนั้นเราสามารถกำหนดจำนวนปัจจัยหลักได้ 9 ปัจจัย



รูปที่ 2 ค่าเวกเตอร์เจาะจง (eigenvalue) ของ PCA

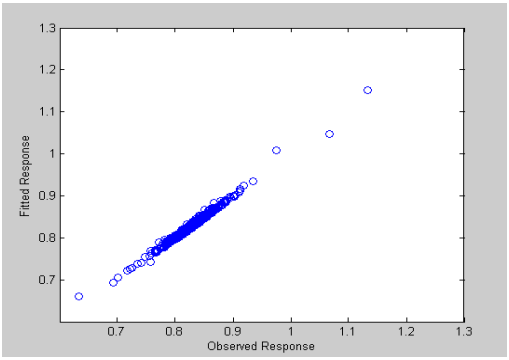
4.2.2 วิธี Partial Least square Regression

ผลลัพธ์ที่ได้จากการสกัดปัจจัยด้วยวิธี Partial Least Square Regression สามารถกำหนดจำนวนปัจจัยตัวใหม่ได้จากผลของการทดลอง ดังรูปที่ 3 ค่าปัจจัยหลักที่มีความสัมพันธ์กับตัวแปรเกณฑ์มากที่สุดนับจากปัจจัยที่ 1 ถึง ปัจจัยที่ 6 ส่วนปัจจัยที่ 7 ถึงปัจจัยที่18 ไม่ทำให้ค่าตัวแปรเกณฑ์เปลี่ยนแปลง เพราะฉะนั้นเราสามารถกำหนดจำนวนปัจจัยหลักได้ 6 ปัจจัย



รูปที่ 3 ค่าองค์ประกอบหลัก ของ PLSR ที่เหมาะสมในการพยากรณ์ตัวแปรเกณฑ์

ตัวแปรพยากรณ์หรือปัจจัยใหม่และตัวแปรเกณฑ์ทั้งสองมีความสัมพันธ์กันมากดังรูปที่ 4



รูปที่ 4 ความสัมพันธ์ระหว่างตัวแปรพยากรณ์ และตัวแปรเกณฑ์

4.2.3 วิธี PCA-Stepwise

จากการทดลองพบว่า การหาปัจจัยที่มีความสัมพันธ์กับราคาทองคำด้วยวิธีนี้ได้ 6 ปัจจัย โดยกำหนดได้จากค่า P-value [1] ของแต่ละปัจจัยซึ่งกำหนดเอาเฉพาะค่า P-value < 0.05 และ Status เป็น 'In' ดังรูปที่ 5

'Coeff'	'Std.Err.'	'Status'	'P'
[-1.0812e-004]	[1.8675e-005]	'In'	[1.0795e-008]
[-0.0010]	[1.0532e-004]	'In'	[0]
[8.6697e-004]	[1.0582e-004]	'In'	[1.2212e-015]
[-2.2454e-004]	[2.1844e-005]	'In'	[0]
[6.8950e-005]	[2.0451e-005]	'In'	[7.9028e-004]
[-4.7278e-006]	[9.7489e-006]	'Out'	[0.6279]
[1.3792e-004]	[1.6525e-005]	'In'	[4.4409e-016]
[-2.5917e-005]	[2.2274e-005]	'Out'	[0.2450]
[2.4320e-005]	[1.7846e-005]	'Out'	[0.1734]

รูปที่ 5 ค่าสัมประสิทธิ์ (coefficient) ความสัมพันธ์ของแต่ละปัจจัยและค่า P-value ในขั้นตอน PCA-Stepwise

4.2.4 วิธี Enter Selection

จากการทดลองพบว่า การหาปัจจัยที่มีความสัมพันธ์กับราคาทองคำด้วยวิธีนี้ได้ 5 ปัจจัย โดยการกำหนดได้จากค่า P-value [1] ของแต่ละปัจจัยซึ่งกำหนดเอาเฉพาะปัจจัยที่มีค่า P-value < 0.05 ดังรูปที่ 6

Coef	StdErr	tStat	pVal
753.38	546.81	1.3778	0.16874
-0.071639	0.33627	-0.21304	0.83136
-0.5127	0.32603	-1.5726	0.1163
0.2029	0.33404	0.60739	0.5438
1.1151	0.32906	3.3888	0.00074332
0.69702	0.33495	2.081	0.037817
-0.2873	0.33111	-0.86771	0.38587
0.049803	0.33534	0.14852	0.88198
-0.064845	0.33488	-0.19364	0.84652
-0.44641	0.070025	-6.375	3.422e-010
0.25116	0.069197	3.6296	0.00030563
-0.2567	0.16668	-1.54	0.12403
-43.089	94.094	-0.45794	0.64715
-47.323	106.19	-0.44566	0.65599
0.17316	0.40594	0.42656	0.66984
0.35065	0.13274	2.6416	0.0084455
0.070549	0.069378	1.0169	0.30958
-3.7832	11.754	-0.32187	0.74765
7.3583	11.384	0.64637	0.51826

รูปที่ 6 ค่าสัมประสิทธิ์ (coefficient) ความสัมพันธ์ของแต่ละปัจจัยและค่า P-value ในขั้นตอน Enter Selection

4.2.5 วิธี Stepwise Selection

จากการทดลองพบว่า การหาปัจจัยที่มีความสัมพันธ์กับราคาทองคำด้วยวิธีนี้ได้ 9 ปัจจัย โดยการกำหนดได้จากค่า P-value [1] ของแต่ละปัจจัยซึ่งกำหนดเอาเฉพาะปัจจัยที่มีค่า P-value < 0.05 และ Status เป็น 'In' ดังรูปที่ 7

'Coeff'	'Std.Err.'	'Status'	'P'
[0.1456]	[0.0785]	'In'	[0.0642]
[-0.3494]	[0.1896]	'Out'	[0.0658]
[-0.2044]	[0.1935]	'Out'	[0.2912]
[0.6115]	[0.0675]	'In'	[0]
[0.7498]	[0.0778]	'In'	[0]
[-0.2593]	[0.1901]	'Out'	[0.1731]
[-0.1604]	[0.1934]	'Out'	[0.4073]
[-0.3586]	[0.0685]	'In'	[2.1903e-007]
[-0.4505]	[0.0687]	'In'	[1.0687e-010]
[0.2479]	[0.0680]	'In'	[2.8819e-004]
[-0.2680]	[0.1206]	'In'	[0.0266]
[-53.2199]	[20.0997]	'In'	[0.0083]
[-0.1088]	[78.9543]	'Out'	[0.9989]
[0.3236]	[0.2673]	'Out'	[0.2265]
[0.3528]	[0.1032]	'In'	[6.6853e-004]
[0.0427]	[0.0418]	'Out'	[0.3076]
[2.9824]	[6.6272]	'Out'	[0.6528]
[11.1386]	[8.8936]	'Out'	[0.2108]

รูปที่ 7 ค่าสัมประสิทธิ์ (coefficient) ความสัมพันธ์ของแต่ละปัจจัยและค่า P-value ในขั้นตอน Stepwise Selection

4.3 เปรียบเทียบผลการสกัดปัจจัย

จากการทดลองวัดผลของการสกัดปัจจัยแต่ละวิธีสามารถนำมาเปรียบเทียบจำนวนปัจจัยตัวใหม่ได้ตามตารางที่ 2

การสกัดปัจจัยด้วยวิธี	จำนวนปัจจัยตัวใหม่ที่ได้
PCA	9
PLSR	6
PCA-Stepwise	6
Enter Selection	5
Stepwise Selection	9

ตารางที่ 2 เปรียบเทียบจำนวนปัจจัยตัวใหม่และเวลา ในการทำงานของทั้งสามวิธี

4.4 เปรียบเทียบผลการพยากรณ์โดยโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับและแบบฟังก์ชันรัศมีฐาน

ในงานวิจัยนี้ได้ทำการวัดประสิทธิ์ภาพของการพยากรณ์ด้วยค่าร้อยละของความคลาดเคลื่อนสมบูรณ์ (MAPE)

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\left(\frac{x_{(t)} - y_{(t)}}{x_{(t)}}\right) * 100\right| \qquad (15)$$

โดยที่ $x_{(t)}$ คือ ค่าข้อมูลจริง ณ เวลา t

$y_{(t)}$ คือ ค่าพยากรณ์ ณ เวลา t

N คือ จำนวนข้อมูลทั้งหมด

ปัจจัยที่ได้จากข้อที่ 4.3 มาทำการแบ่งเป็นสองส่วนคือ

ส่วนที่1 เป็นข้อมูลที่ใช้ในการเรียนรู้ของตัวแบบโดยใช้วิธีการป้อนชุดข้อมูลทดสอบและชุดเป้าหมายเข้าในเครือข่ายเพื่อให้โครงข่ายประสาทเทียมมีการเรียนรู้และปรับค่าน้ำหนักเพื่อให้เกิดความผิดพลาดน้อยที่สุด

ส่วนที่ 2 ข้อมูลในช่วงเวลาเดียวกันสำหรับใช้ในการทดสอบโดยเฉพาะ

สำหรับพารามิเตอร์ที่ใช้ในการพยากรณ์จะกำหนดให้มีค่าเท่ากันทุกตัวแบบโดยการทดลองปรับเปลี่ยนค่าสัมประสิทธิ์ให้กับตัวแบบเพื่อหาค่าพารามิเตอร์ที่ให้ค่าการพยากรณ์สูงสุดของโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ พบว่าค่าที่เหมาะสมสามารถแสดงดังตารางที่ 3

ค่าโมเมนต้มในชั้นซ่อนและชั้นแสดงผล	0.9
ค่าอัตราการเรียนรู้ในชั้นซ่อนและชั้นแสดงผล	0.3
ฟังก์ชันกระตุ้นในชั้นซ่อน	Hyperbolic Tangent
ฟังก์ชันกระตุ้นในชั้นผลลัพธ์	ผลรวมเชิงเส้น
จำนวนรอบของการเรียนรู้	3000

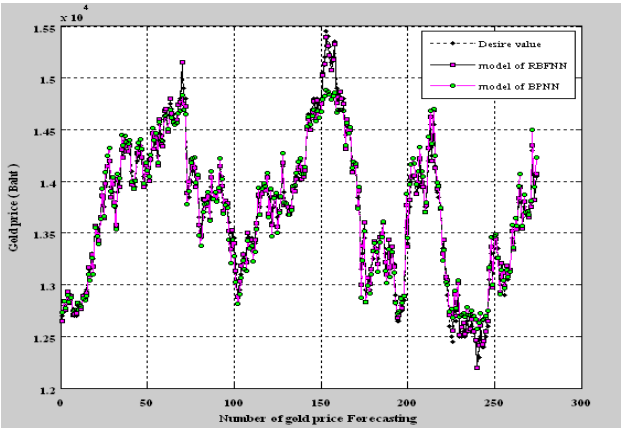
ตารางที่ 3 ค่าพารามิเตอร์ซึ่งกำหนดให้เท่ากันทุกตัวแบบของโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ

ผลการทดลองหาค่าพารามิเตอร์ให้ RBFNN สามารถกำหนดพารามิเตอร์การกระจาย (Spread) เท่ากับ 33.2 จำนวนจุดศูนย์กลางค่งที่ฟังก์ชันการกระตุ้นในชั้นซ่อนคือฟังก์เกาส์เซียน เปรียบเทียบผลที่ได้จากการพยากรณ์ของทั้งสองโครงข่ายประสาทเทียมดังตารางที่ 4

โครงข่ายประสาทเทียม	วิธีการสกัดปัจจัย				
	PCA	PCA-Stepwise	PLSR	Enter Selection	Stepwise Selection
BPNN	1.285	1.277	0.686	1.115	1.052
RBFNN	1.259	1.230	***0.384***	1.049	1.065

ตารางที่ 4 เปรียบเทียบประสิทธิภาพ MAPE ระหว่างสองโครงข่าย BPNN และ RBFNN โดยใช้ปัจจัยที่ได้จากข้อที่ 4.3

ผลการทดลองพบว่าปัจจัยที่ผ่านการสกัดด้วยวิธี PLSR จะให้ผลการพยากรณ์ด้วยตัวแบบ BPNN เป็น 6:2:1 และแบบ RBFNN เป็น 6:6:1 ที่มีอัตราส่วนข้อมูลฝึกสอนต่อทดสอบเป็น 60 : 40 ให้ค่า MAPE เท่ากับ 0.686% และ 0.384% ตามลำดับ และสามารถพยากรณ์ได้ 274 วัน ดังนั้นจึงนำตัวแบบนี้ไปพยากรณ์ราคาทองคำได้โดยให้ความแม่นยำสูง



รูปที่ 8 การเปรียบเทียบราคาทองคำที่ได้จากการพยากรณ์กับราคาทองคำจริง

5. สรุปผลการทดลอง

งานวิจัยนี้ได้นำเสนอการเปรียบเทียบวิธีการสกัดปัจจัยเพื่อให้การพยากรณ์ที่มีประสิทธิ์ภาพและความแม่นยำสูง โดยการรวบรวมปัจจัยทั้งหมดรวมทั้งราคาทองคำได้ 18 ปัจจัย นำมาใช้เป็นข้อมูลนำเข้าวิธีการสกัดปัจจัย 5 วิธีการ จากนั้นนำผลที่ได้จากการสกัดปัจจัยแต่ละวิธีมาทำการเปรียบเทียบกัน จากการทดลองพบว่าการสกัดปัจจัยด้วยวิธี PLSR ได้จำนวนปัจจัยตัวใหม่เท่ากับวิธี PCA-Stepwise และมีจำนวนน้อยกว่าการสกัดด้วยวิธี PCA และ วิธี Stepwise Selection วิธี Enter Selection เมื่อนำปัจจัยที่ได้จากแต่ละวิธีไปพยากรณ์เพื่อวัดประสิทธิภาพของวิธีการสกัดปัจจัยด้วย BPNN และ RBFNN พบว่าปัจจัยที่ได้จากวิธี PLSR จะให้ประสิทธิ์ภาพความแม่นยำในการพยากรณ์สูงกว่า ทุกวิธีที่กล่าวมา ด้วย RBFNN ตัวแบบ 6 : 6 : 1 โดยให้ค่า MAPE เท่ากับ 0.384%

จากผลการทดลองที่ได้ศึกษา ผู้วิจัยได้วิธีการในการเลือกปัจจัยสำหรับการพยากรณ์ราคาทองคำ ในการทำวิจัยในลำดับถัดไป คณะผู้วิจัยจะศึกษาตัวแบบการพยากรณ์ที่เหมาะสมสำหรับการพยากรณ์ระยะสั้นและ ระยะยาว ตลอดถึงการประยุกต์ใช้วิธีการที่ได้จากงานวิจัยนี้กับข้อมูลชนิดอื่นๆ ต่อไป

เอกสารอ้างอิง

[1] กัลยา วานิชย์บัญชา. *การวิเคราะห์สถิติ:สถิติสำหรับบริหารและวิจัย* พิมพ์ครั้งที่ 9 จุฬาลงกรณ์มหาวิทยาลัย. กรุงเทพฯ, pp.354-355, 2550

[2] สมาคมค้าทองคำแห่งประเทศไทย. ราคาทองคำ [ออนไลน์] [อ้างเมื่อ 20 สิงหาคม 2551]. จาก http://www.goldtraders.or.th/price.php.

[3] ปรีดา อนุสรณ์ธีรกุล และ พุธษดี ศิริแสงตระกูล. “การเพิ่มประสิทธิ์ภาพการพยากรณ์ราคาทอง คำโดยวิธีพยากรณ์รวมด้วยวิธีโครงข่ายประสาทเทียม”, *Proceedings of The Conference on Knowledge and Smart Technology*, Burapha University, pp.1-6, 2552

[4] ปิยะวิทย์ หมดมลทิน, “ปัจจัยที่มีอิทธิพลต่อราคาทองรูปพรรณ” วิทยานิพนธ์ คณะเศรษฐศาสตร หลักสูตรปริญญาเศรษฐศาสตรมหาบัณฑิต มหาวิทยาลัยรามคำแหง, 2549

[5] วิภาดา เวทย์ประสิทธิ์ และ พรพิมล ณ นคร, “การสกัดปัจจัยข้อมูลสำหรับการพยากรณ์ฝนโดยใช้โครงข่ายประสาทเทียม”, *Proceedings of The Joint Conference on Computer Science and Software Engineering*, pp.151-156, 2005

[6] Alexandra Debiolles, Latifa Oukhellou and Patrice Aknin. “Combined Used of Partial Least Square Regression and Neural Network for Diagnosis Tasks”, *Proceedings of the 17th International Conference on Pattern Recognition*, pp.1051-4651, 2004

[7] Simon Haykin, *Neural Network: A Comprehensive Foundation*, (2nd ed), Prentice Hall, 1998

[8] Ying Sai ,Shiwei Zhu และ Tao Zhang. “Predicting financial distress of Chinese listed corporate by a hybrid PCA-RBFNN”, *Fourth International Conference on Natural Computation*, pp.277-281, 2008

[9] Z. Ismail, A. Yahya and A. Shabri. “Forecasting Gold Price Using Multiple Linear Regression Method”, *American Journal of Applied Sciences*, 6(8): pp. 1509-1514, 2009