

การใช้โปรแกรม smart PLS สำหรับการวิจัยด้วยตัวแบบสมการ โครงสร้าง

บรรยายในงาน มหกรรมวิจัยแห่งชาติ 2562 (Thailand Research Expo 2019)
10 เมษายน 2562 ณ โรงแรมเซ็นทาราแกรนด์ เซ็นทรัลเวิลด์ กรุงเทพมหานคร

โดย

รองศาสตราจารย์ ดร.มนตรี พิริยะกุล

ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยรามคำแหง

Email: mpiriyakul@yahoo.com line ID: mpiriyakul

PLS คืออะไร

PLS (partial least square) เป็นซอฟต์แวร์ที่รันชุดของสมการถดถอยวนเวียนจนได้รับคำตอบ (convergence) ด้วยวิธี least square พัฒนาอัลกอริทึมโดย Herman Wold (1982) โปรแกรม Smart PLS พัฒนาโดย Christian M. Ringle, Sven Wende, and Alexander Will เวอร์ชัน 2.0 ในปี 2005

Smart PLS คืออะไร

โปรแกรม smart PLS 3 เป็นหนึ่งในหลายโปรแกรมในกลุ่ม variance-based SEM ดาวน์โหลดมาทดลองใช้ได้ฟรีนาน 1 เดือนจาก <https://www.smartpls.com/downloads> แต่ smart PLS 2 สามารถดาวน์โหลดมาใช้นาน 3 เดือนและจะต่อไลเซนส์ให้ฟรี

หมายเหตุ Variance-based คือสถานการณ์ที่เราหาค่าประมาณการถดถอยจากการ minimize $\sum_i^n e_i^2$

$$e_i^2 = (Y_i - \hat{Y}_i)^2 \quad \text{โดยที่} \quad \hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \dots + \hat{\beta}_1 X_1 + u_i$$

ทั้งนี้ $\sum_i^n e_i^2$ คือ sum square error ถ้าหารด้วย degree of freedom คือ $n-(k-1)$ ก็คือ variance เรียกว่า error variance

โปรแกรม PLS บางส่วน

PLS-Graph (Chin, 2003),
Visual PLS (Fu, 2006),
SPAD-PLS (Test & Go, 2006),
XLSTAT (Henseler & Dijkstra, 2015),
PLS-GUI (Hubona, 2015),
Warp PLS (Kock, 2015),
Smart PLS (Ringle et al., 2015),
semPLS (Monecke & Heisch, 2012),
plsmpm (Sanchez, Trinchera and Russolillo, 2015)
LVPLS (Lohmöller, 1984),
PLSPath (Sellin, 1989)

PLS algorithm

PLS จะการวนเวียนคำนวณด้วยวิธี least square estimation ซ้ำจนกว่าจะได้รับความตอบ โดยกระทำกับ สมการถดถอยที่มีจำนวนมากเท่ากับจำนวนสมการ โครงสร้าง (structural model, inner model) และ จำนวนสมการมาตรวัด (measurement model, outer model) รวมกัน

- 1) ประมาณค่าสัมประสิทธิ์ถดถอยสมการมาตรวัด โดยรันทีละบล็อก
- 2) สร้างค่าของตัวแปรแฝงของบล็อก ค่าของ LV ถูกสร้างขึ้นจาก MV ในบล็อกของตนและให้ นอร์มอลไลซ์ให้มี $SD = 1$
- 3) ประมาณค่าสัมประสิทธิ์เส้นทางของสมการ โครงสร้าง
- 4) พยากรณ์ค่าตัวแปรแฝงขึ้นมาใหม่จากสมการ โครงสร้าง
- 5) ประมาณค่าสัมประสิทธิ์ถดถอยสมการมาตรวัดใหม่ระหว่าง LV ค่าใหม่กับ MV เดิม
- 6) ทำข้อ 2) ถึง 5)

Formative-Reflective indicator

MV (manifest variable) หรือ indicator มีได้ 2 แบบคือ formative indicator กับ reflective indicator

1. **formative indicator** ตัวชี้วัดจะเป็นตัวแทนจากทุกส่วน (domain/dimension/reality/parcel) ของตัวแปรแฝงนั้น ข้อถามต้องมาจากต่างส่วนกันของตัวแปรแฝง จึงไม่สัมพันธ์กันและตัดทิ้งไม่ได้เพราะเท่ากับกับตัด โดเมนทิ้งไป ข้อที่เหลือไม่อาจให้สารสนเทศแทนกันได้

Formative-Reflective indicator

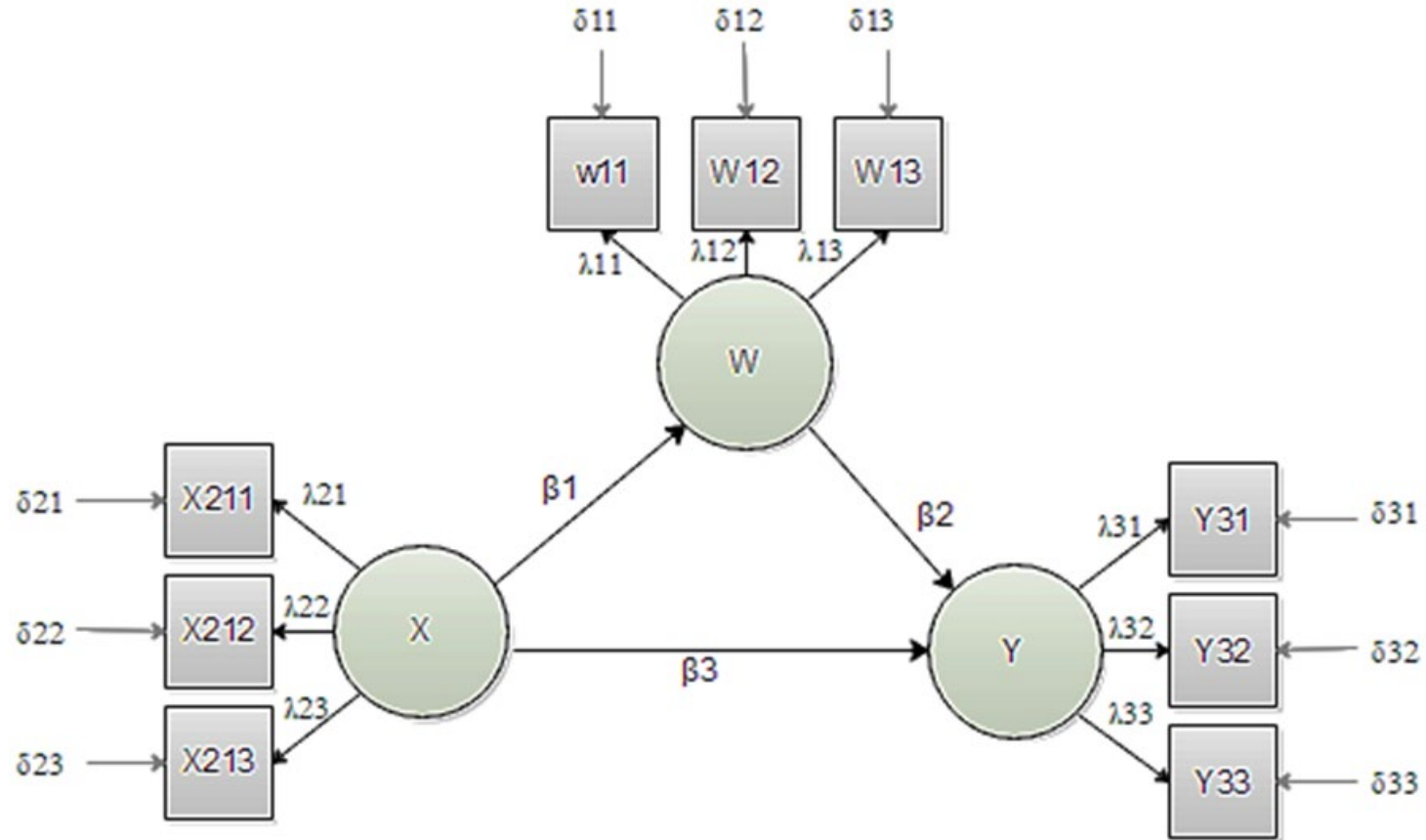
2. **reflective indicator** ตัวชี้วัดจะเป็นตัวแทนของตัวแปรแฝง เป็นตัวอย่างส่วนหนึ่งของตัวชี้วัดของแปรแฝงจึงมีความสัมพันธ์กันสูง (co-vary) เพราะวัดเรื่องเดียวกันและสามารถทดแทนกัน (interchangeability) ได้ คือตัดทิ้งไปบางข้อได้ โดยข้อที่เหลืออยู่ซึ่งสัมพันธ์กันกับข้อที่ถูกตัดทิ้งไป (ที่จริงคือสัมพันธ์กับตัวแปรแฝง) ก็สามารถให้สารสนเทศแทนข้อที่ตัดทิ้งไปได้ เรากำหนดว่าตัวชี้วัดแบบนี้มีความสัมพันธ์ทางบวกกับตัวแปรแฝง ค่า loading จึงต้องเป็นปริมาณบวก

Formative-Reflective indicator

ตัวอย่างเช่น มาตรการวัดความพึงพอใจในโรงแรมที่พัก กรณี **reflective indicator** จะวัดรวมๆ ทั่วไปหมด เช่น ความชอบ ความรู้สึกสะดวกสบาย ความตั้งใจบอกต่อ แต่ในกรณี **formative indicator** จะวัดแยกเป็น โดเมน เช่นด้านพนักงาน ด้านห้องพัก ด้านความสงบ ด้านสิ่งบริการเสริมเช่นฟิตเนส

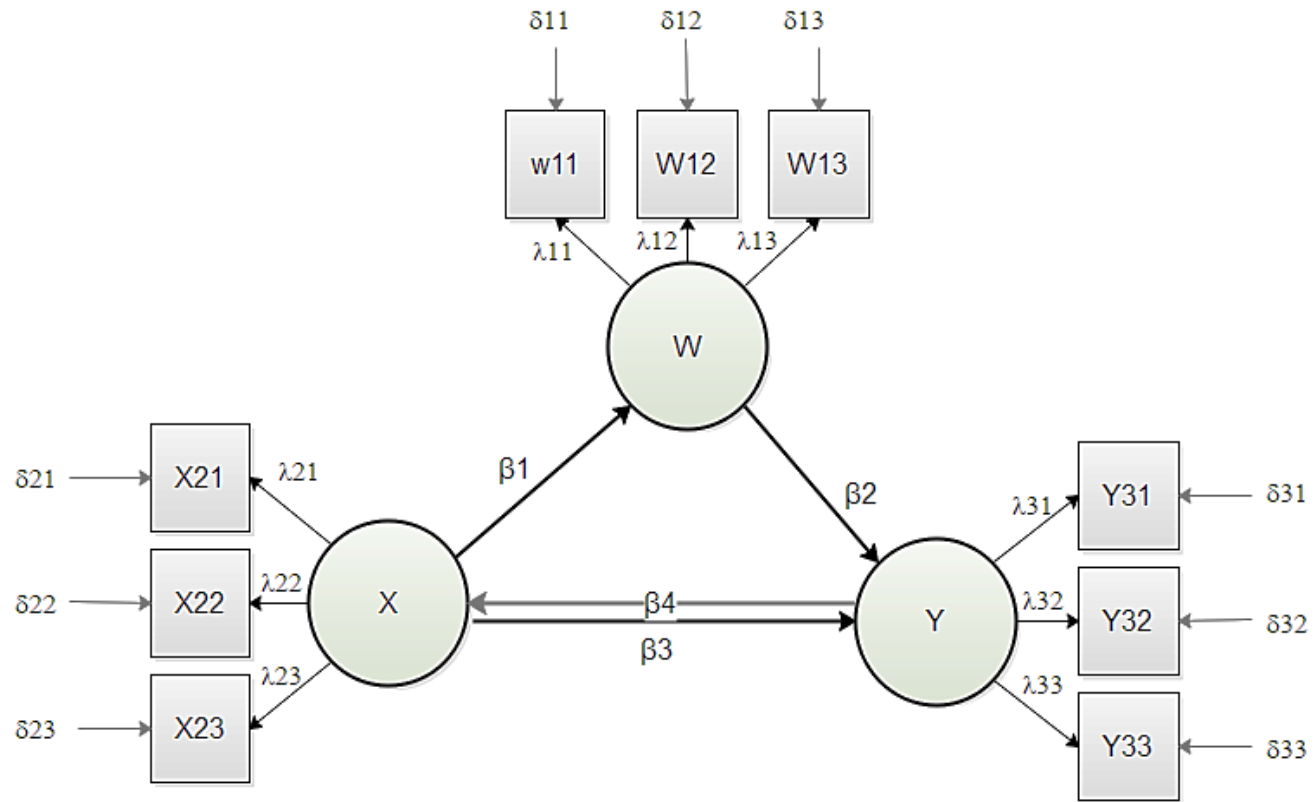
Research Design

1. Causality ตัวแปรแฝงโยงถึงกันเป็นเครือข่ายโดยแต่ละเส้นทางมีทฤษฎีกำกับ



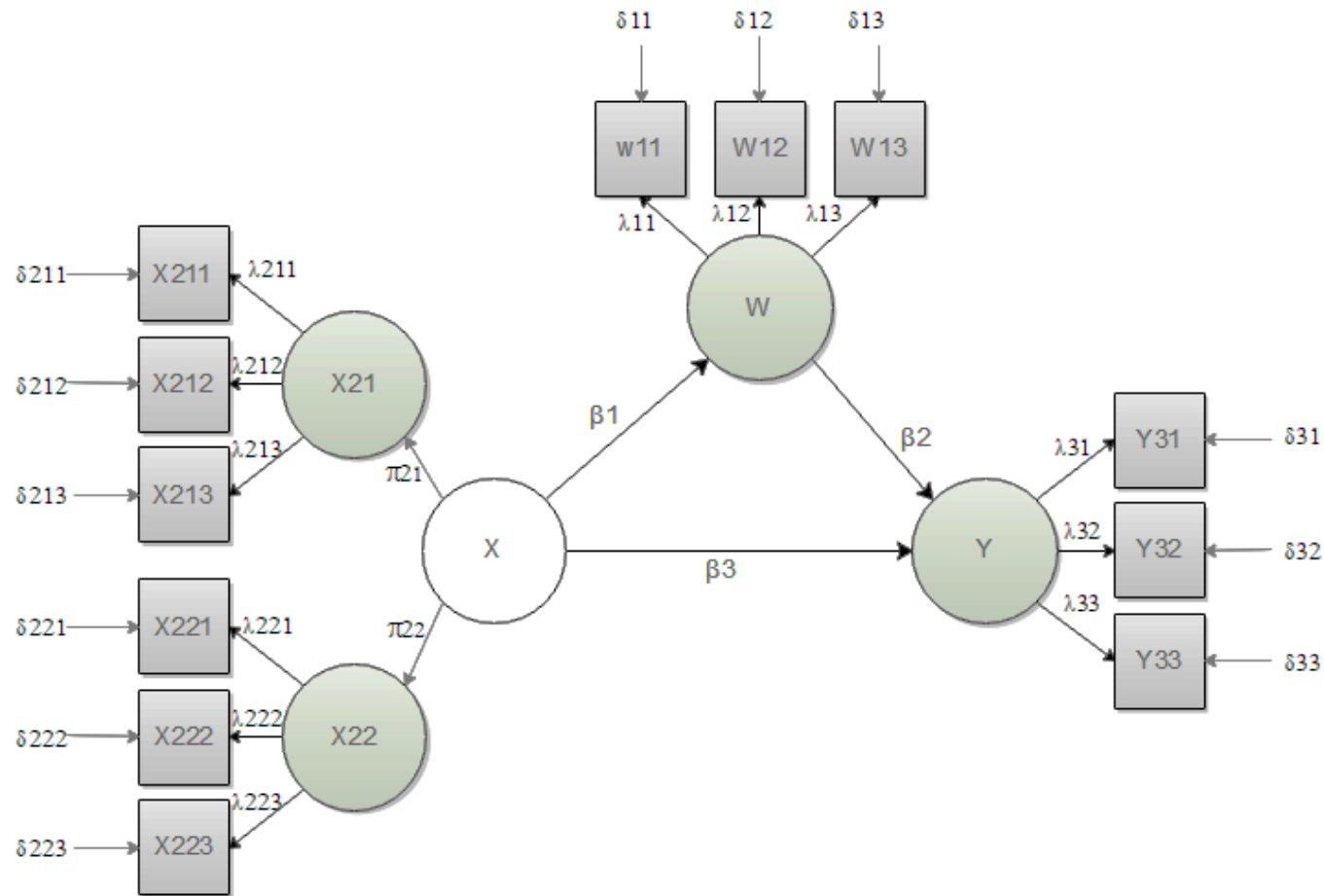
Research Design

2. non-recursive model/reciprocal model ตัวแปรแฝงโยงถึงกันเป็นเครือข่ายโดยที่แต่ละเส้นทางมีทฤษฎีกำกับกับ บางเส้นทางเป็นความสัมพันธ์ย้อนกลับ



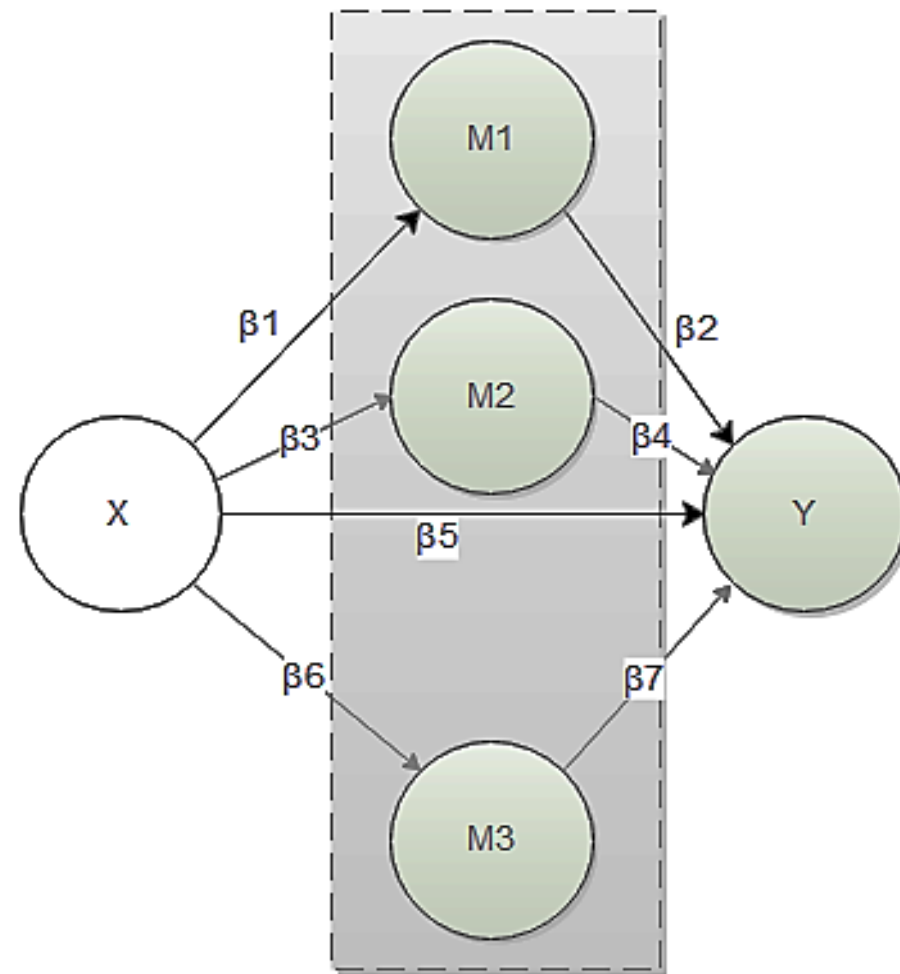
Research Design

3. **Second order model** ตัวแปรแฝง โยงถึงกันเป็นเครือข่าย โดยแต่ละเส้นทางมีทฤษฎีกำกับกับ ตัวแปรแฝงบางตัวหรือทุกตัวมี sub-construct ซึ่งก็คือ sub-domain ของเนื้อหา



Research Design

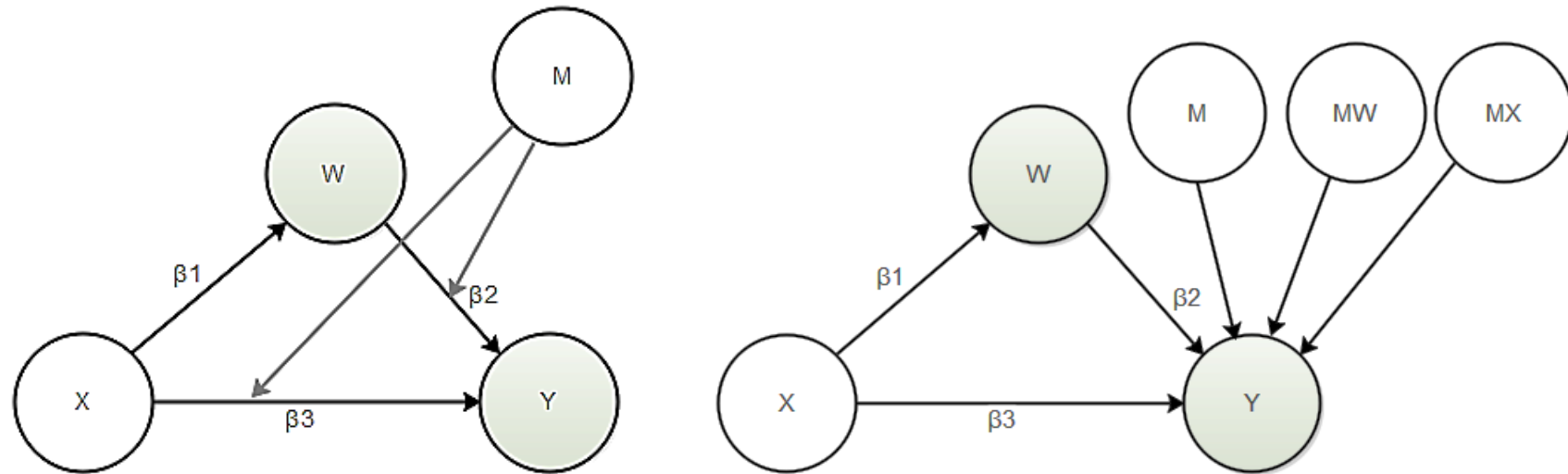
4. **Mediation model** ตัวแบบมีตัวแปรคั่นกลาง ตัวแปรคั่นกลางเป็นตัวแปรที่ซ่อนเร้นเชื่อมโยงระหว่างตัวแปรต้นทางกับตัวแปรปลายทาง (hidden transmitting variables) ซึ่งนักวิจัยขุดค้นนำมาแสดงให้ประจักษ์เป็นตัวแปรที่มีอิทธิพลต่อตัวแปรปลายทางแต่อาจไม่เกี่ยวข้องข้องกับตัวแปรต้นทาง ถ้ามีหลายตัวเรียกว่า multiple mediation model อาจเป็น parallel mediation หรือ serial mediation ขึ้นอยู่กับทฤษฎีหรือ sequential exploratory analysis



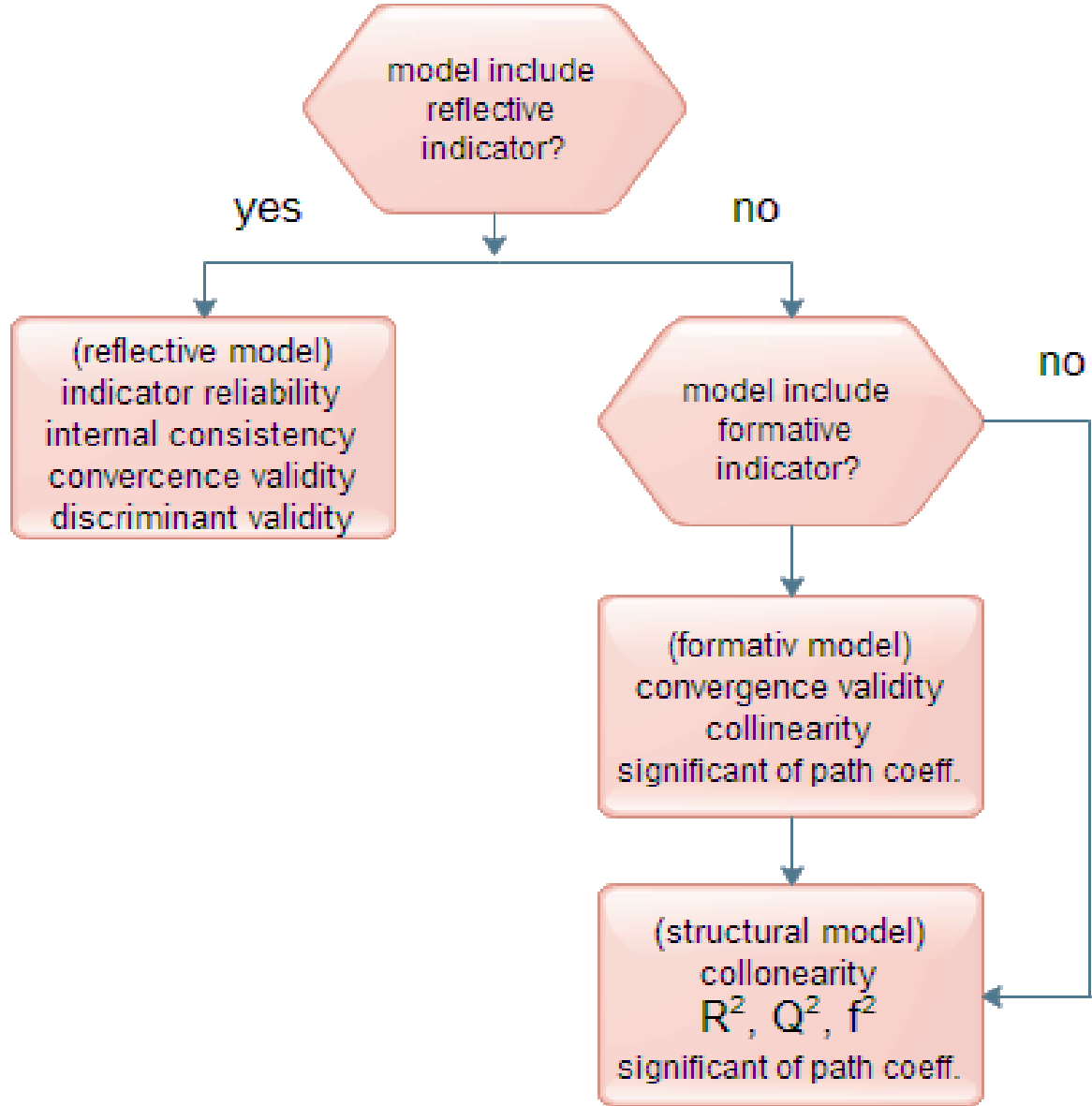
Research Design

5. **Moderation model** ตัวแบบมีตัวแปรกำกับ ตัวแปรกำกับเป็นตัวแปรที่ซ่อนเร้นควบคุมความสัมพันธ์ระหว่างตัวแปรต้นทางกับตัวแปรปลายทาง คือมีปฏิสัมพันธ์กับตัวแปรต้นทางแล้วส่งผลต่อตัวแปรปลายทางซึ่งนักวิจัยชุดคั่นนำมาแสดงให้เห็นประจักษ์ ตัวแปรกำกับมักขึ้นอยู่กับ sequential exploratory analysis มีได้หลายลักษณะตามประสบการณ์ของผู้เชี่ยวชาญ

ถ้าเป็นตัวแปรนามบัญญัติให้วิเคราะห์แบบพหุกลุ่ม ถ้าเป็นตัวแปรปริมาณให้วิเคราะห์ตามปกติแต่ต้องจัดรูปกรอบการวิจัยเป็น statistical model เสียก่อนและต้องวิเคราะห์ conditional effect ด้วย



PLS running process



การรัน smart PLS 3

1. การเตรียมข้อมูล อาจเตรียมผ่าน SPSS หรือ Excel ไฟล์ใช้งานจะเป็น text file จึงให้เซฟเป็น filename.csv หรือ filename.raw
2. ต้องไม่มี missing data หากมีให้แก้ปัญหาลงให้เสร็จก่อน

เกณฑ์คุณภาพ

Cronbach's alpha และ Composite reliability

1. Cronbach's alpha

Cronbach's alpha ≥ 0.60 for exploratory purpose

Cronbach's alpha ≥ 0.70 acceptable scale

Cronbach's alpha ≥ 0.80 good scale

2. Composite reliability (CR)

CR ≥ 0.60 for exploratory purpose

CR ≥ 0.70 acceptable confirmatory purpose

CR ≥ 0.80 good confirmatory purpose

CR ≥ 0.90 check for minor wording invariant

Average Variance Extract (AVE)

3. Average variance extracted (AVE) = $\frac{1}{m_i} \sum_j^{m_i} \text{loading}_j^2$ คือค่าเฉลี่ยความผันแปรในกลุ่มตัวชี้วัดที่ LV สามารถอธิบายตัวชี้วัดได้

AVE ใช้แสดง LV reliability

เกณฑ์คือ $AVE_i \geq 0.50$ และต้องมากกว่า cross loading

$AVE_i = 0.50$ แสดงว่า LV อธิบายความผันแปรตัวชี้วัดได้โดยเฉลี่ยร้อยละ 50

(Note: Reliability is a capability to measure the right concept. Reliability is the accuracy in measurements when the measurements are repeated)

Loading

4. **Loading** คืออิทธิพลที่ LV ถ่ายทอดไปยังตัวชี้วัด

1) ต้องเป็นปริมาณบวก เพราะ SEM ถือว่า MV ต้องมีความสัมพันธ์ทางบวกกับ LV

2) ต้องมีค่าขั้นต่ำ 0.707 และต้องมีนัยสำคัญ (Henseler, Ringle and Sarstedt, 2012)

อาจต่ำกว่า 0.707 ได้แต่ต้องไม่ต่ำกว่า 0.50 และ AVE_i ต้องไม่ต่ำกว่า 0.50 (Hair et al., 2014)

ทั้งนี้ $loading_j^2$ ใช้แสดง indicator reliability เฉพาะข้อที่ j (Hair et al., 2014)

loading

5. **Indicator reliability** ให้พิจารณาทั้ง loading และ weigh ร่วมกัน

1) Loading ต้องมีค่าไม่ต่ำกว่า 0.707 และมีนัยสำคัญ (คือ $\text{loading}_j^2 \geq 0.50$)

2) ถ้า loading ต่ำกว่า 0.50 แต่มีนัยสำคัญ และ weight ไม่มีนัยสำคัญให้ทิ้งข้อถามนั้นถ้าเห็นว่าไม่ใช่ข้อที่มีความสำคัญ (theoretical importance)

3) ถ้า loading สูงและมีนัยสำคัญ แต่ weight ไม่มีนัยสำคัญให้คงข้อถามนั้นถ้าเห็นว่ามีมีความสำคัญ (theoretical importance)

หมายเหตุ ตัวอย่างของ theoretical importance เช่นมาตรวัดของความตั้งใจจะต้องมีเงื่อนไขอยู่ด้วย ถ้ามีบางข้อ ไม่มีเงื่อนไขอยู่ในเนื้อความก็ให้พิจารณาตัดทิ้ง

data cleaning

โดยปกติ เมื่อพบเหตุการณ์ตามข้อ 2) นักวิจัยต้องพิจารณาดำเนินการ data cleaning ก่อน เช่น

แก้ปัญหา outlier ตามวิธี Winsorizing หรือ
แทนที่ข้อมูลสูญหายด้วยวิธีที่เหมาะสมกว่าเดิม หรือ
พิจารณาตัดหน่วยวิเคราะห์ที่มีข้อมูลไม่น่าเชื่อถือทิ้งไป เช่นกรณีแบบสอบถาม
เป็นชนิดไลเกอร์ท 5 คำตอบ ถ้าข้อมูลรายการใด (case, record) มีค่าเฉลี่ยประมาณ 3
และ CV ต่ำมากให้ตัดหน่วยนั้นทิ้ง

Standardized Root Mean Square Residual

6. Standardized Root Mean Square Residual (SRMR) คือ mean square ของผลต่างของสหสัมพันธ์ระหว่าง observed correlation กับ model-implied correlation เกณฑ์ขั้นสูงคือ $SRMR \leq 0.08$ (Smart PLS กำหนดให้ใช้ $SRMR \leq 0.12$)

Cross –loading

7. **Cross –loading** ใช้เป็นทางเลือกของการใช้ \sqrt{AVE} ในการวิเคราะห์ความเที่ยงตรงเชิงจำแนก (discriminant validity)

เกณฑ์คือ cross-loading $\leq 0.30, 0.40$

ถ้ามากกว่านี้แสดงว่ามาตรวัดขาดความเที่ยงตรงเชิงจำแนก ทั้งนี้เพราะตัวแบบที่ดีนั้น ตัวชี้วัดจะต้องสัมพันธ์ (load) กับ LV ของตนเองมากกว่ากับ LV อื่น ค่า cross-loading จึงควรมีค่าต่ำ

Heterotrait-Monotrait Correlation Ratio (HTMT)

8. Heterotrait-Monotrait Correlation Ratio (HTMT) ใช้วัดความเที่ยงตรงเชิงจำแนก โดยเปรียบเทียบเป็นคู่ๆระหว่างคู่ตัวแปรแฝง ค่านี้ต้องน้อยกว่า 1 ถ้ามามากกว่า 1 แสดงว่าขาดความเที่ยงตรงเชิงจำแนก

$$\text{HTMT}(ij) \leq 1.00, 0.90, 0.85; i \neq j$$

$$\text{HTMT}_{ij} = \underbrace{\frac{1}{K_i K_j} \sum_{g=1}^{K_i} \sum_{h=1}^{K_j} r_{ig,jh}}_{\text{average heterotrait-heteromethod}} \div \underbrace{\left(\frac{2}{K_i(K_i-1)} \cdot \sum_{g=1}^{K_i-1} \sum_{h=g+1}^{K_i} r_{ig,ih} \cdot \frac{2}{K_j(K_j-1)} \cdot \sum_{g=1}^{K_j-1} \sum_{h=g+1}^{K_j} r_{jg,jh} \right)^{\frac{1}{2}}}_{\text{geometric mean of the average monotrait-heteromethod correlation of construct } \xi_i \text{ and the average monotrait-heteromethod correlation of construct } \xi_j}$$

Fornell-Larcker Criterion

9. Fornell-Larcker Criterion

$$\text{Average variance extracted (AVE)} = \frac{1}{m_i} \sum_j^{m_i} \text{loading}_j^2$$

loading_j^2 คือค่าเฉลี่ยของ indicator reliability ใช้เป็น LV reliability

ถ้าให้ r_{ij} คือสหสัมพันธ์ของคะแนนปัจจัยระหว่างตัวแปรแฝงที่ i กับตัวแปรแฝงที่ j
ดังนั้น r_{ij}^2 คือสัดส่วนความผันแปรที่ตัวแปรแฝงทั้งสองต่างอธิบายกันและกัน

การแปลผล Fornell-Larcker Criterion

1. $\sqrt{AVE_i} > r_{ij}$; all $i \neq j$ แสดงว่าความสัมพันธ์ระหว่างตัวชี้วัดภายในกลุ่มเดียวกันสูงกว่าความสัมพันธ์ระหว่างคะแนนปัจจัยของตัวแปรแฝงที่สนใจกับตัวแปรแฝงอื่น แสดงว่ามีความเที่ยงตรงเชิงจำแนก หรือ
2. $AVE_i > r_{ij}^2$ แสดงถึง indicator reliability (คือค่าถัวเฉลี่ย $loading_j^2$ ภายในกลุ่มตัวแปรแฝงที่ i คือ AVE_i) สูงกว่าสัดส่วนความผันแปรที่ตัวแปรแฝงที่ i กับตัวแปรแฝงที่ j ต่างควบคุมความผันแปรของกันและกัน แสดงว่ามีความเที่ยงตรงเชิงจำแนก

multicollinearity

10. ภาวะร่วมเส้นตรงพหุ (multicollinearity) ใช้แสดงให้เห็นว่าตัวแบบมีตัวแปรสาเหตุที่สัมพันธ์กันสูงเกินไป ถ้าปรับแก้แล้วจึงจะเป็น well-fitting model เกณฑ์คือ

$$VIF_j = \frac{1}{1-R_j^2} < 4 \text{ หรือ } 5$$

เกณฑ์นี้กำหนดค่อนข้างเคร่งครัดเพราะกำหนดให้ $R_j^2 = 0.75, 0.80$ (สามารถผ่อนผันลงได้เป็น 0.90) โดยวัดกับตัวแปรผลลัพธ์ทุกตัวในตัวแบบ

multicollinearity

Multicollinearity คือความสัมพันธ์เชิงเส้นระหว่างปัจจัยสาเหตุ หากเกิดปัญหานี้ขึ้นแสดงว่าปัจจัยสาเหตุมีความสัมพันธ์กันสูงเกินระดับยินยอม

ผลกระทบ เราไม่อาจใช้ปัจจัยสาเหตุนั้นเป็นสื่ออธิบายปัจจัยผลลัพธ์ได้อย่างถูกต้อง และไม่อาจแน่ใจได้ว่าตัวแปรสาเหตุคู่ที่สัมพันธ์กันสูงนั้นตัวใดกันแน่ที่มีอิทธิพลต่อตัวแปรผลลัพธ์ หรือหรือตัวชี้วัดคู่ที่สัมพันธ์กันสูงนั้นตัวใดกันแน่ที่มีอิทธิพลต่อตัวแปรแฝง (formative measurement model) หรือเป็นผลสะท้อนของตัวแปรแฝง (reflective measurement model)

อาการที่แสดงออก คือ t-statistics มีค่าต่ำ (เพราะ SE มีค่าสูงมากจึงทำให้ $t_j = \frac{\beta_j}{se_{\beta_j}}$ มีค่าต่ำ) ทำให้สัมประสิทธิ์เส้นทางหลายเส้นทางไม่มีนัยสำคัญ หรือเครื่องหมายผิด

วิธีแก้ปัญหา multicollinearity

วิธีแก้ปัญหา multicollinearity คือ

- 1) รวมตัวแปรแฝงที่สัมพันธ์กันสูงเป็นกลุ่มเดียวกัน อาจรวมในชื่อใหม่และจัดให้ตัวแปรแฝงเดิมแต่ละตัวทำหน้าที่เป็น parcel/sub-construct/sub-domain คือแก้ไขให้การวิจัยแบบ second order model
- 2) ตัดตัวแปรแฝงที่ซ้ำซ้อนกันทิ้งไป
- 3) ตามไปดู multicollinearity ของตัวชี้วัดแล้วหาทางตัดตัวชี้วัดที่สัมพันธ์กันสูงเกินไปทิ้ง (item/indicator/MV)
- 4) ตัดรายการข้อมูล (case/record)
- 5) ถ้าเป็น path model อาจใช้วิธี mean-centered

Residual variance

12. **Residual variance** ต้องมีค่าต่ำและไม่มีนัยสำคัญ ($H_0: \sigma_e^2 = 0$ vs $H_1: \sigma_e^2 \neq 0$) ถ้ามีค่าสูงและมีนัยสำคัญแสดงว่าตกหล่นตัวแปรสำคัญ ให้ตรวจสอบวรรณกรรม

Goodness of fit ของ PLS-PM

Goodness of fit ของ PLS-PM วัดแบบเดียวกับในสมการถดถอย ดังนี้

1. R^2 เรียกว่า overall effect size เกณฑ์คือ $R^2 = 0.19, 0.33, 0.67$ แสดงว่าปัจจัยสาเหตุมีอิทธิพลร่วมกันต่อปัจจัยผลลัพธ์ต่ำ ปานกลาง และสูงตามลำดับ (Chin, 1998; Hock and Ringle, 2006)
2. Effect size (f^2) ใช้แสดงอิทธิพลของตัวแปรสาเหตุที่มีต่อตัวแปรผลลัพธ์เดียวกัน ค่านี้สามารถใช้ร่วมกับค่าสัมประสิทธิ์เส้นทาง (โดยปกติจะเสนอเป็นค่าปรับมาตรฐาน) ตัวแปรใดมีค่าสัมประสิทธิ์เส้นทางสูงกว่าแสดงว่ามีอิทธิพลต่อตัวแปรผลลัพธ์มากกว่า

Effect size

Effect size ของปัจจัยใดมีค่า f_i^2 สูงกว่าแสดงว่าปัจจัยนั้นมีอิทธิพลมากกว่า เกณฑ์ระดับอิทธิพลคือ $f_i^2 = 0.02, 0.15, 0.35$ แสดงว่ามีอิทธิพลน้อย ปานกลาง และสูง ตามลำดับ (Cohen, 1988)

ข้อควรระวัง การตัดตัวแปรกรณี multicollinearity โดยเฉพาะตัวแปรที่มีค่า f_i^2 สูงจะก่อให้เกิดความเสียหายมาก

ตัวอย่างการคำนวณหาค่า f_i^2

จากสมการ $Y = f(X_1, X_2) + u$ สามารถคำนวณหาค่า f_i^2 ตามสูตร $f_i^2 = \frac{R_{\text{original}}^2 - R_{\text{omitted}}^2}{1 - R_{\text{original}}^2}$ ได้ดังนี้

antecedent	R^2	ΔR_i^2	f_i^2
X1, X2	0.4308	-	-
X1 (X2 omitted)	0.3828	$0.4328 - 0.3828 = 0.0480$	$\frac{0.0480}{0.5692} = 0.0843$
X2 (X1 omitted)	0.2186	$0.4328 - 0.2186 = 0.2122$	$\frac{0.2122}{0.5692} = 0.3728$
$1 - R_{\text{original}}^2 = 1 - 0.4308 = 0.5692$			

residual

3. Standardized residual ถ้าพบว่ามีค่าเกินขอบเขต ± 1.96 คือมี outlier แสดงว่าตกหล่นตัวชี้วัดที่สำคัญ
4. Residual covariance ควรมีค่าต่ำและไม่มีนัยสำคัญ หากมีค่าสูงและมีนัยสำคัญแสดงว่าตกหล่นตัวชี้วัดที่สำคัญ

Vanishing Tetrad/Confirmatory Tetrad Analysis (CTA-PLS)

5. **Vanishing tetrad** ใช้ร่วมพิจารณาตัดสินว่าตัวชี้วัดควรเป็นแบบใด

1) ถ้า Vanishing tetrad ไม่มีนัยสำคัญแสดงว่าตัวชี้วัดเป็น reflective indicator

2) ถ้า Vanishing tetrad มีนัยสำคัญแสดงว่าตัวชี้วัดเป็น formative indicator

สมมุติทางสถิติคือ H_0 : reflective indicator vs H_1 : formative indicator

Stone-Gleisser Q^2

6. Stone-Gleisser Q^2 ใช้ช่วยประเมินว่าตัวแปรสาเหตุตัวใดมีผลในทางช่วยพยากรณ์ค่าตัวชี้วัดของตัวแปรผลลัพธ์ (predictive relevance) มากน้อยเพียงใด

$$Q^2 = 1 - \frac{SSE}{SSO} \quad \text{โดยที่ } SSE = \text{sum square error} \text{ และ } SSO = \text{sum square observed}$$

เกณฑ์ตัดสินใจคือ $Q^2 = 0.02, 0.15, 0.35$

(เรียกว่า effect size) แสดงว่าตัวแปรสาเหตุมีผลในทางช่วยการพยากรณ์ค่าตัวชี้วัดของตัวแปรผลลัพธ์ (indicator cross-validated redundancy) หรือตัวแปรผลลัพธ์ (construct cross-validated redundancy) ใต้น้อย ปานกลาง และมาก ตามลำดับ

$Q^2 \leq 0$ แสดงว่าตัวแปรสาเหตุไม่มีส่วนช่วยคาดคะเนเลย

Stone-Gleisser Q^2

เราจะคำนวณหาค่า Q^2 ของทุกตัวแปรที่เป็นสาเหตุของทุกตัวแปรผลลัพธ์ที่มีในตัวแบบ

กระบวนการวิเคราะห์ใช้วิธี blindfolding cross validation โดย

1. ทำให้ค่าสังเกตจริงของตัวชี้วัดทุกตัวของตัวแปรผลลัพธ์หายไปชั่วคราว D หน่วย
2. แทนที่ missing data ด้วย mean
3. รัน SEM แล้วพยากรณ์ค่าที่สูญหายไปนี้ด้วยค่าตัวแปรต้นทางคูณกับค่าสัมประสิทธิ์เส้นทาง แล้วพยากรณ์ต่อไปจนถึงค่าของตัวชี้วัดที่เป็นค่าสูญหายแล้วหาค่าผลต่างระหว่างค่าจริง (ที่ส่งให้หายไปชั่วคราวเรียกว่าค่า observe) กับค่าที่พยากรณ์ได้ เรียกว่า error
4. วนทำเช่นนี้ทุกระยะเว้นระหว่าง case โดยแนะนำให้เว้น $D = 7$ case วนเวียนเรื่อยไปจนถึงสิ้นสุดไฟล์และครบทุก case
5. คำนวณหา $Q^2 = 1 - \frac{SSE}{SSO}$

Q² VS R²

Q² ใช้แสดงว่าตัวแปรสาเหตุทุกตัวช่วยกันพยากรณ์ค่าของตัวชี้วัดของตัวแปรผลลัพธ์ได้ดีเพียงใด

q² ใช้แสดงว่า ตัวแปรสาเหตุแต่ละตัวสามารถพยากรณ์ค่าของตัวชี้วัดของตัวแปรผลลัพธ์ได้ดีเพียงใด

Q² และ q² คิดตามแนวทางเดียวกับกับ R² และ f² ต่างกันที่

R² ใช้แสดงว่าตัวแปรสาเหตุทุกตัวช่วยกันอธิบายความผันแปรของตัวแปรผลลัพธ์ได้ดีเพียงใด

f² ใช้แสดงว่าตัวแปรสาเหตุแต่ละตัวอธิบายความผันแปรของตัวแปรผลลัพธ์ได้ดีเพียงใด

Importance-Performance Map Analysis (IPMA)

7. Importance-Performance Map Analysis (IPMA) ใช้ช่วยตรวจสอบว่าตัวแปรสาเหตุ
ตัวใดมีความสำคัญหรือไม่
ตัวแปรที่ large Importance แต่ low performance จะต้องได้รับการปรับปรุงแก้ไข

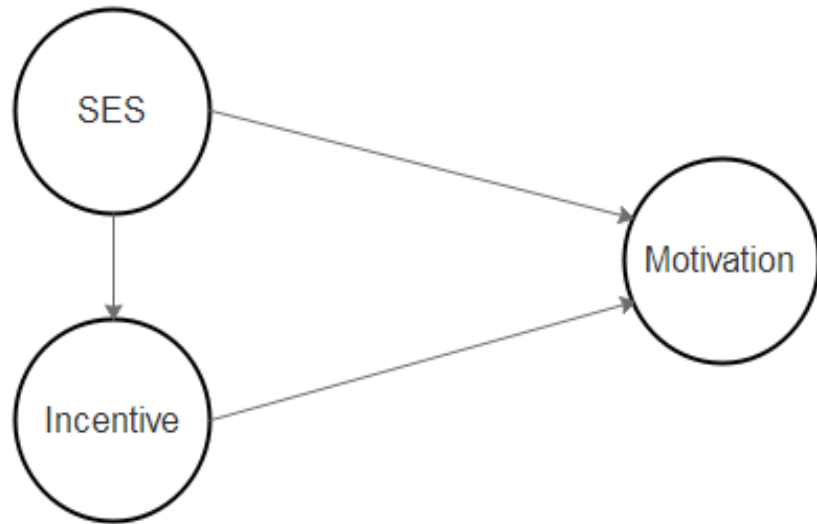
Importance-Performance Map Analysis (IPMA)

Importance = absolute total effect (คือ ไม่ต้องสนใจเครื่องหมายว่าลบหรือบวก สนใจแต่ปริมาณ) ของตัวแปรสาเหตุของตัวแปรผลลัพธ์

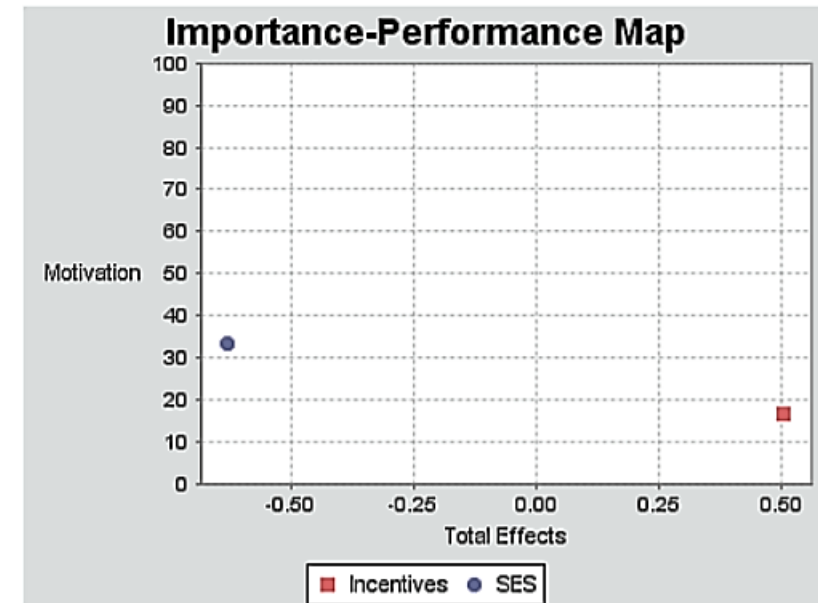
Performance = ค่าเฉลี่ยของ unstandardized latent variable score ที่ได้ normalize และทำเป็นร้อยละแล้ว กล่าวคือ

$$\text{Normalized latent score variable} = \frac{x - \min}{\max - \min} * 100$$

Importance-Performance Map Analysis (IPMA)



Construct performance	
Incentives	16.558
SES	33.030
Construct importance	
Incentives	0.502
SES	-0.628



Goodness of Fit (GoF)

8. $GoF = \sqrt{R^2_{com}}$ (Tenenhaus et al., 2005) โดยที่ $com = AVE$ เป็น global criteria และ R^2 คือ global effect size

ใช้แสดงว่าตัวแปรต้นทางทั้งหมดสามารถคาดคะเนความผันแปรตัวชี้วัดของตัวแปรผลลัพธ์ได้ดีเพียงใด

จุดอ่อน ไม่เหมาะใช้แสดงความเหมาะสมของตัวแบบ (Henseler & Sarstedt, 2013) และใช้ไม่ดีกับตัวแบบที่ไม่ถูกต้อง (mispecified model) (Hair et al., 2014)

Goodness of Fit (GoF)

GoF มีเกณฑ์ดังนี้

$$\text{GoF} = \sqrt{R^2 \text{ com}} = \sqrt{0.19 * 0.50} = 0.308 \text{ ตัวแบบมีความเหมาะสมต่ำ}$$

$$\text{GoF} = \sqrt{R^2 \text{ com}} = \sqrt{0.35 * 0.50} = 0.406 \text{ ตัวแบบมีความเหมาะสมปานกลาง}$$

$$\text{GoF} = \sqrt{R^2 \text{ com}} = \sqrt{0.67 * 0.50} = 0.578 \text{ ตัวแบบมีความเหมาะสมสูง}$$

Redundancy

9. **Redundancy** ของตัวแปรผลลัพธ์ที่ i ใช้แสดงว่าตัวแปรต้นทางสามารถคาดคะเนความผันแปรตัวชี้วัดของตัวแปรผลลัพธ์ที่ i ได้ดีเพียงใด

$$\text{Redundancy}_i = \text{Com}_i * R_i^2$$

Redundancy criterion

Redundancy of ith endogenous variable	ตัวแปรต้นทางสามารถคาดคะเน ความผันแปรตัวชี้วัดของตัวแปร ผลลัพธ์ที่ i ได้
$\text{Redundancy}_i = \text{Com}_i * R_i^2 = 0.50 * 0.19 = 0.095$	ต่ำ
$\text{Redundancy}_i = \text{Com}_i * R_i^2 = 0.50 * 0.35 = 0.175$	ปานกลาง
$\text{Redundancy}_i = \text{Com}_i * R_i^2 = 0.50 * 0.67 = 0.335$	สูง

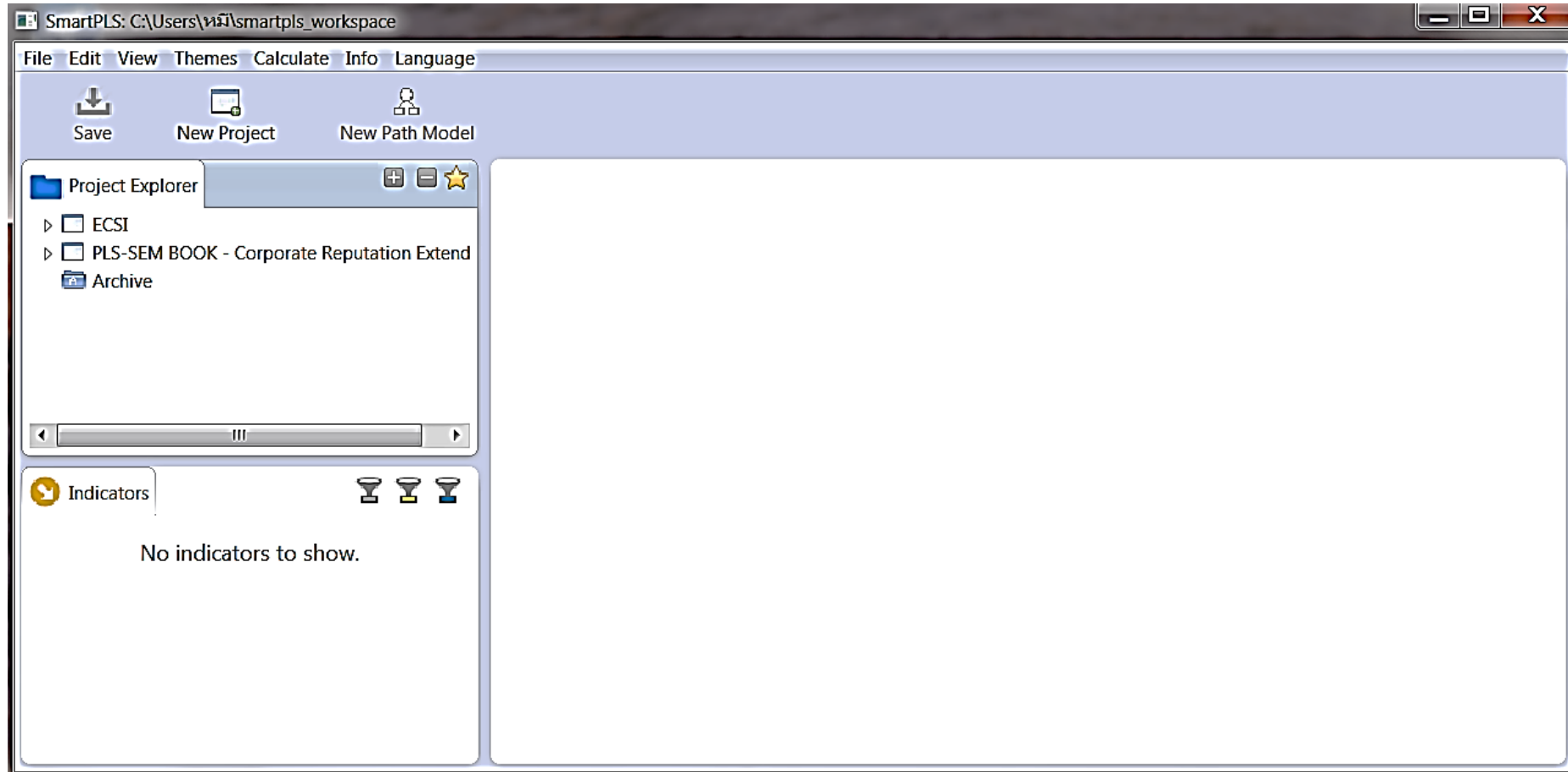
หมายเหตุ Redundancy และ GoF ไม่มีแสดงใน smart PLS3

การรันโปรแกรม Smart PLS

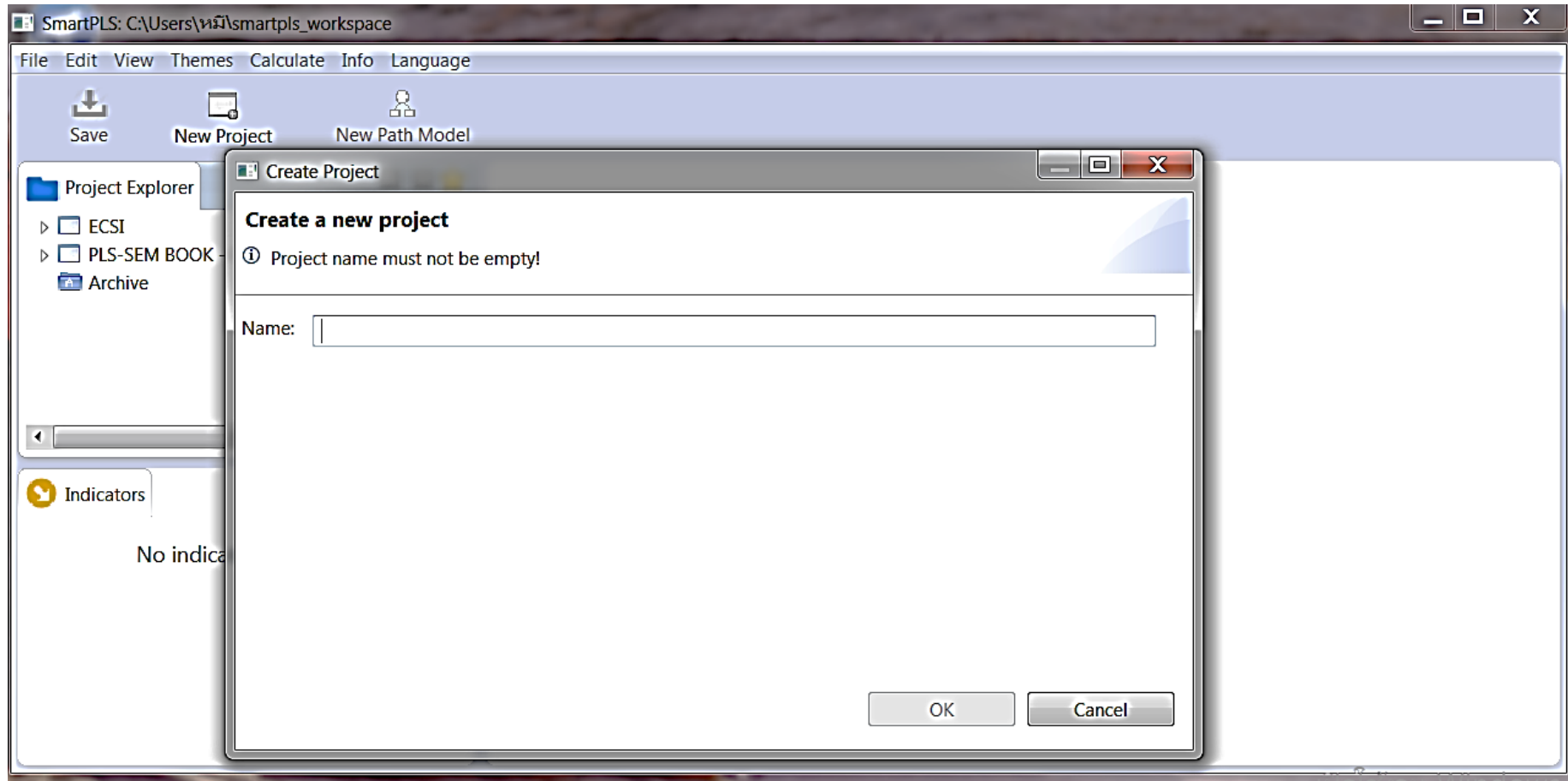
คลิกที่ไอคอนซอฟต์แวร์



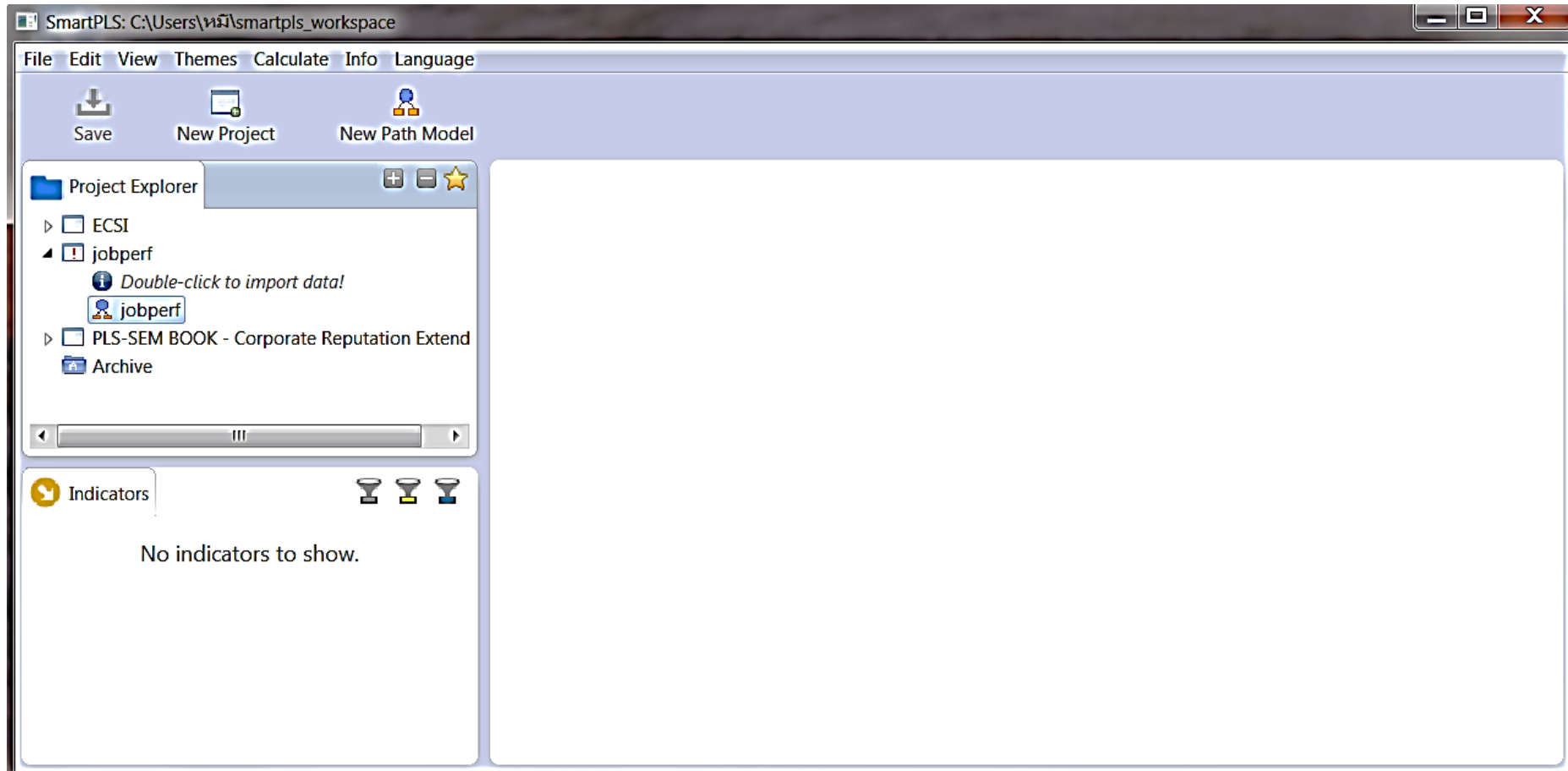
จะพบหน้าจอซอฟต์แวร์ มี 3 หน้าต่าง



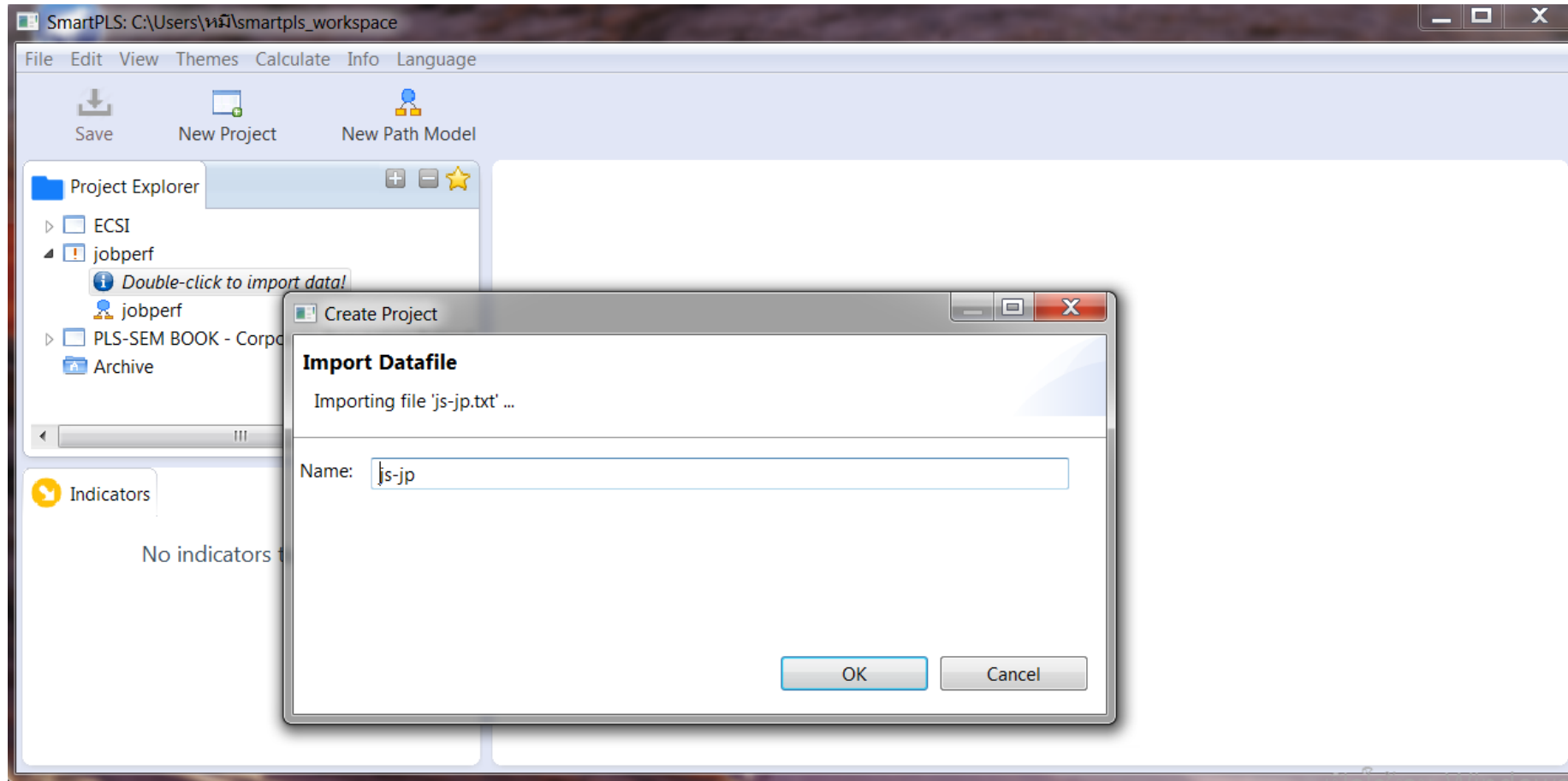
กด New Project > พิมพ์ชื่อโครงการ



มี 2 ไอคอนเพิ่มเข้ามาคือ *Double click to import data* กับ ไอคอนชื่อโปรเจกต์



Db1 click ที่ไอคอน double click to import data > ใส่ชื่อเพิ่มข้อมูล

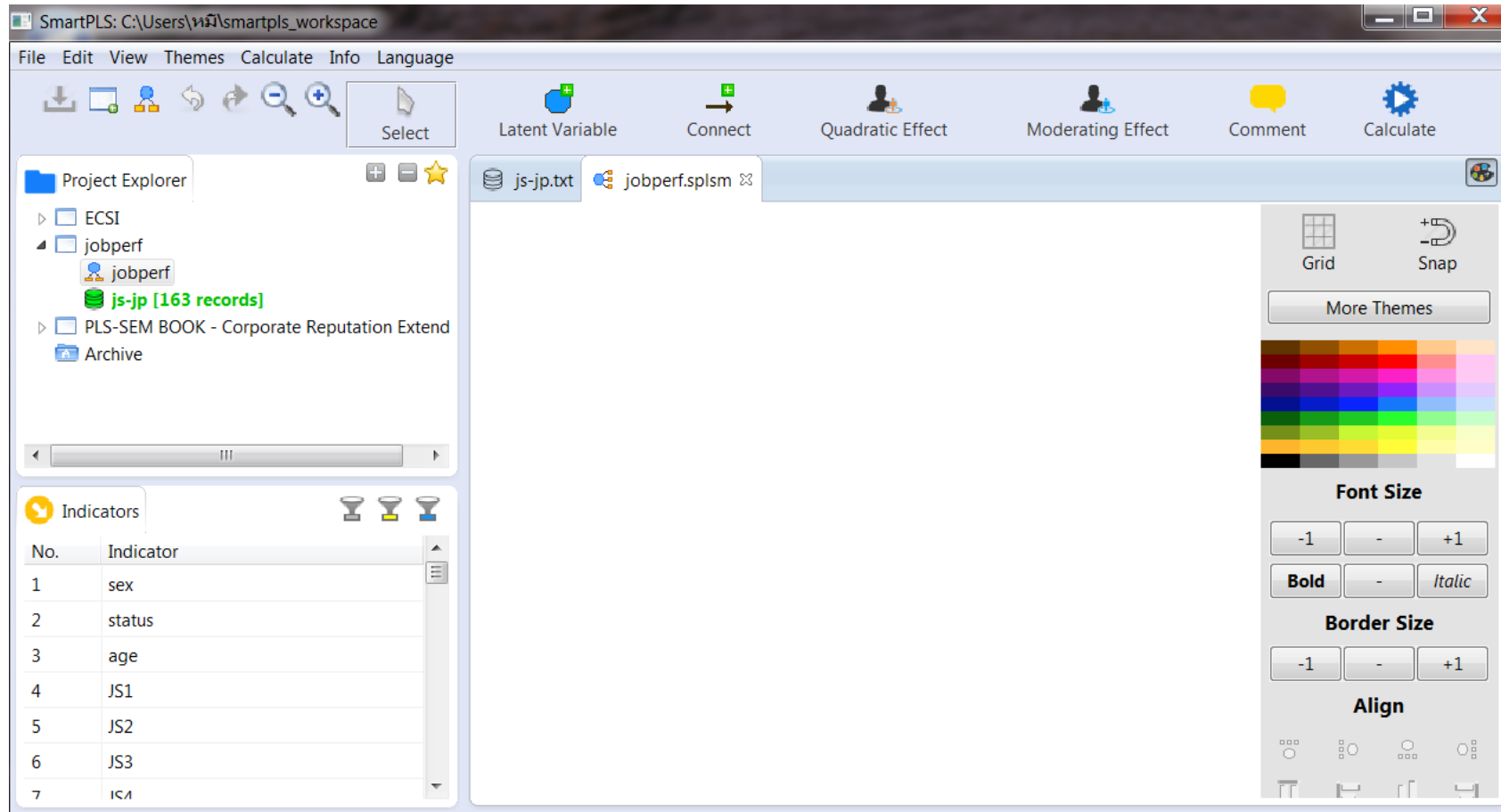


คลิกไอคอนชื่อเพิ่มข้อมูลจะแสดงข้อมูล ถ้าไม่แสดงแปลว่าเพิ่มข้อมูลมีปัญหา missing data หรือมีข้อมูลผิดปกติ เช่นมีข้อความปนอยู่

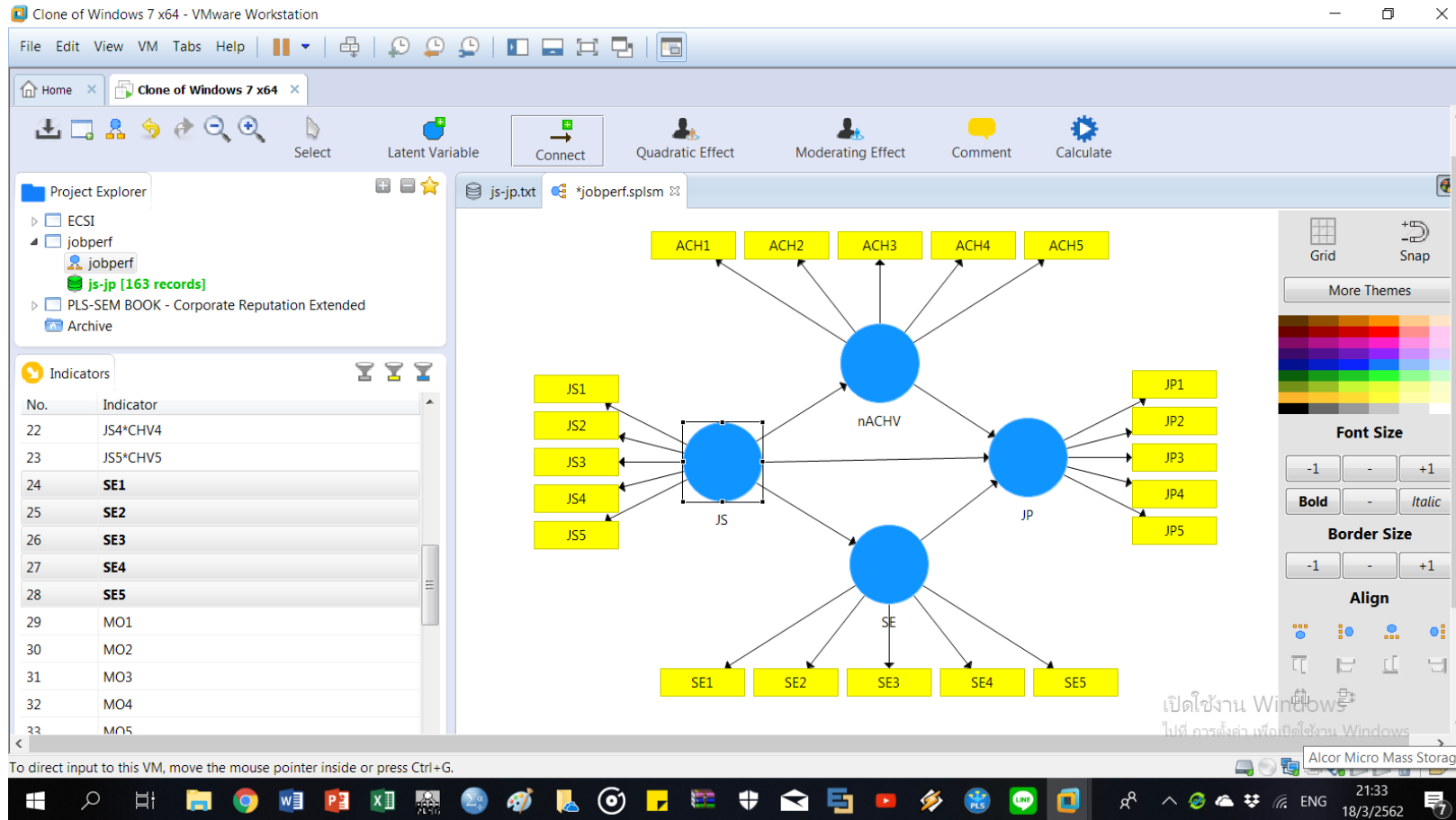
The screenshot shows the SmartPLS software interface. The main window displays the 'js-jp.txt' data file with the following settings: Delimiter: Tabulator, Encoding: UTF-8, Value Quote Character: None, Sample size: 163, Number Format: US (e.g. 1,000.23), Indicators: 55, and Missing Value Marker: None. Below these settings is a table of indicators with their respective statistics.

Indicators:	Indicator	Correlations	Raw File						
	No.	Missing	Mean	Median	Min	Max	Standard D...	Exc...	
sex	1	0	1.601	2.000	1.000	2.000	0.490		
status	2	0	1.393	1.000	1.000	2.000	0.488		
age	3	0	33.552	32.000	23.000	50.000	7.335		
JS1	4	0	4.209	4.000	2.000	5.000	0.747		
JS2	5	0	4.196	4.000	1.000	5.000	0.820		
JS3	6	0	3.850	4.000	1.000	5.000	0.938		
JS4	7	0	4.104	4.000	1.000	5.000	0.811		

คลิกไอคอนชื่อโปรเจกต์ ข้อมูลหน้าต่างที่ 2 จะแสดงเป็นชื่อตัวแปรและหน้าต่างที่ 3 พร้อมใช้งาน ถ้าหน้าต่างที่ 2 ไม่แสดงอะไรแปลว่ามีปัญหาเรื่องข้อมูลหรือไลเซนส์



Shift click ตัวชี้วัดของตัวแปรในหน้าต่างที่ 2 แล้วลากไปวางในหน้าต่างที่ 3 เชื่อมโยง และตั้งชื่อตัวแปรแฝง



กด calculation แนะนำให้เลือกใช้ Consistent PLS Algorithm

The screenshot displays the SmartPLS software interface. The main window shows a path model with three latent variables: JS (Job Satisfaction), nACHV (Need for Achievement), and JP (Job Performance). JS is measured by indicators JS1 through JS5. nACHV is measured by indicators ACH1 through ACH5. JP is measured by indicators JP3 through JP5. There are also indicators SE1 through SE5, which are measured by the latent variable SE. The model shows paths from JS to nACHV, from nACHV to JP, and from SE to nACHV and JP. A context menu is open over the 'Calculation' button, listing various algorithms and options. The 'Consistent PLS Algorithm' is highlighted. The 'Indicators' table on the left lists the indicators and their corresponding latent variables. The 'Best correlation' section at the bottom indicates a correlation of 0.680 between SE5 and SE3.

No.	Indicator
22	JS4*CHV4
23	JS5*CHV5
24	SE1
25	SE2
26	SE3
27	SE4
28	SE5
29	MO1
30	MO2
31	MO3
32	MO4
33	MO5
34	age1

Best correlation
SE5 -> SE3 : 0.680

ได้อะล็อก setting เปิดขึ้นให้กำหนดเอง ถ้าไม่ประสงค์จะกำหนดก็ปล่อยให้ทำงานตาม default แต่ควรกำหนดจำนวนการวนหาคำตอบเป็น 5,000-10,000 รอบ

Consistent PLS

The consistent PLS (PLSc) algorithm performs a correction of reflective constructs' correlations to make results consistent with a factor-model.

[Read more!](#)

The screenshot shows the 'Consistent PLS' settings window in SmartPLS. The interface is divided into three tabs: 'Setup', 'Partial Least Squares', and 'Weighting'. The 'Weighting' tab is active. The settings are organized into 'Basic Settings' and 'Advanced Settings' sections.

Basic Settings

- Weighting Scheme: Centroid Factor Path
- Maximum Iterations: 300
- Stop Criterion (10^{-X}): 7

Advanced Settings

- Initial Weights: Use Lohmoeller Settings
or configure [individual initial weights](#)

Stop Criterion

The PLS algorithm stops when the change in the outer weights between two consecutive iterations is smaller than this stop criterion value (or the maximum number of iterations is reached). This value should be sufficiently small (e.g., 10^{-5} or 10^{-7}).

Advanced Settings

Initial Outer Weights

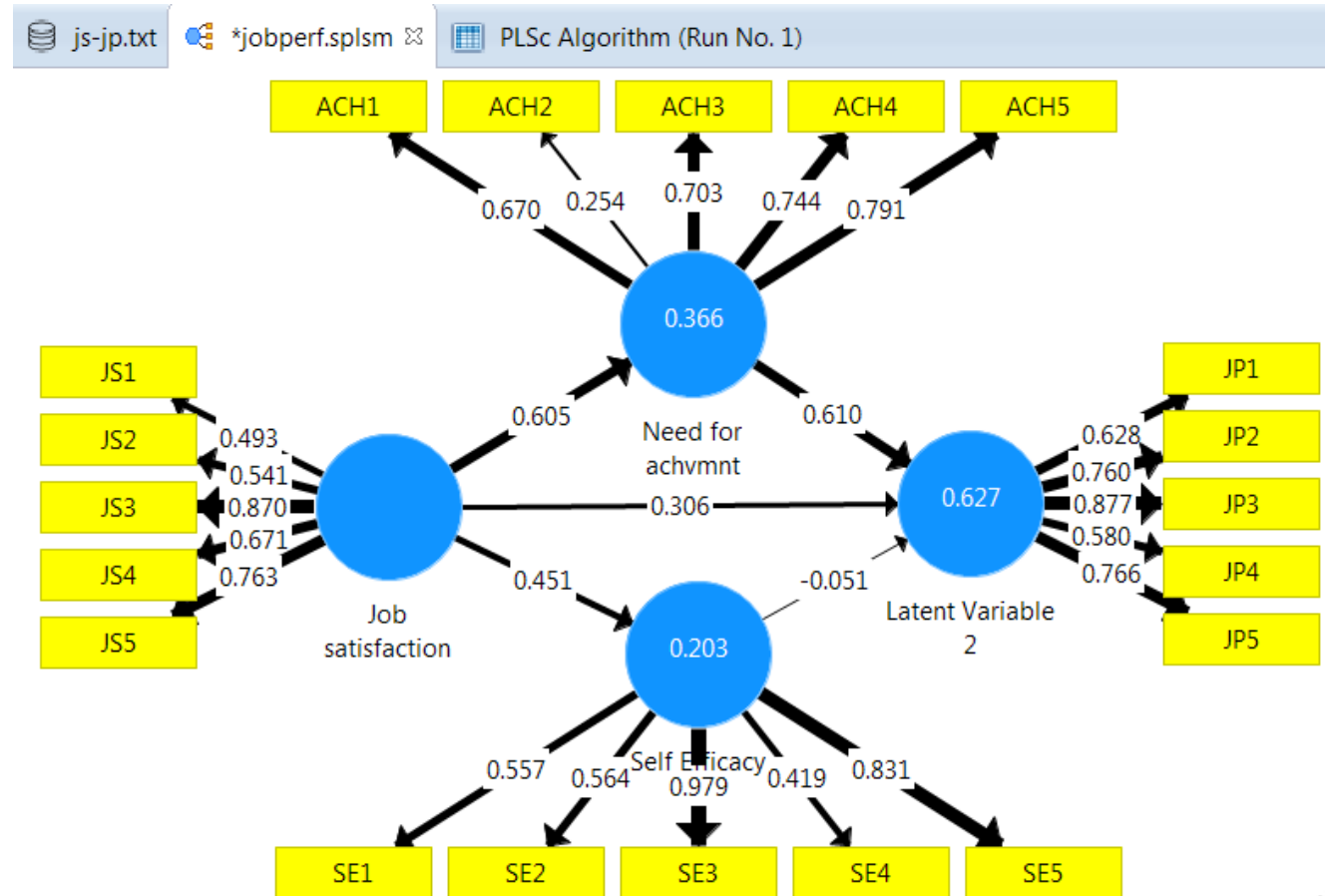
As the default (i.e., the SmartPLS settings), the initial outer weights are set to +1. However, the following alternatives are available:

- Lohmöller Settings:** Lohmöller suggested using +1 as initial outer weight for all indicators per measurement model except the last one, which uses an initial outer weight of -1. Thereby, the PLS-SEM algorithm converges faster. However, this kind of initialization can lead to counterintuitive signs of estimated PLS path coefficients in the measurement models and/or in the structural model.
- Individual Settings:** SmartPLS to define individual initial outer weights for every indicator in the PLS path model. For example, are particularly important indicator can obtain a +1 (e.g., when the strong an positive relationship with the latent variable is assumed a prior), while the other indicators of the same measurement model obtain a 0.

[Link to Literature](#)

After Calculation: [Open Full Report](#) [Close](#) [Start Calculation](#)

กด start calculation จากนั้นให้ดูผลจากภาพและ report ว่าตรงตามเกณฑ์คุณภาพหรือไม่ ถ้าไม่ตรง
 เกณฑ์ต้องปรับแก้ให้ตรงเกณฑ์และตรงตามวรรณกรรม
 ผลการรัน consistent PLS Algorithm ปราบกฏดังนี้



เก็บ

ผลการรัน consistent PLS Algorithm

The screenshot shows the SmartPLS software interface. The main window displays the 'Path Coefficients' table, which is a lower triangular matrix of path coefficients, standard errors (SE), and R-squared values for the variables JobPerf, JobSat, SE, and needACHV. The bottom panel shows the 'Final Results' and 'Quality Criteria' section, which includes links to various diagnostic statistics and model fit measures.

Path Coefficients

	JobPerf	JobSat	SE	needACHV
JobPerf				
JobSat	0.306		0.451	0.605
SE	-0.051			
needACHV	0.610			

Final Results

Final Results	Quality Criteria	Interim Results	Base Data
Path Coefficients	R Square	Stop Criterion Changes	Setting
Indirect Effects	f Square	Correction Value C	Inner Model
Total Effects	Construct Reliability and Validity		Outer Model
Outer Weights	Discriminant Validity		Indicator Data (Original)
Outer Loadings	Collinearity Statistics (VIP)		Indicator Data (Standardized)
Latent Variable	Model Fit		Indicator Data (Correlations)
Residuals			

การคำนวณเลือก Bootstrapping

Bootstrapping

Bootstrapping is a nonparametric procedure that allows testing the statistical significance of various PLS-SEM results such path coefficients, Cronbach's alpha, HTMT, and R² values.

[Read more!](#)

Setup Partial Least Squares Weighting

Basic Settings

Subsamples:

Do Parallel Processing

Sign Changes: No Sign Changes
 Construct Level Changes
 Individual Changes

Amount of Results: Basic Bootstrapping
 Complete Bootstrapping

Advanced Settings

Confidence Interval Method: Percentile Bootstrap
 Studentized Bootstrap
 Bias-Corrected and Accelerated (BCa) Bootstrap
 Davision Hinkley's Double Bootstrap
 Shi's Double Bootstrap

Test Type: One Tailed Two Tailed

Significance Level:

Basic Settings

Subsamples

In bootstrapping, subsamples are created with observations randomly drawn (with replacement) from the original set of data. To ensure stability of results, the number of subsamples should be large. For an initial assessment, one may use a smaller number of bootstrap subsamples (e.g., 500). For the final results preparation, however, one should use a large number of bootstrap subsamples (e.g., 5,000).
Note: Larger numbers of bootstrap subsamples increase the computation time.

Do Parallel Processing

This option runs the bootstrapping routine on multiple processors (if your computer device offers more than one core). Using parallel computing will reduce computation time.

Sign Changes

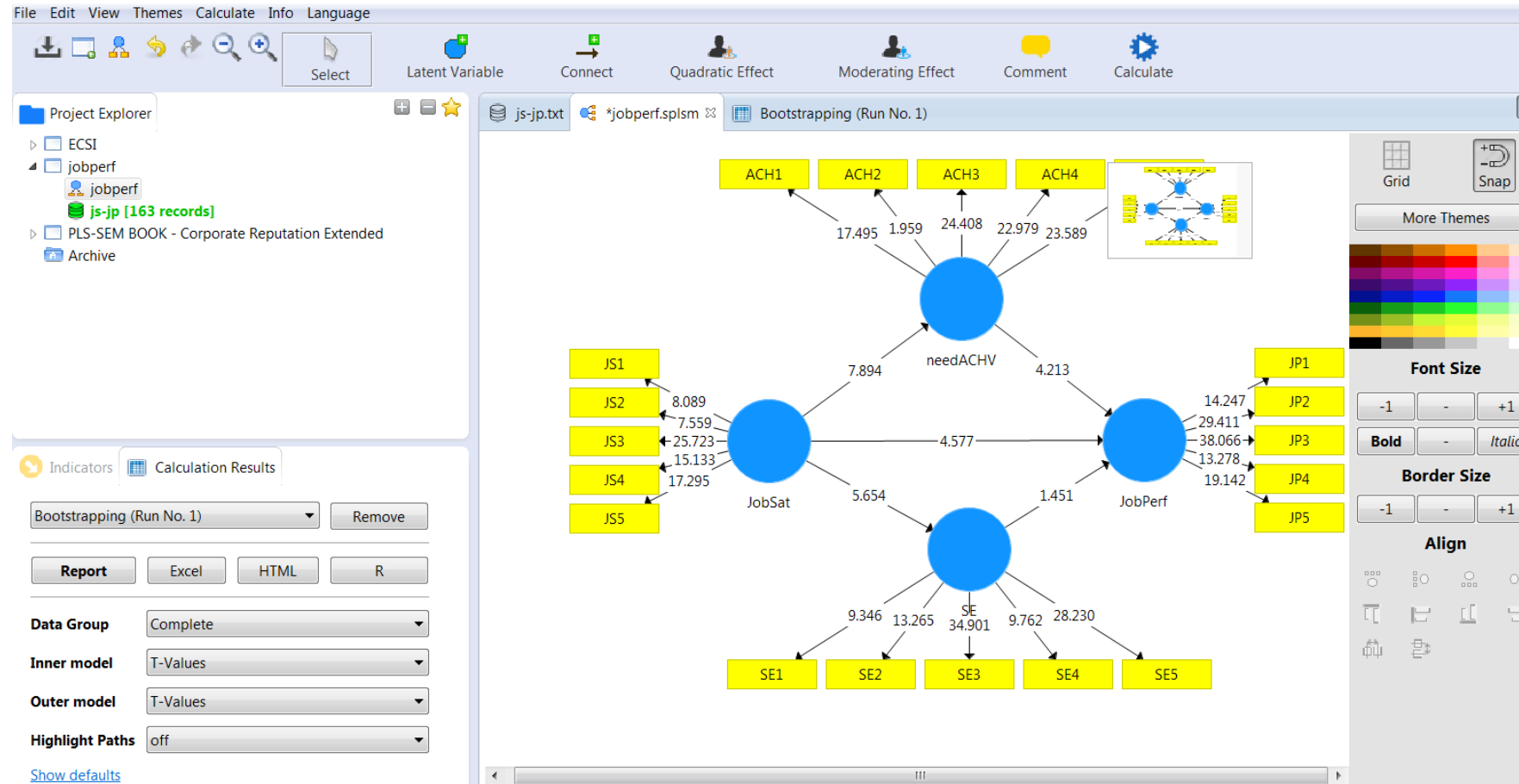
Sets the method for dealing with sign changes during the bootstrap iterations. The following options are available:

(1) No Sign Changes (*default*)
Sign changes in the resamples will be ignored and the results are taken as they are. This is the most conservative estimation option and the recommended choice when running the bootstrapping routine.

(2) Construct Level Changes
The signs of a group of coefficients (e.g., all outer loadings of a specific latent variable) in a bootstrapping subsample are compared with the signs of the original PLS path model estimation. If the majority of signs need to be reversed in a bootstrap run to better match the signs of the model estimation using the original sample, all signs are reversed in that bootstrap run.

After Calculation:

ผลการรัน Bootstrapping



ผลการรัน Bootstrapping

File Edit View Themes Calculate Info Language

Save New Project New Path Model Hide Zero Values Increase Decimals Decrease Decimals Export to Excel Export to Web Export to R

Project Explorer

- ECSI
- jobperf
 - js-jp [163 records]
- PLS-SEM BOOK - Corporate Reputation Extended
 - Archive

Indicators

No.	Indicator
49	ZMO
50	ZJSZnCH
51	JSf
52	JPf
53	nCHVf
54	SEf
55	MOF

Best correlation
SE5 -> SE3 : 0.680

js-jp.txt *jobperf.splsm Bootstrapping (Run No. 1)

Path Coefficients

	Original Sa...	Sample Me...	Standard D...	T Statistics (...)	P Values
JobSat -> J...	0.311	0.319	0.068	4.577	0.000
JobSat -> SE	0.386	0.398	0.068	5.654	0.000
JobSat -> n...	0.501	0.508	0.063	7.894	0.000
SE -> JobPerf	0.127	0.123	0.087	1.451	0.147
needACHV ...	0.386	0.387	0.092	4.213	0.000

Mean, STDEV, T-Values, P... Confidence Intervals Confidence Intervals Bias... Samples Copy to Clipboard: Excel Format R Format

Final Results	Quality Criteria	Histograms	Base Data
Path Coefficients	R Square	Path Coefficients Histogram	Setting
Indirect Effects	R Square Adjusted	Indirect Effects Histogram	Inner Model
Total Effects	f Square	Total Effects Histogram	Outer Model
Outer Loadings	Average Variance Extracted (AVE)		Indicator Data (Original)
Outer Weights	Composite Reliability		Indicator Data (Standardized)
	rho A		
	Cronbach's Alpha		

เอกสารอ้างอิง

Chin, W. W. (1998). The partial least squares approach for structural equation modeling. Pp. 295-336 in Macoulides, G. A., ed. Modern methods for business research. Mahwah, NJ: Lawrence Erlbaum Associates.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Mahwah, NJ: Lawrence Erlbaum

Garson, G. D. (2016). Partial Least Squares: Regression and Structural Equation Models. Asheboro, NC: Statistical Associates Publishers.

Hair, Joseph F., Jr.; Hult, G. Tomas M.; Ringle, Christian M.; & Sarstedt, Marko (2014). A primer on partial least squares structural equation modeling (PLSSEM). Thousand Oaks, CA: Sage Publications.

Henseler, Jörg; Ringle, C. M.; & Sarstedt, M. (2016). Testing measurement invariance of composites using partial least squares. International Marketing Review. Forthcoming

เอกสารอ้างอิง

- Henseler, Jörg & Sarstedt, Marko (2013). Goodness of fit indices for partial least squares path modeling. *Computational Statistics* 28(2): 565-580.
- Höck, Michael & Ringle, Christian M. (2006). Strategic networks in the software industry: An empirical analysis of the value continuum. IFSAM VIIIth World Congress, Berlin 2006. Retrieved 2/22/2009 from <http://www.iblunihh.de/IFSAM06.pdf>.
- Tenenhaus M., Esposito Vinzi V., Chatelin Y.M., Lauro N.C. (2005): PLS Path Modeling. *Computational Statistics and Data Analysis*, 48, 159-205.
- Temme, D., Kreis and Hildebrandt, L. (2010). A Comparison of Current PLS Path Modeling Software: Features, Ease-of-Use, and Performance. In book: *Handbook of Partial Least Squares*.
- Wold, Herman (1982). Soft modeling: The basic design and some extensions. Pp. 1-54 in Jöreskog, K. G. & Wold, H., eds., *Systems under indirect observations*., Part II. Amsterdam: North-Holland.