การใช้โปรแกรม smart PLS สำหรับการวิจัยด้วยตัวแบบสมการโครงสร้าง

บรรยายในงาน มหกรรมวิจัยแห่งชาติ 2562 (Thailand Research Expo 2019)

10 เมษายน 2562 ณ โรงแรมเซ็นทาราแกรนด์ เซ็นทรัลเวิลด์ กรงเทพมหานคร

โดย

รองศาสตราจารย์ ดร.มนตรี พิริยะกุล

ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยรามคำแหง

Email: mpiriyakul@yahoo.com line ID: mpiriyakul

PLS คืออะไร

PLS (partial least square) เป็นซอฟท์แวร์ที่รันชุดของสมการถดถอยวนเวียนจนได้รับคำตอบ (convergence) ด้วยวิธี least square พัฒนาอัลกอริธึมโดย Herman Wold (1982) โปรแกรม Smart PLS พัฒนาโดย Christian M. Ringle, Sven Wende, and Alexander Will เวอร์ชั่น 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 สามารถดาวน์โหลดมาใช้ได้นาน 3 เดือนและจะต่อไลเซนส์ให้ฟรี

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

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

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),

plspm (Sanchez, Trinchera and Russolillo, 2015)

LVPLS (Lohm¨oller, 1984),

PLSPath (Sellin, 1989)

PLS algorithm

PLS จะการวนเวียนคำนวณด้วยวิธี least square estimation ซ้ำจนกว่าจะได้รับคำตอบโดยกระทำกับสมการถดถอยที่มีจำนวนมากเท่ากับจำนวนสมการโครงสร้าง (structural model, inner model) และจำนวนสมการมาตรวัด (measurement model, outer model) รวมกัน

1) ประมาณค่าสัมประสิทธ์ถดถอยสมการมาตรวัดโดยรันทีละบล็อก

2) สร้างค่าของตัวแปรแฝงของบล็อก ค่าของ LV ถูกสร้างขึ้นจาก MV ในบล็อกของตนและให้นอร์มอลไลซ์ให้มี SD =1

3) ประมาณค่าสัมประสิทธิ์เส้นทางของสมการโครงสร้าง

4) พยากรณ์ค่าตัวแปรแฝงขึ้นมาใหม่จากสมการโครงสร้าง

5) ประมาณค่าสัมประสิทธ์ถดถอยสมการมาตรวัดใหม่ระหว่าง 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 จึงไม่สัมพันธ์กันและตัดทิ้งไม่ได้เพราะเท่ากับกับตัดโดเมนทิ้งไป ข้อที่เหลือไม่อาจให้สารสนเทศแทนกันได้

Formative-Reflective indicator

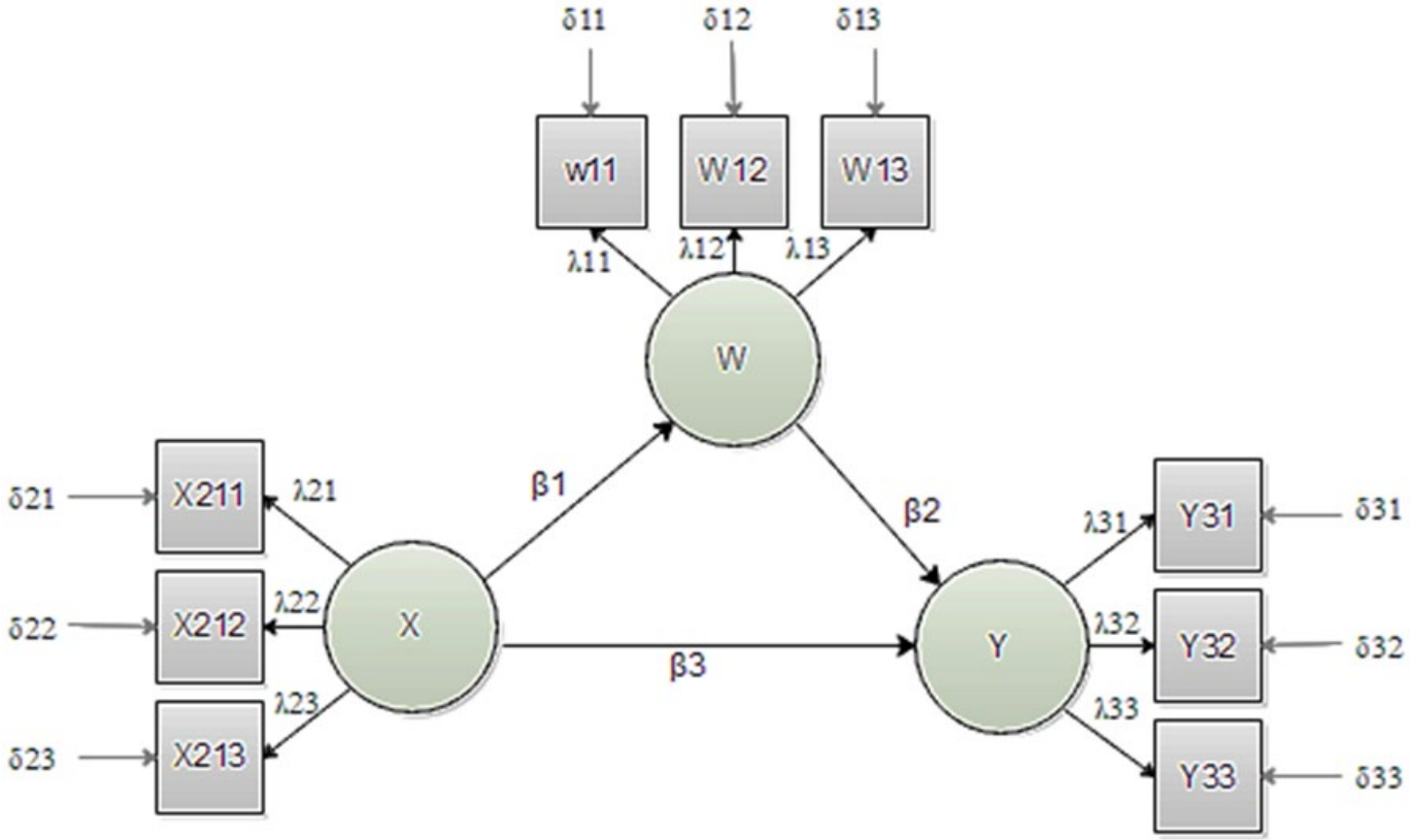
2. **reflective indicator** ตัวชี้วัดจะเป็นตัวแทนของตัวแปรแฝง เป็นตัวอย่างส่วนหนึ่งของตัวชี้วัดของแปรแฝงจึงมีความสัมพันธ์กันสูง (co-vary) เพราะวัดเรื่องเดียวกันและสามารถทดแทนกัน (interchangeability) ได้ คือตัดทิ้งไปบางข้อได้โดยข้อที่เหลืออยู่ซึ่งสัมพันธ์กันกับข้อที่ถูกตัดทิ้งไป (ที่จริงคือสัมพันธ์กับตัวแปรแฝง)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

Formative-Reflective indicator

ตัวอย่างเช่น มาตรวัดความพึงพอใจในโรงแรมที่พัก กรณี reflective indicator จะวัดรวมๆ ทั่วไปหมด เช่น ความชอบ ความรู้สึกสะดวกสบาย ความตั้งใจบอกต่อ แต่ในกรณี formative indicator จะวัดแยกเป็นโดเมน เช่นด้านพนักงาน ด้านห้องพัก ด้านความสงบ ด้านสิ่งบริการเสริมเช่นฟิตเนส

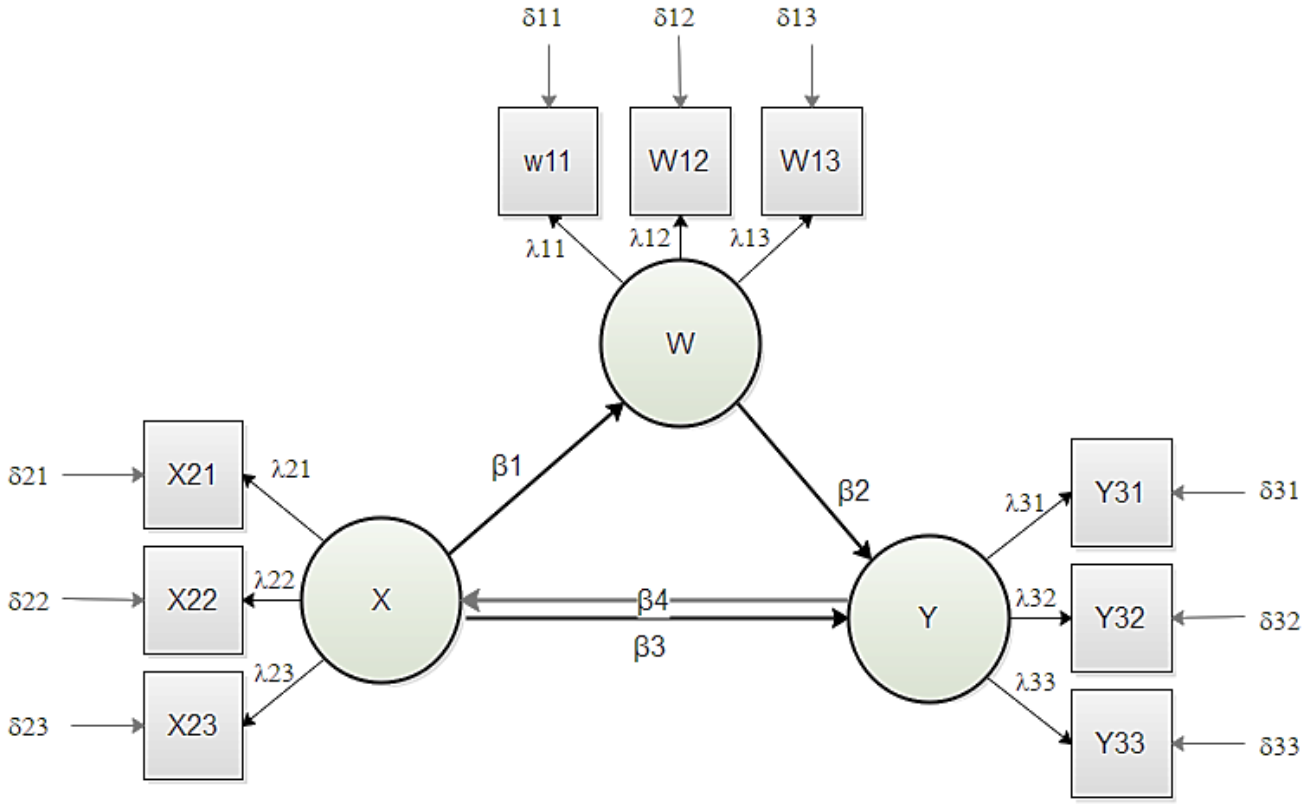
Research Design

1. Causality ตัวแปรแฝงโยงถึงกันเป็นเครือข่ายโดยแต่ละเส้นทางมีทฤษฎีกำกับ



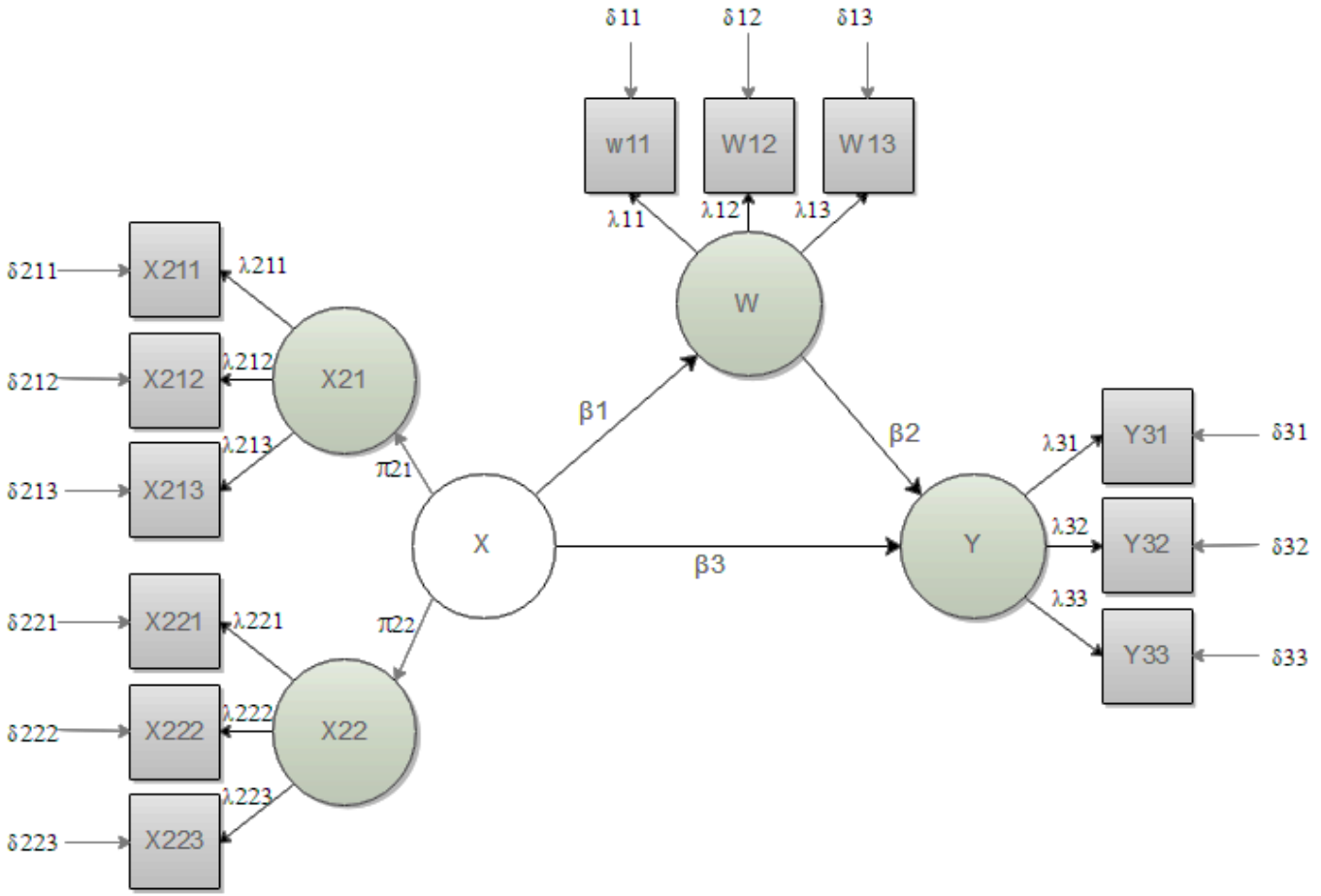
Research Design

2. non-recursive model/reciprocal model ตัวแปรแฝงโยงถึงกันเป็นเครือข่ายโดยที่แต่ละเส้นทางมีทฤษฎีกำกำกับ บางเส้นทางเป็นความสัมพันธ์ย้อนกลับ



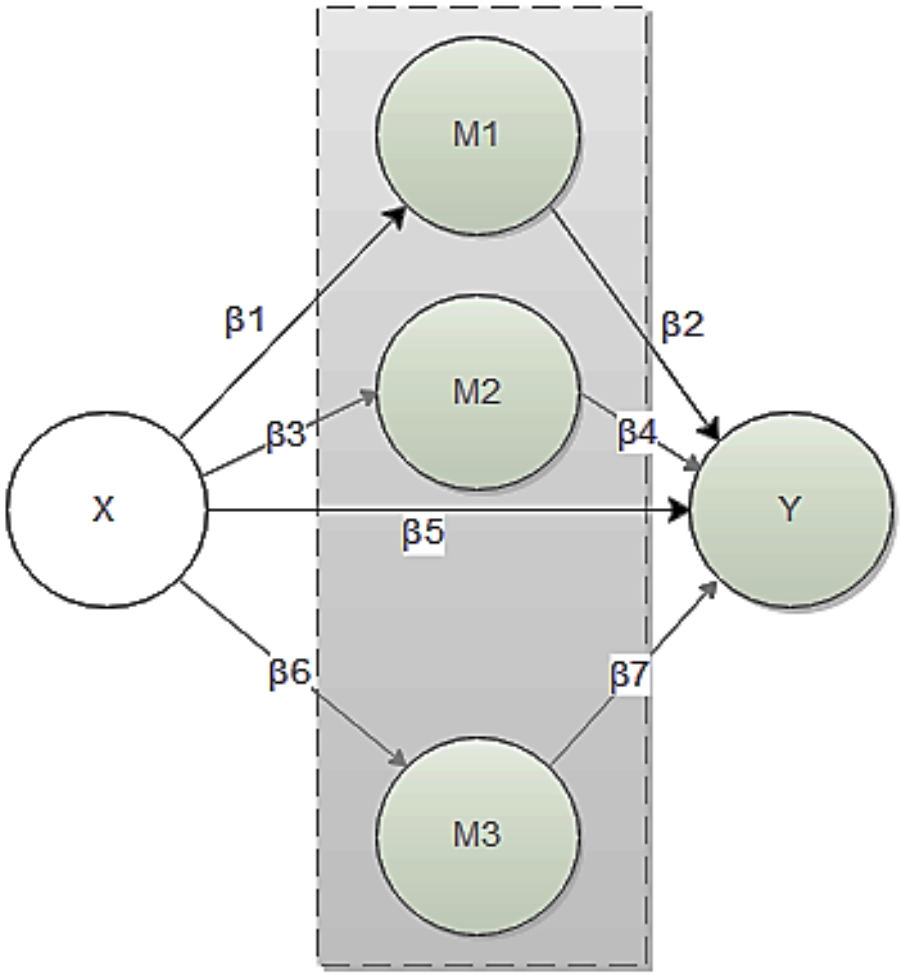
Research Design

3. Second order model ตัวแปรแฝงโยงถึงกันเป็นเครือข่ายโดยแต่ละเส้นทางมีทฤษฎีกำกำกับ ตัวแปรแฝงบางตัวหรือทุกตัวมี sub-construct ซึ่งก็คือ sub-domain ของเนื้อหา



Research Design

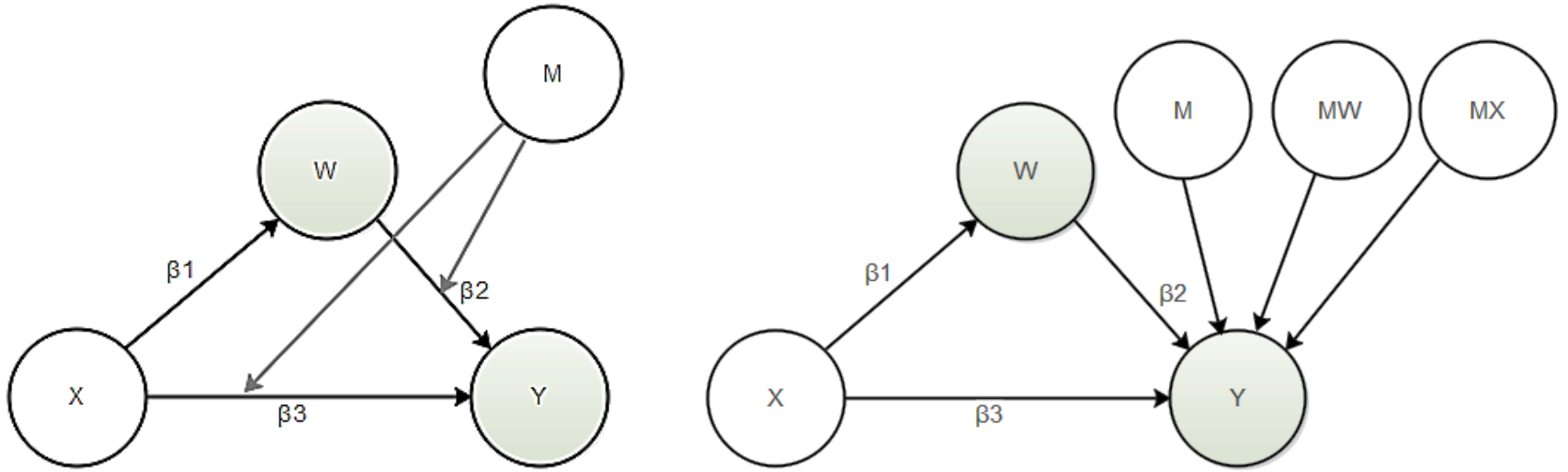
4. Mediation model ตัวแบบมีตัวแปรคั่นกลาง ตัวแปรคั่นกลางเป็นตัวแปรที่ซ่อนเร้นเชื่อมโยงระหว่างตัวแปรต้นทางกับตัวแปรปลายทาง (hidden transmitting variables) ซึ่งนักวิจัยขุดค้นนำมาแสดงให้ประจักษ์ เป็นตัวแปรที่มีอิทธิพลต่อตัวแปรปลายทางแต่อาจไม่ค่อยเกี่ยวข้องกับตัวแปรต้นทาง ถ้ามีหลายตัวเรียกว่า multiple mediation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



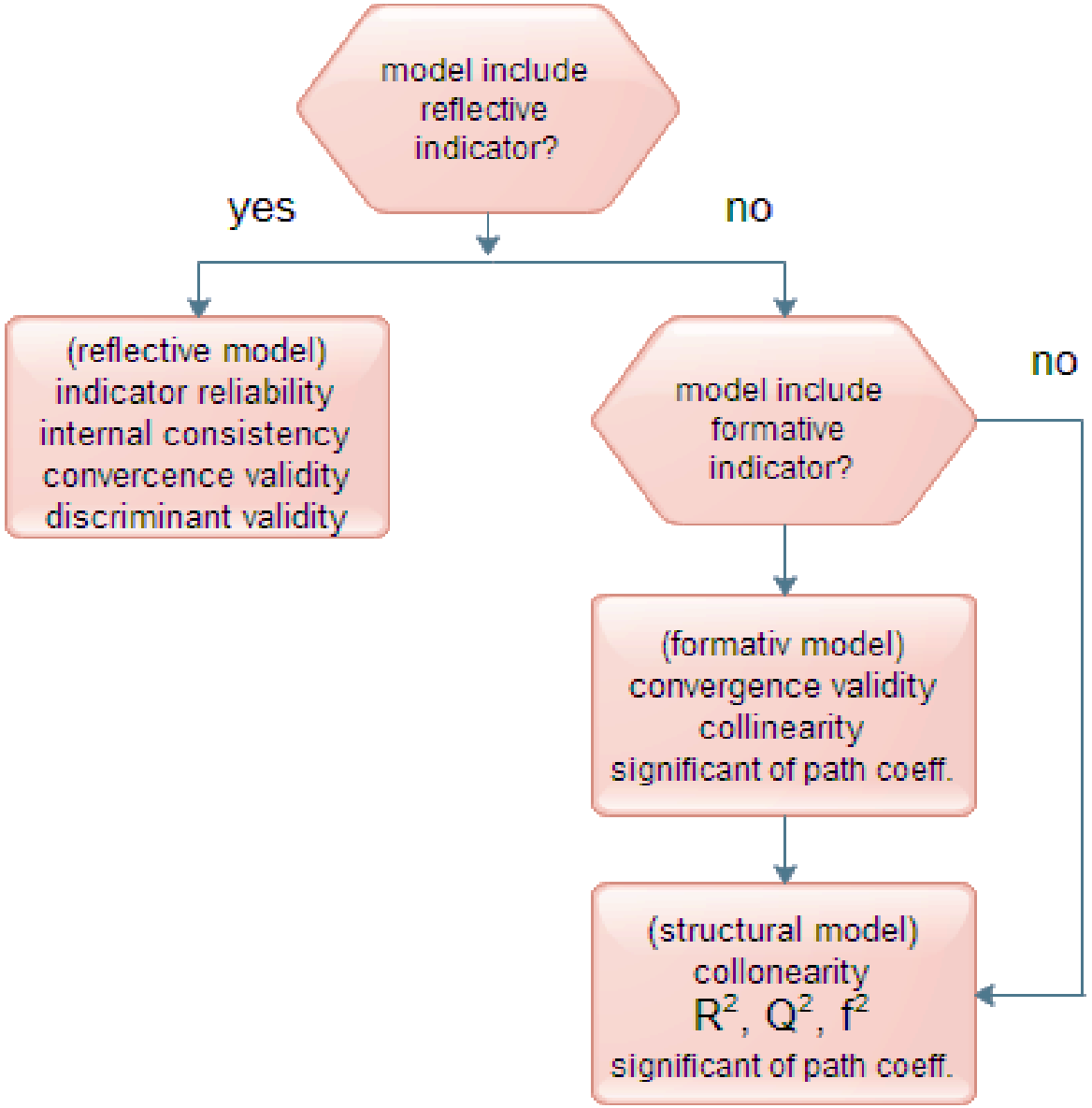
Research Design

5. Moderation model ตัวแบบมีตัวแปรกำกับ ตัวแปรกำกับเป็นตัวแปรที่ซ่อนเร้นควบคุมความสัมพันธ์ระหว่างตัวแปรต้นทางกับตัวแปรปลายทาง คือมีปฏิสัมพันธ์กับตัวแปรต้นทางแล้วส่งผลต่อตัวแปรปลายทางซึ่งนักวิจัยขุดค้นนำมาแสดงให้ประจักษ์ 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 หากมีให้แก้ปัญหาให้เสร็จก่อน

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 ไม่มีนัยสำคัญให้คงข้อถามนั้นถ้าเห็นว่ามีความสำคัญ (theoretical importance)

หมายเหตุ ตัวอย่างของ theoretical importance เช่นมาตรวัดของความตั้งใจจะต้องมีเงื่อนไขอยู่ด้วย ถ้ามีบางข้อไม่มีเงื่อนไขอยู่ในเนื้อความก็ให้พิจารณาตัดทิ้ง

data cleaning

โดยปกติ เมื่อพบเหตุการณ์ตามข้อ 2) นักวิจัยต้องพิจารณาดำเนินการ data cleaning ก่อน เช่น

แก้ปัญหา outlier ตามวิธี Winsorizing หรือ

แทนที่ข้อมูลสูญหายด้วยวิธีที่เหมาะสมกว่าเดิม หรือ

พิจารณาตัดหน่วยวิเคราะห์ที่มีข้อมูลไม่น่าเชื่อถือทิ้งไป เช่นกรณีแบบสอบถามเป็นชนิดไลเกอร์ท 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 $\sqrt{\text{AVE}}$ ในการวิเคราะห์ความเที่ยงตรงเชิงจำแนก (discriminant validity)

เกณฑ์คือ cross-loading $\leq 0.30, 0.40$

ถ้ามากกว่านี้แสดงว่ามาตรวัดขาดความเที่ยงตรงเชิงจำแนก ทั้งนี้เพราะตัวแบบที่ดีนั้นตัวชี้วัดจะต้องสัมพันธ์ (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 ถ้ามากกว่า 1 แสดงว่าขาดความเที่ยงตรงเชิงจำแนก

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

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

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 เราไม่อาจใช้ปัจจัยสาเหตุนั้นเป็นสิ่งอธิบายปัจจัยผลลัพธ์ได้อย่างถูกต้อง และไม่อาจแน่ใจได้ว่าตัวแปรสาเหตุคู่ที่สัมพันธ์กันสูงนั้นตัวใดกันแน่ที่มีอิทธิพลต่อตัวแปรผลลัพธ์ หรือหรือตัวชี้วัดคู่ที่สัมพันธ์กันสูงนั้นตัวใดกันแน่ที่มีอิทธิพลต่อตัวแปรแฝง (formative measurement model) หรือเป็นผลสะท้อนของตัวแปรแฝง (reflective measurement model)

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

วิธีแก้ปัญหา multicollinearity

วิธีแก้ปัญหา multicollinearity คือ

1) รวมตัวแปรแฝงที่สัมพันธ์กันสูงเป็นกลุ่มเดียวกัน อาจรวมในชื่อใหม่และจัดให้ตัวแปรแฝงเดิมแต่ละตัวทำหน้าที่เป็น parcel/sub-construct/sub-domain คือแก้ให้เป็นการวิจัยแบบ 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 (Ho: $\sigma_e^2 = 0$ vs H1: $\sigma_e^2 \neq 0$) 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(X1, X2) + u สามารถคำนวณหาค่า f_i^2 ตามสูตร $f_i^2 = \frac{R^2_{original} - R^2_{omitted}}{1 - R^2_{original}}$ 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	$\frac{0.2122}{0.5692} = 0.3728$
$1 - R^2_{original}$ = 1- 0.4308 = 0.5692			

residual

3. Standardized residual ถ้าพบว่ามีค่าเกินขอบเขต ±1.96 คือมี outlier แสดงว่าตกหล่นตัวชี้วัดที่สำคัญ

4. Residual covariance ควรมีค่าต่ำและไม่มีนัยสำคัญ หากมีค่าสูงและมีนัยสำคัญแสดงว่าตกหล่นตัวชี้วัดที่สำคัญ

Vanishing Tetrad/Confirmatory Tedtrad Analysis (CTA-PLS)

5. Vanishing tetrad ใช้ร่วมพิจารณาตัดสินว่าตัวชี้วัดควรเป็นแบบใด

1) ถ้า Vanishing tetrad ไม่มีนัยสำคัญแสดงว่าตัวชี้วัดเป็น reflective indicator

2) ถ้า Vanishing tetrad มีนัยสำคัญแสดงว่าว่าตัวชี้วัดเป็น formative indicator

สมมุติทางสถิติคือ Ho: reflective indicator vs H1: formative indicator

Stone-Gleisser Q^2

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

$$Q^2 = 1 - \frac{SSE}{SSO}$$ โดยที่ SSE = sum square error และ SSO = 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) ได้น้อย 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 (ที่สั่งให้หายไปชั่วคราวเรียกว่าค่า observe) กับค่าที่พยากรณ์ได้ เรียกว่า error

4. วนทำเช่นนี้ทุกระยะเว้นระหว่าง case โดยแนะนำให้เว้น D = 7 case วนเวียนเรื่อยไปจนสิ้นสุดไฟล์และครบทุก case

5. คำนวณหา $Q^2 = 1 - \frac{SSE}{SSO}$

Q^2 vs R^2

Q^2 ใช้แสดงว่าตัวแปรสาเหตุทุกตัวช่วยกันพยากรณ์ค่าของตัวชี้วัดของตัวแปรผลลัพธ์ได้ดีเพียงใด

q^2 ใช้แสดงว่า ตัวแปรสาเหตุแต่ละตัวสามารถพยากรณ์ค่าของตัวชี้วัดของตัวแปรผลลัพธ์ได้ดีเพียงใด

Q^2 และ q^2 คิดตามแนวทางเดียวกับกับ R^2 และ f^2 ต่างกันที่

R^2 ใช้แสดงว่าตัวแปรสาเหตุทุกตัวช่วยกันอธิบายความผันแปรของตัวแปรผลลัพธ์ได้ดีเพียงใด

f^2 ใช้แสดงว่าตัวแปรสาเหตุแต่ละตัวอธิบายความผันแปรของตัวแปรผลลัพธ์ได้ดีเพียงใด

Importance-Performance Map Analysis (IPMA)

7. Importance-Performance Map Analysis (IPMA) ใช้ช่วยตรวจสอบว่าตัวแปรสาเหตุตัวใดมีความสำคัญหรือไม่

ตัวแปรที่ large Importance แต่ low performance จะต้องได้รับการปรับปรุงแก้ไข

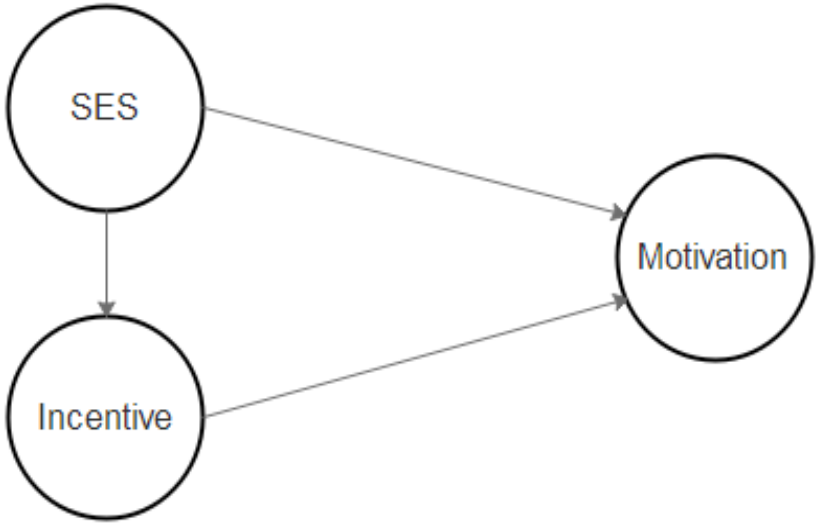
Importance-Performance Map Analysis (IPMA)

Importance = absolute total effect (คือไม่ต้องสนใจเครื่องหมายว่าลบหรือบวก สนใจแต่ปริมาณ) ของตัวแปรสาเหตุของตัวแปรผลลัพธ์

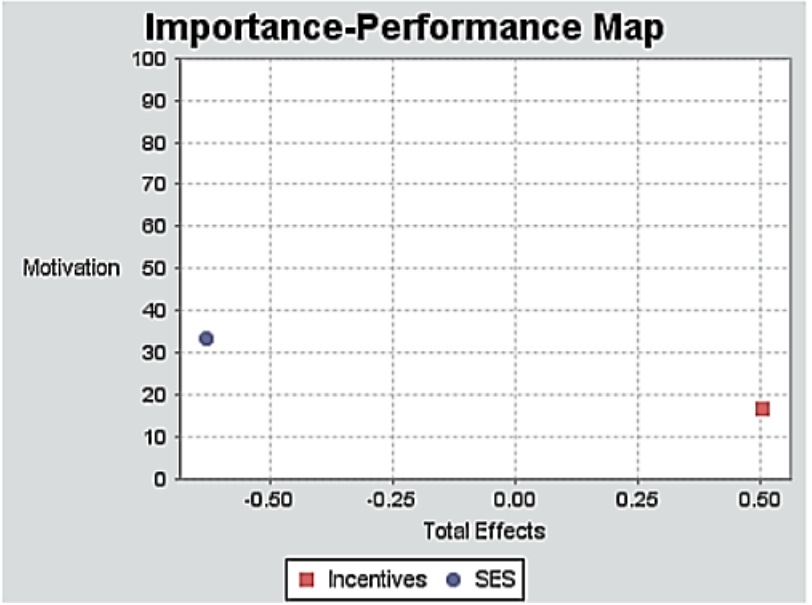
Performance = ค่าเฉลี่ยของ unstandardized latent variable score ที่ได้ normalize และทำเป็นร้อยละแล้ว กล่าวคือ

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

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 = $\sqrt{\overline{R^2}\,\overline{com}}$ (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

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

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

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

Redundancy

9. Redundancy ของตัวแปรผลลัพธ์ที่ i ใช้แสดงว่าตัวแปรต้นทางสามารถคาดคะเนความผันแปรตัวชี้วัดของตัวแปรผลลัพธ์ที่ i ได้ดีเพียงใด

$$Redundancy_i = Com_i * R_i^2$$

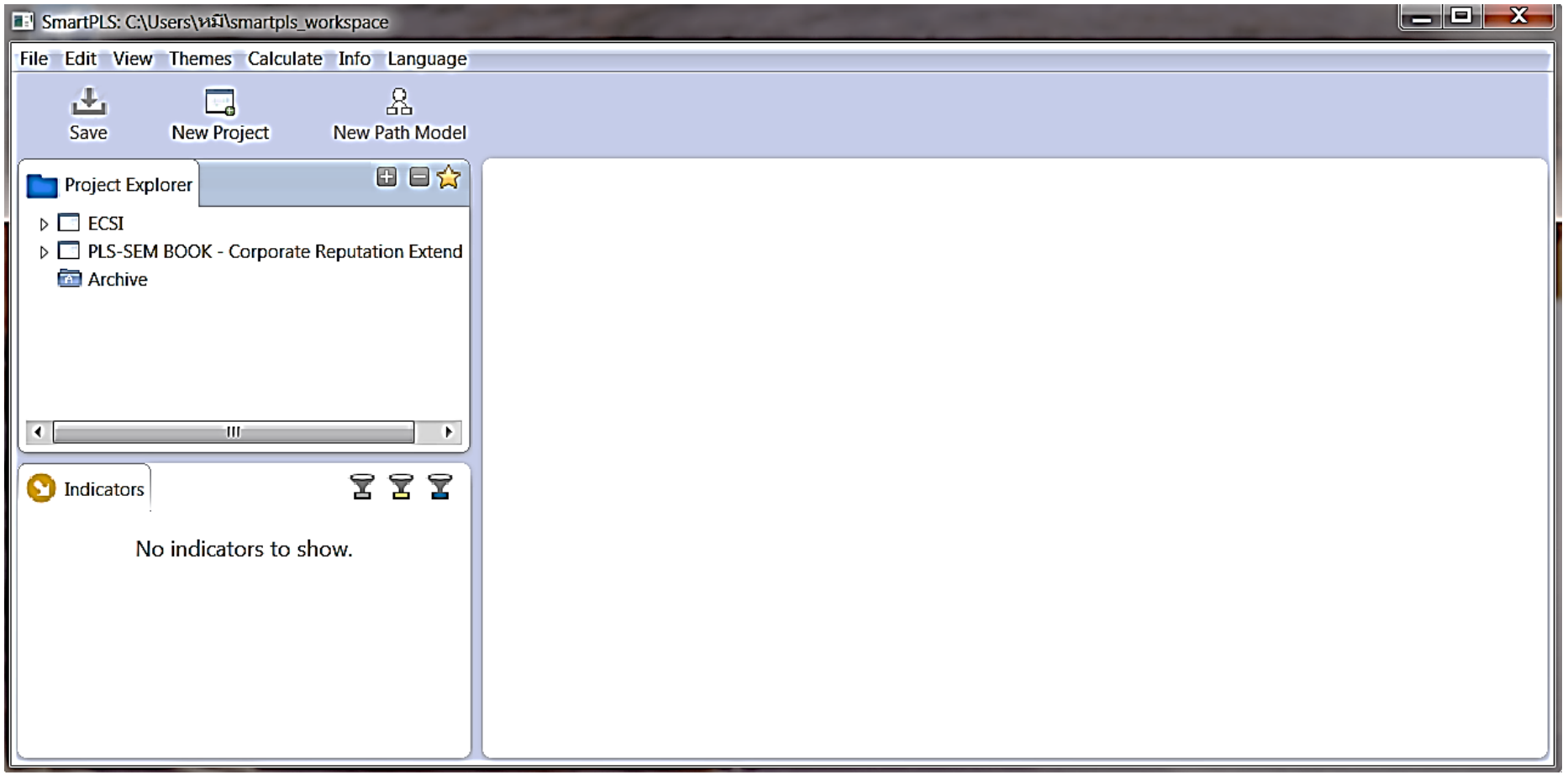
Redundancy criterion

Redundancy of ith endogenous variable	ตัวแปรต้นทางสามารถคาดคะเนความผันแปรตัวชี้วัดของตัวแปรผลลัพธ์ที่ i ได้
$Redundancy_i = Com_i * R_i^2 = 0.50 * 0.19 = 0.095$	ต่ำ
$Redundancy_i = Com_i * R_i^2 = 0.50 * 0.35 = 0.175$	ปานกลาง
$Redundancy_i = Com_i * R_i^2 = 0.50 * 0.67 = 0.335$	สูง

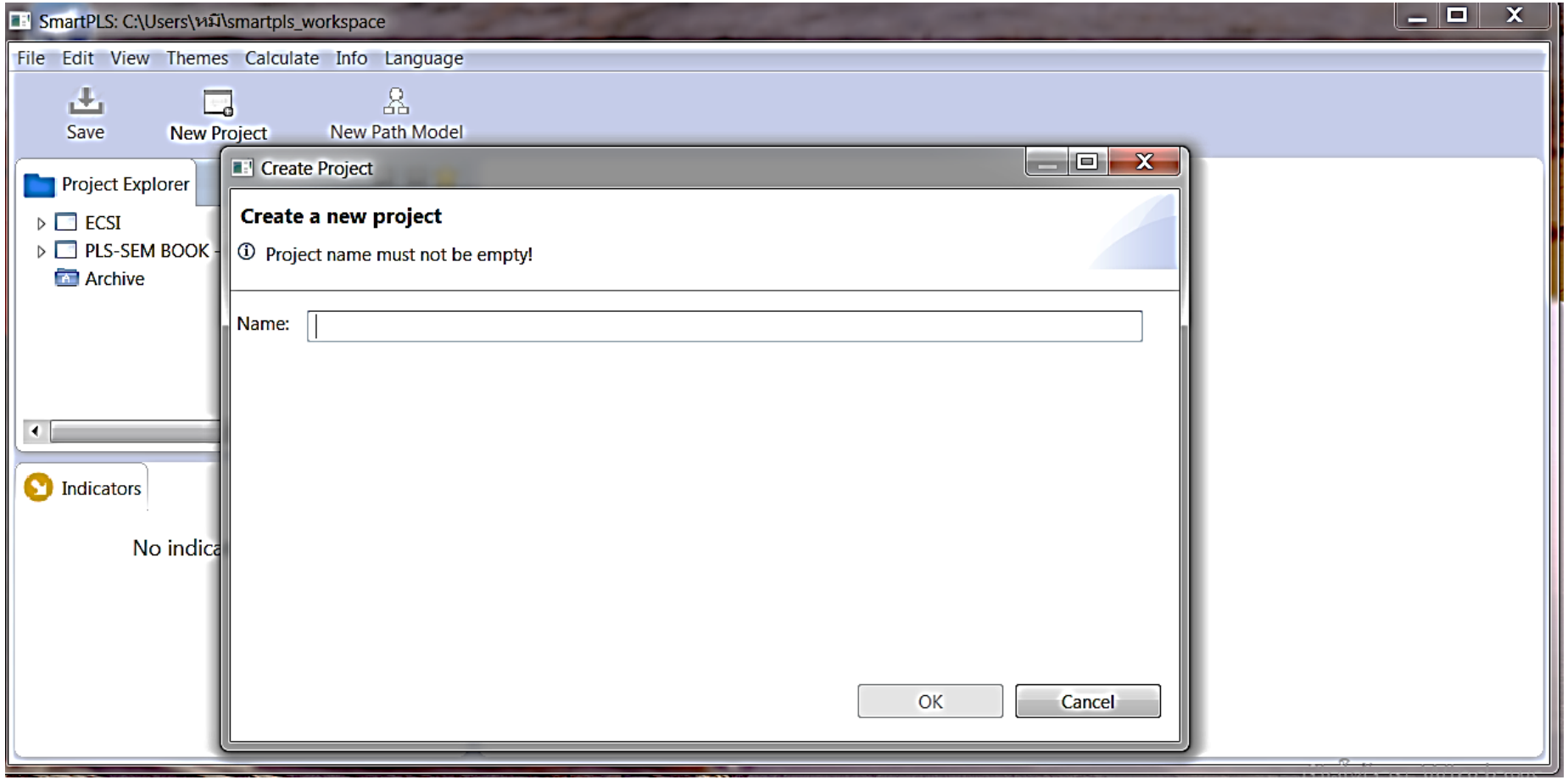
หมายเหตุ Redundancy และ GoF ไม่มีแสดงใน smart PLS3

การรันโปรแกรม Smart PLS

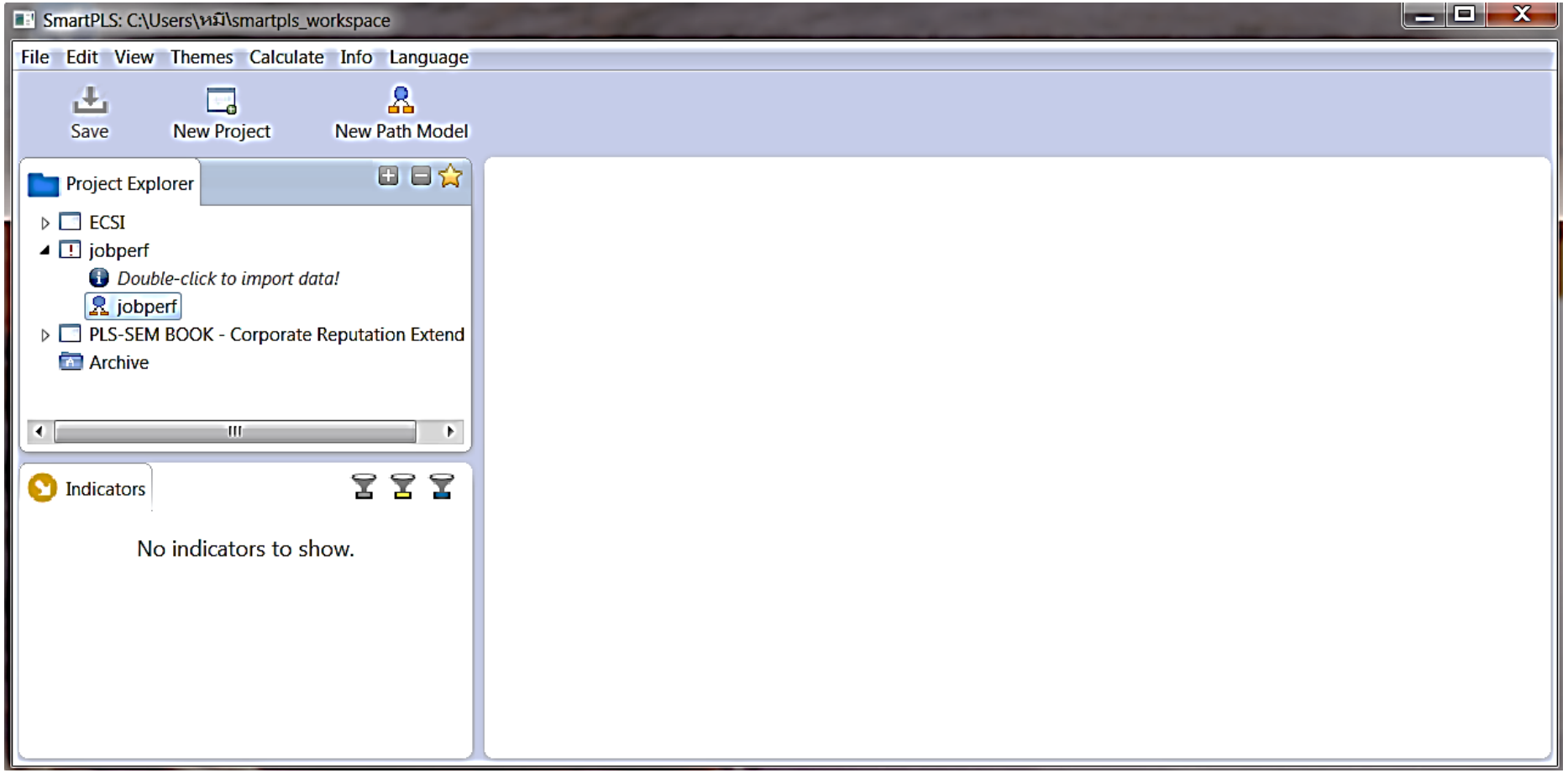
คลิกที่ไอคอนซอฟท์แวร์ จะพบหน้าซอฟท์แวร์ มี 3 หน้าต่าง



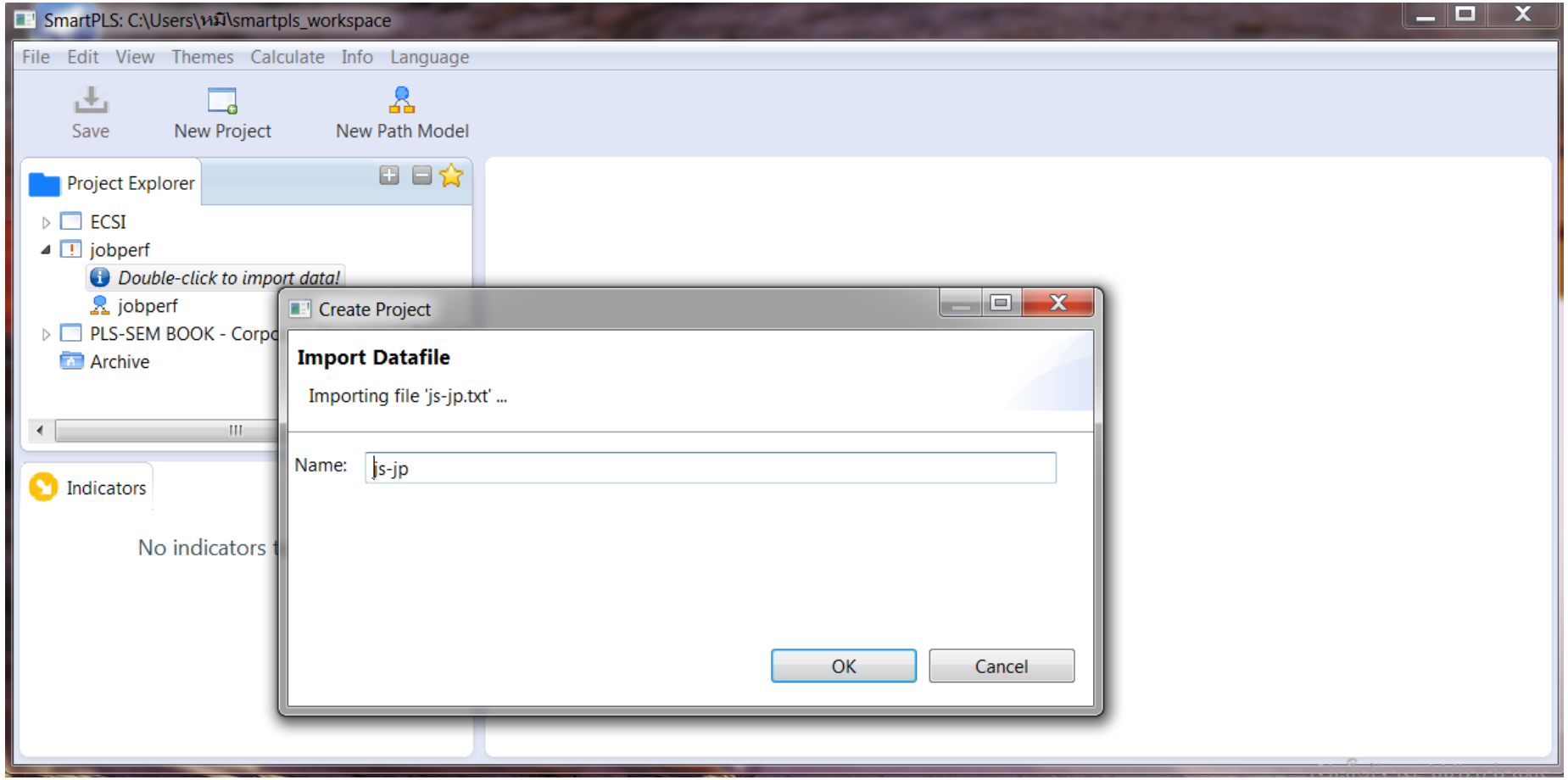
กด New Project > พิมพ์ชื่อโครงการ



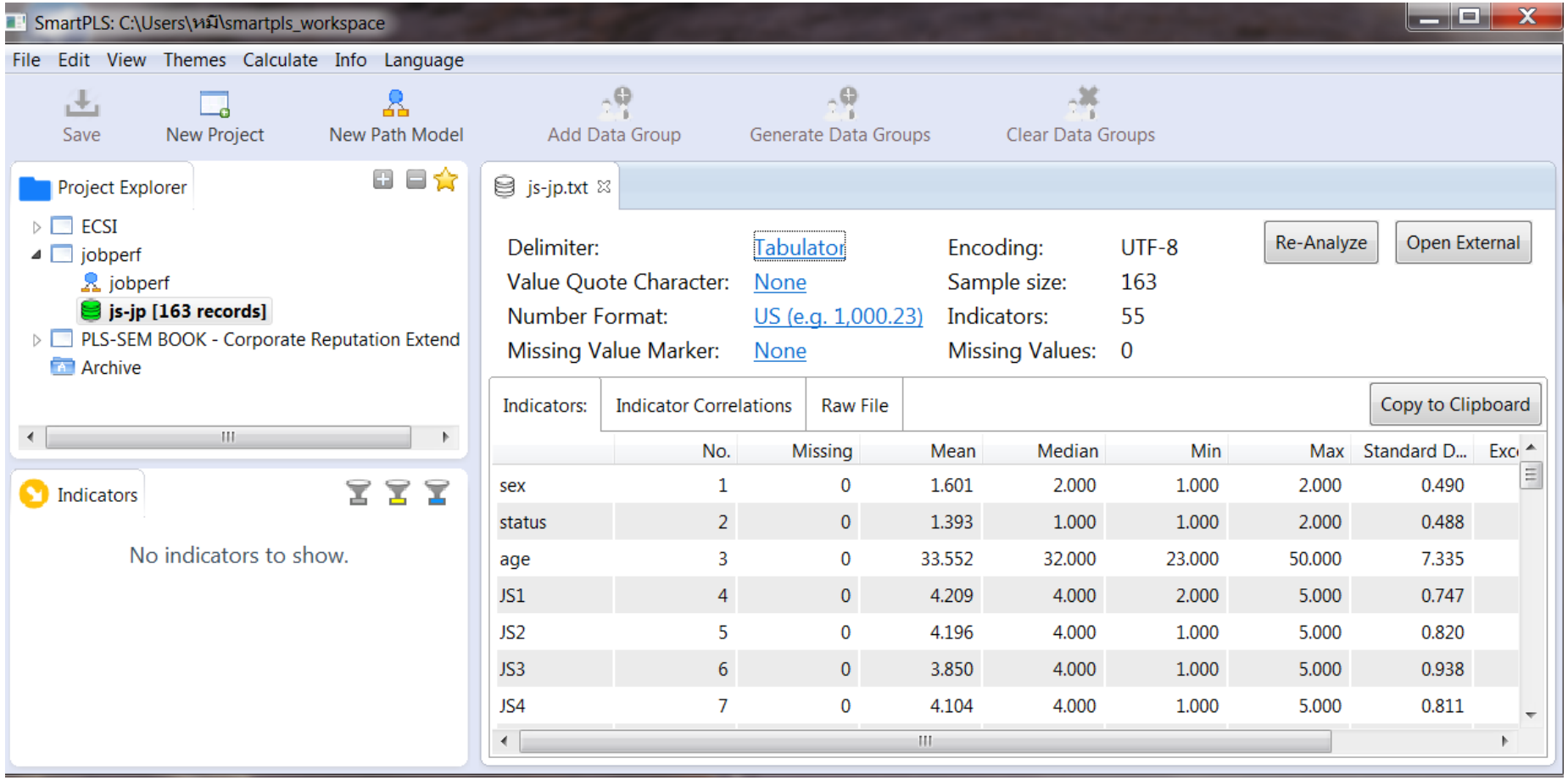
มี 2 ไอคอนเพิ่มเข้ามาคือ *Double click to import data* กับ ไอคอนชื่อโปรเจ็ค



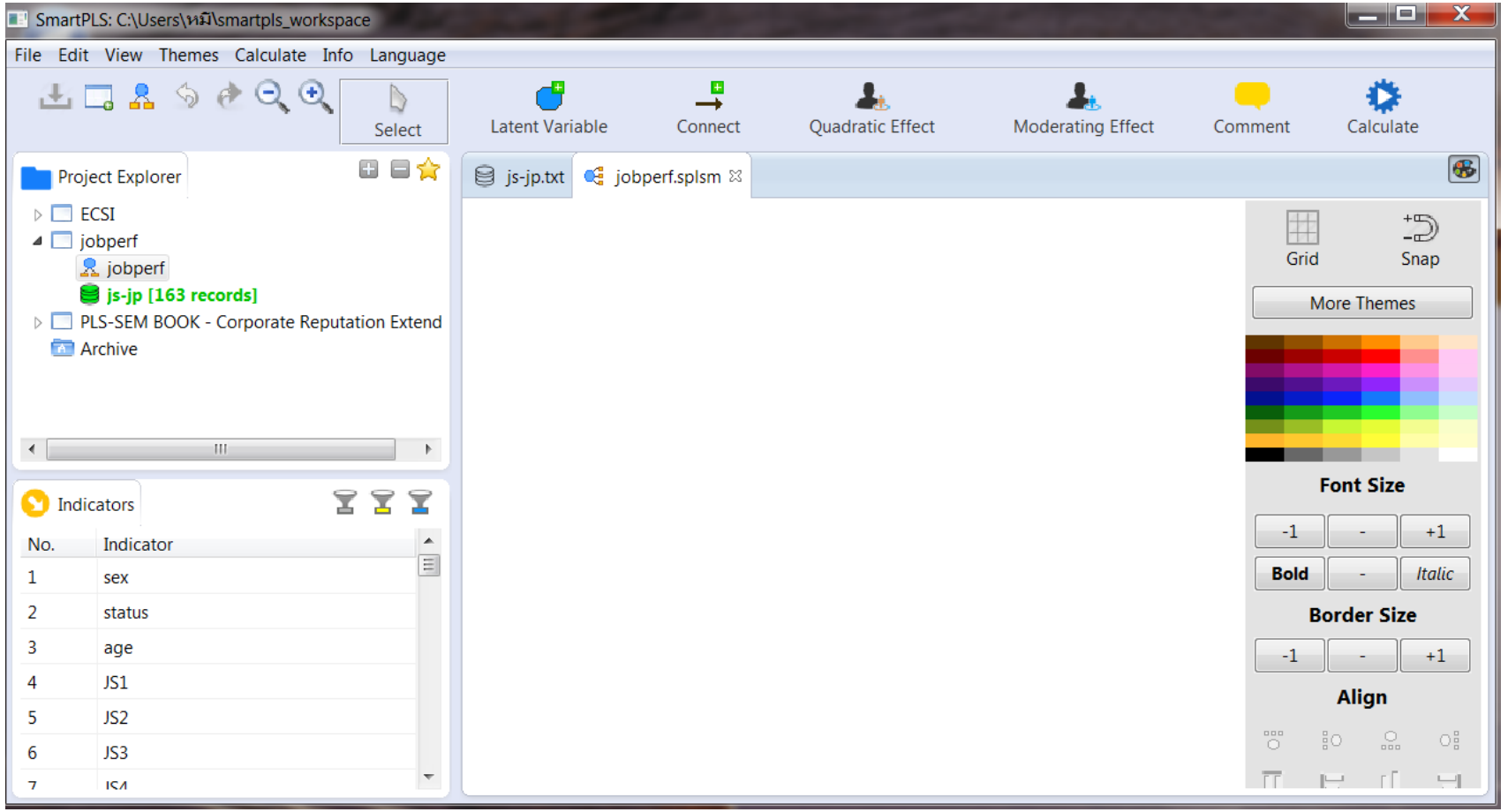
Dbl click ที่ไอคอน double click to import data > ใส่ชื่อแฟ้มข้อมูล



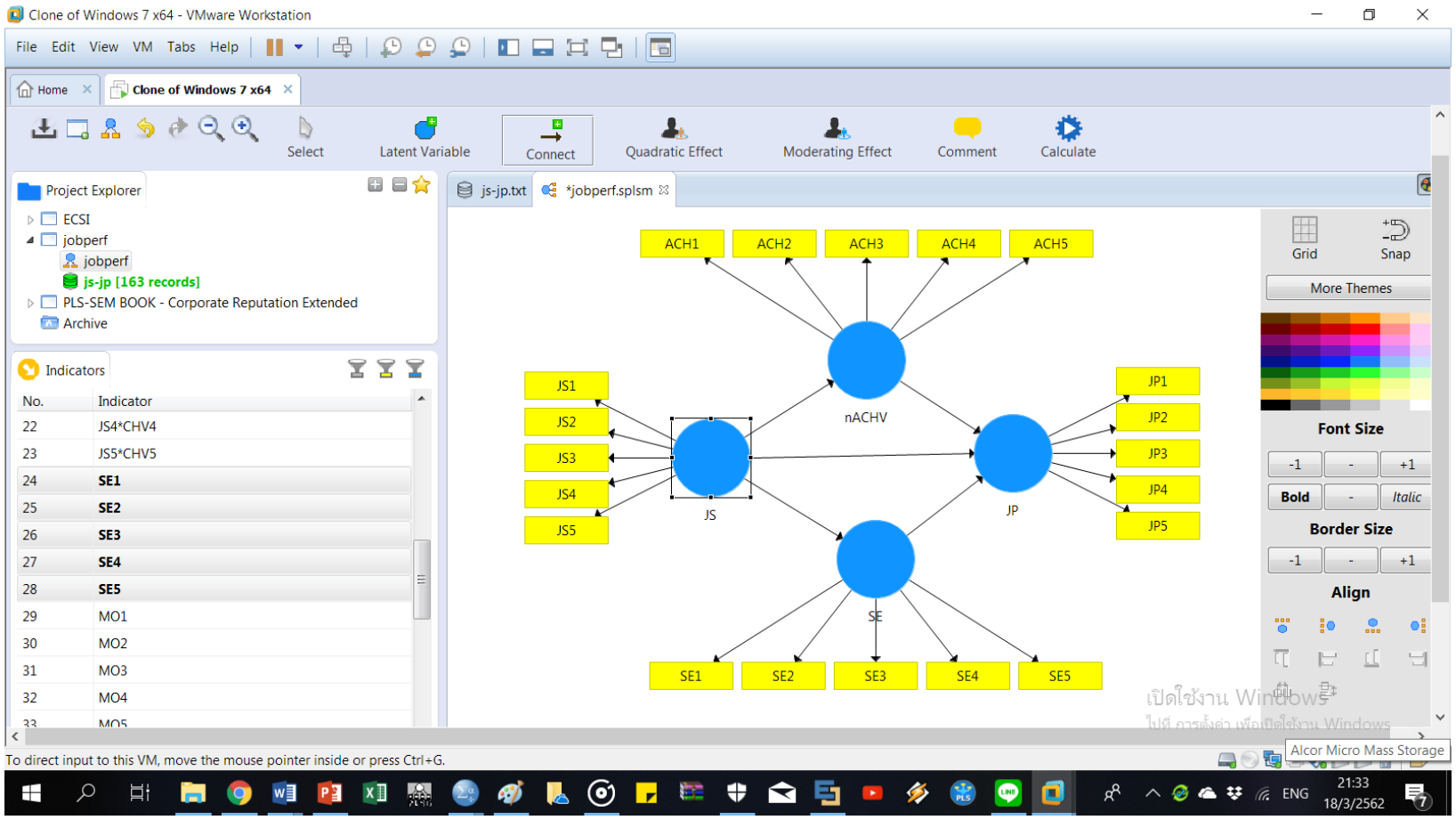
คลิกไอคอนชื่อแฟ้มข้อมูลจะแสดงข้อมูล ถ้าไม่แสดงแปลว่าแฟ้มข้อมูลมีปัญหา missing data หรือมีข้อมูลผิดปกติ เช่นมีข้อความปนอยู่



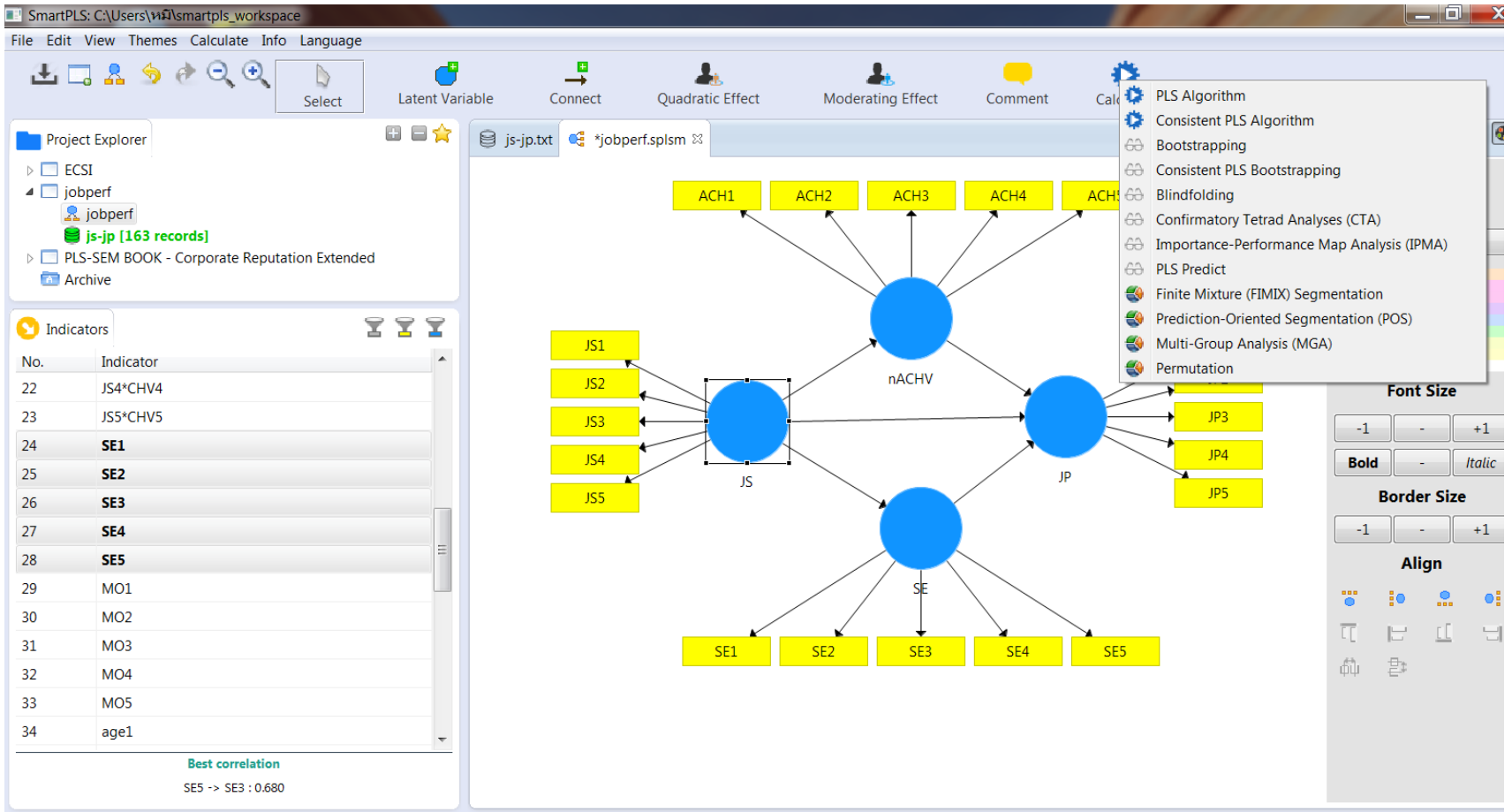
คลิกไอคอนชื่อโปรเจ็ค ข้อมูลหน้าต่างที่ 2 จะแสดงเป็นชื่อตัวแปรและหน้าต่างที่ 3 พร้อมใช้งาน ถ้าหน้าต่างที่ 2 ไม่แสดงอะไรแปลว่ามีปัญหาเรื่องข้อมูลหรือไลเซนต์



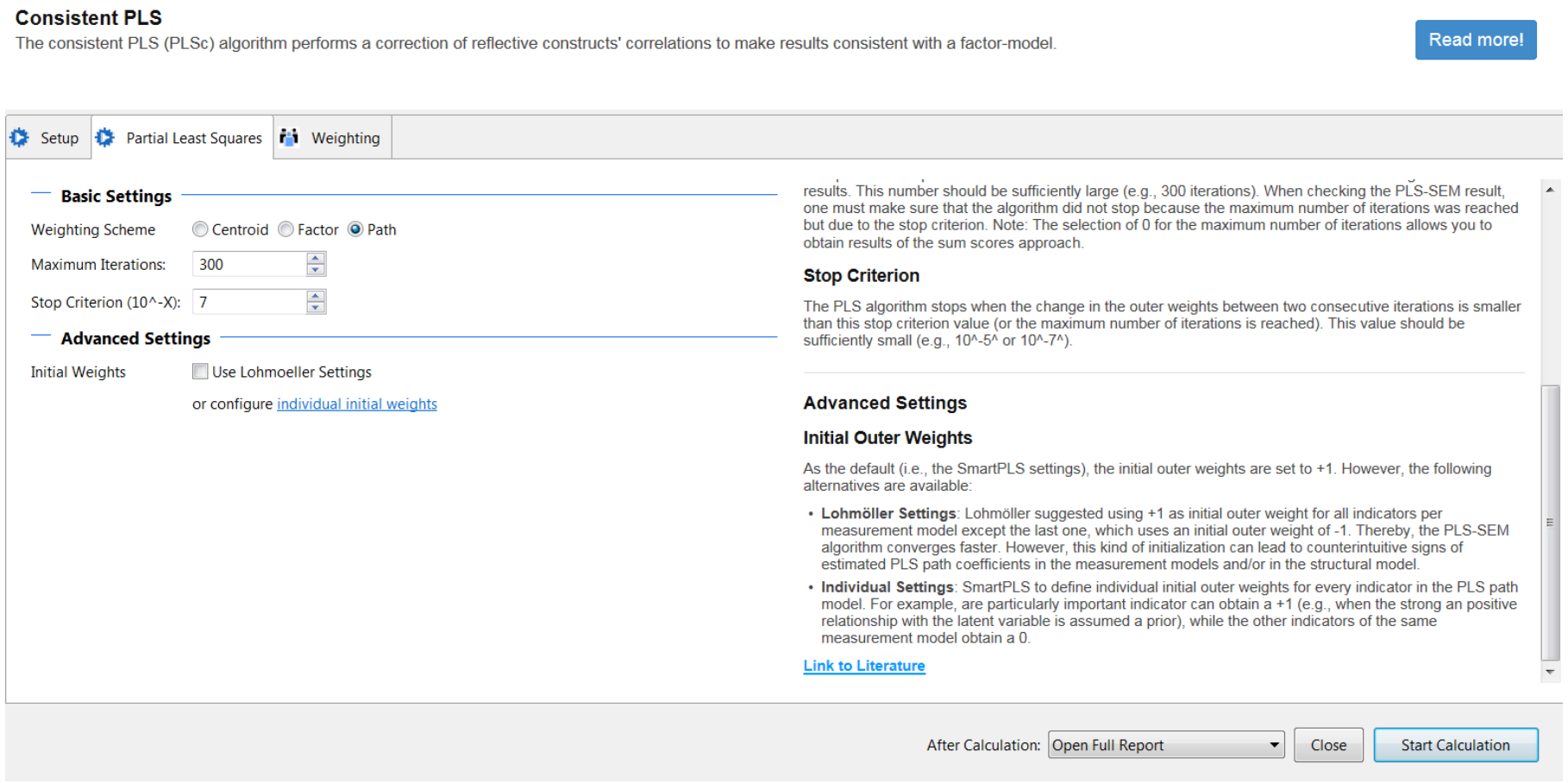
Shift click ตัวชี้วัดของตัวแปรในหน้าต่างที่ 2 แล้วลากไปวางในหน้าต่างที่ 3 เชื่อมโยง และตั้งชื่อตัวแปรแฝง



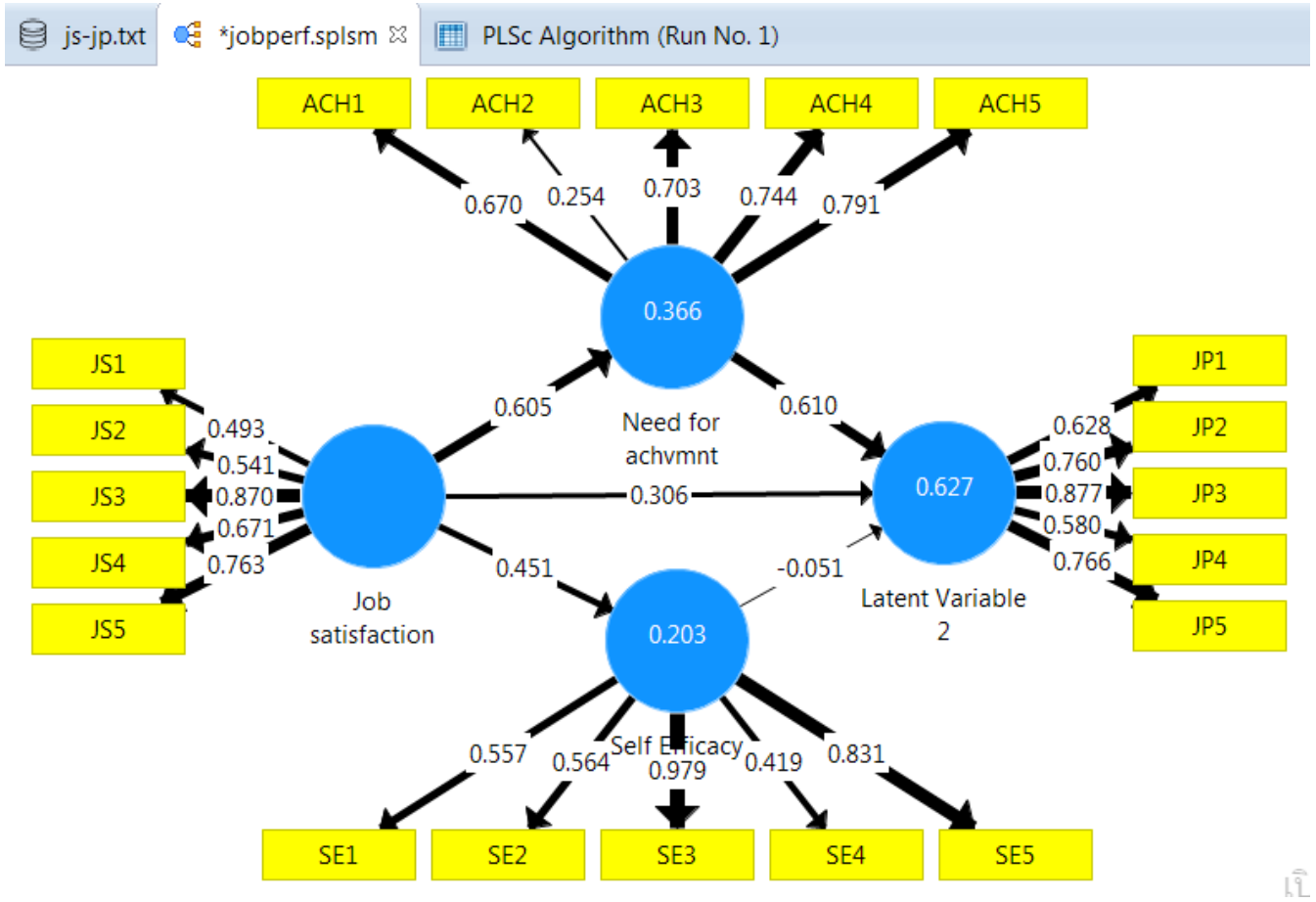
กด calculation แนะนำให้เลือกใช้ Consistent PLS Algorithm



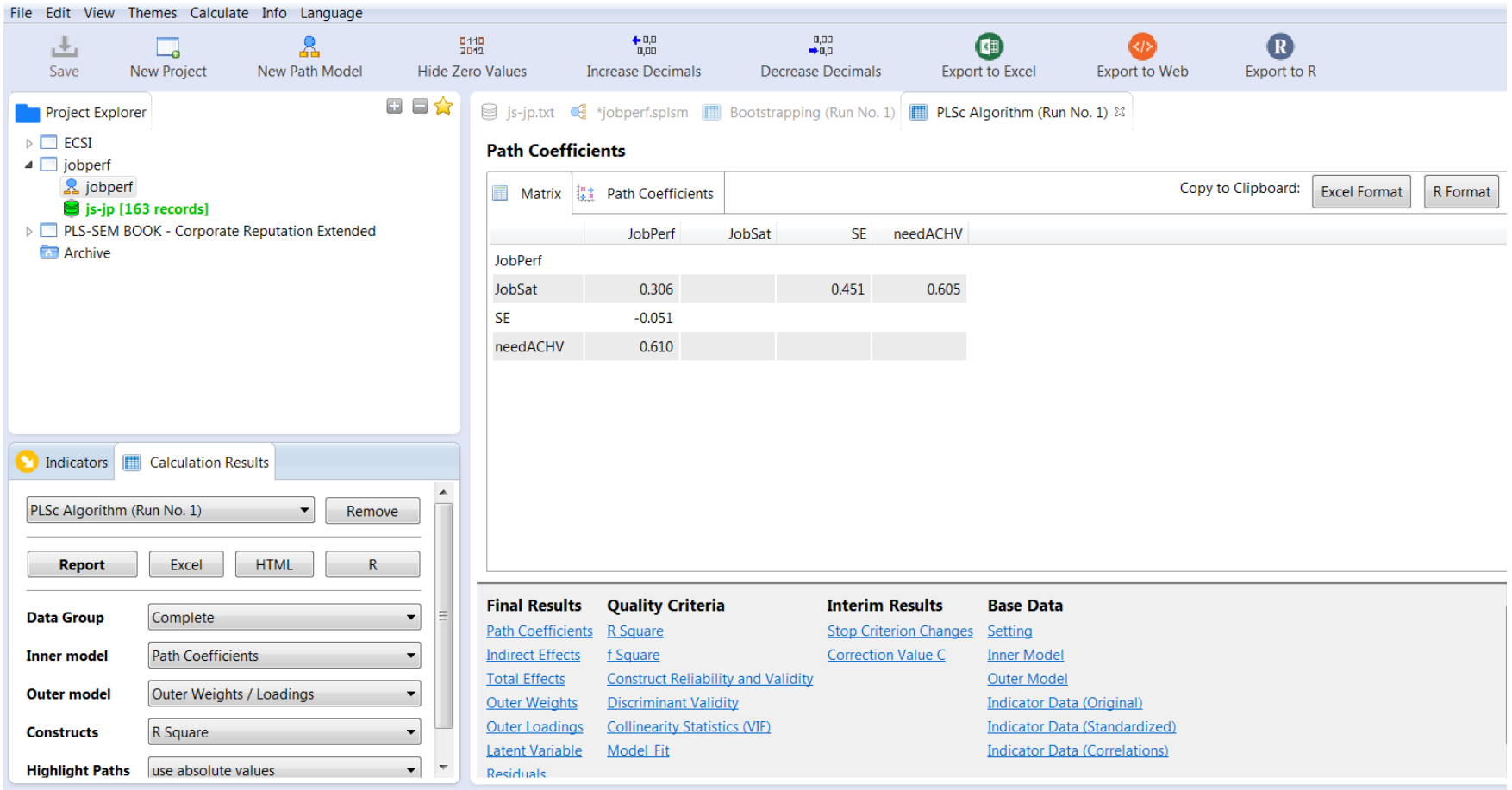
ไดอะล็อก setting เปิดขึ้นให้กำหนดเอง ถ้าไม่ประสงค์จะกำหนดก็ปล่อยให้ทำงานตาม default แต่ควรกำหนดจำนวนการวนหาคำตอบเป็น 5,000-10,000 รอบ



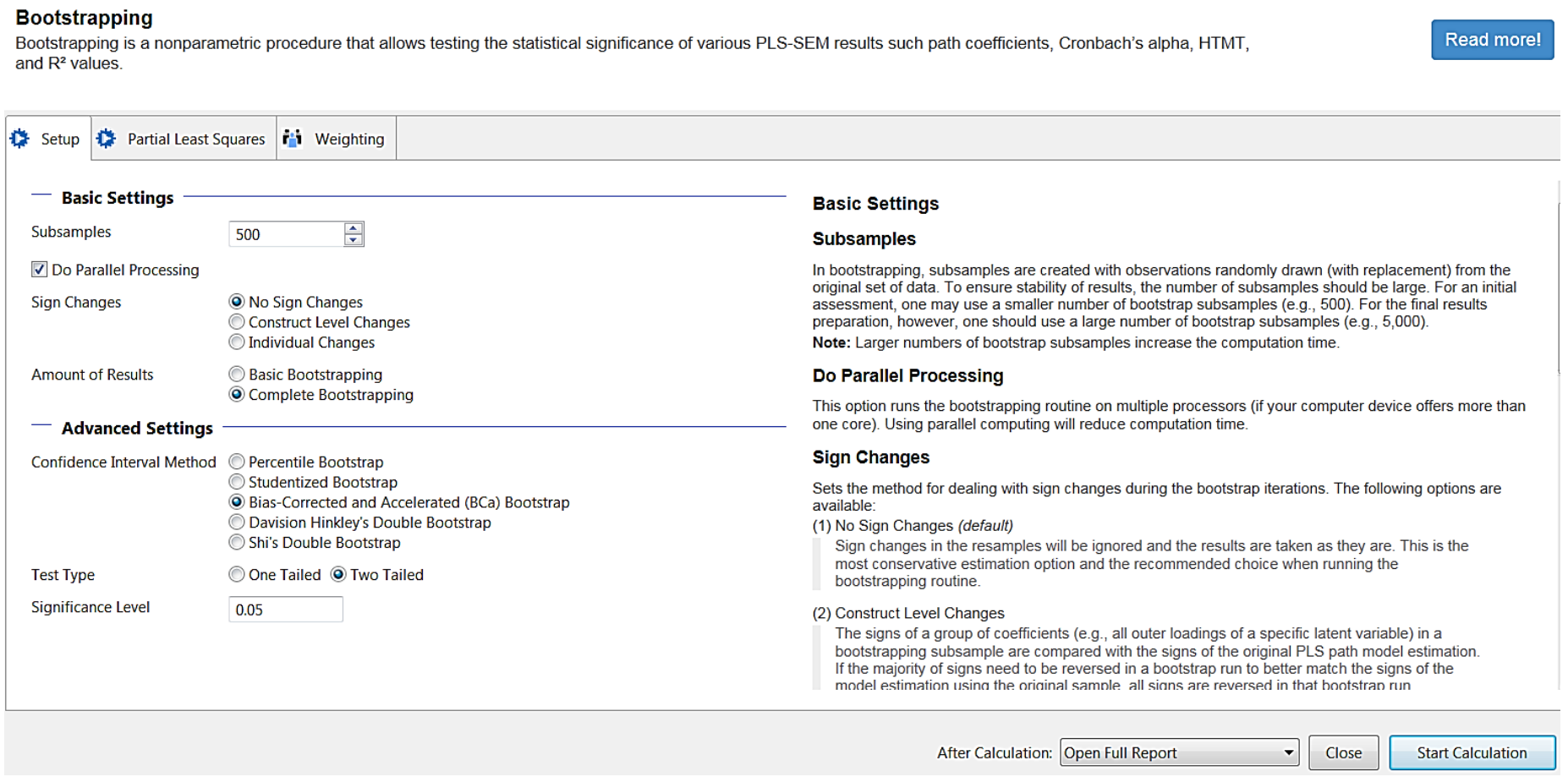
กด start calculation จากนี้ให้ดูผลจากภาพและ report ว่าตรงตามเกณฑ์คุณภาพหรือไม่ ถ้าไม่ตรงเกณฑ์ต้องปรับแก้ให้ตรงเกณฑ์และตรงตามวรรณกรรม ผลการรัน consistent PLS Algorithm ปรากฏดังนี้



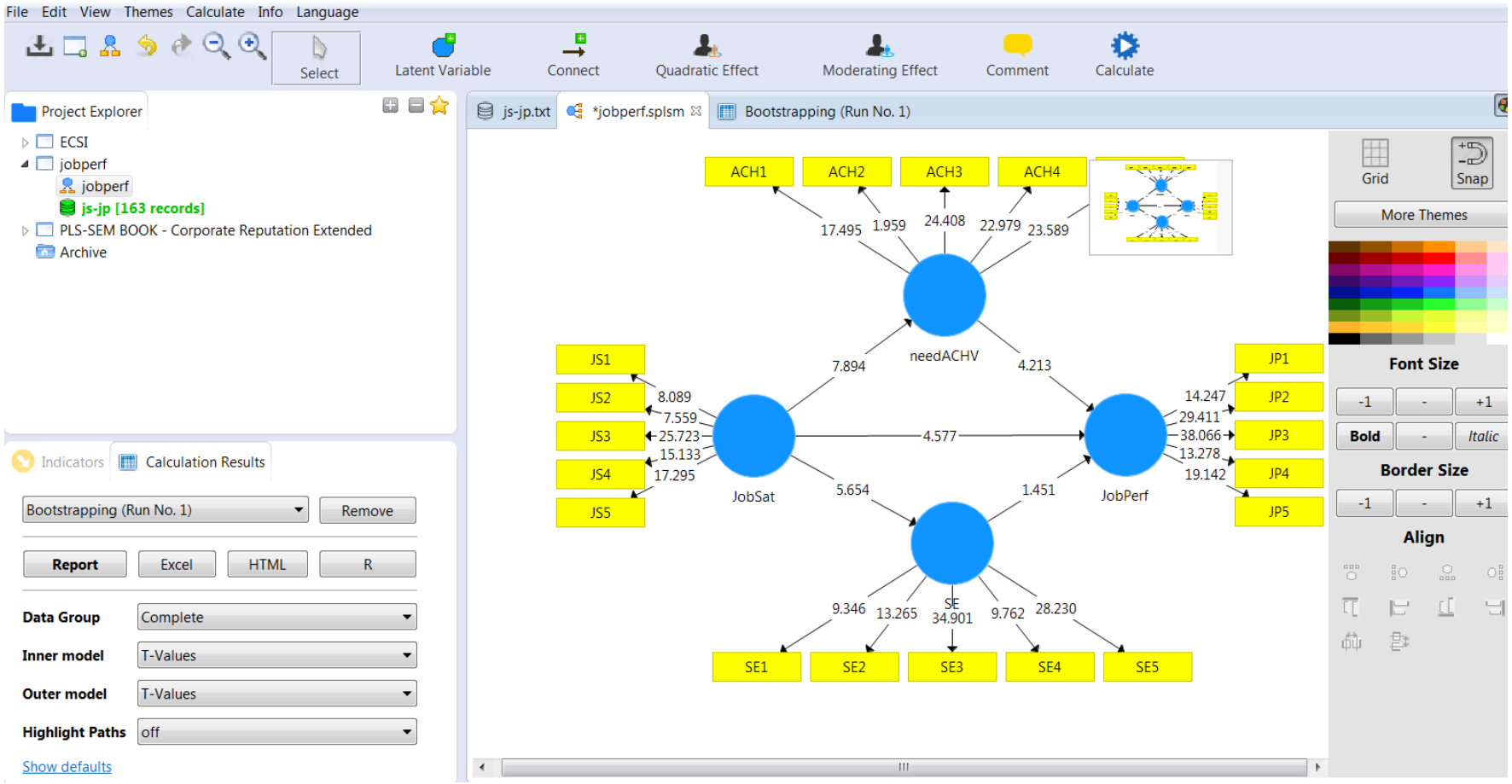
ผลการรัน consistent PLS Algorithm



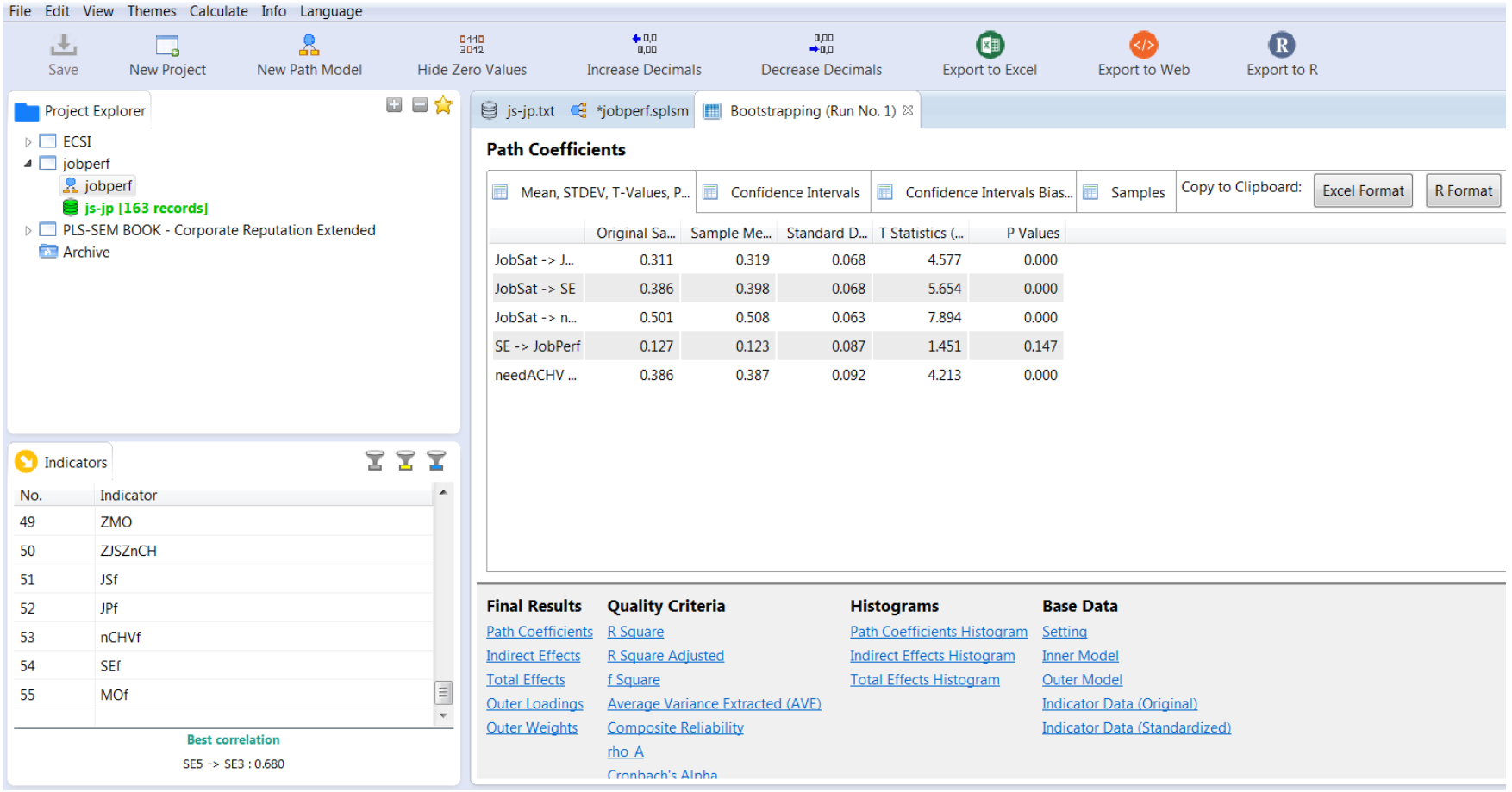
กด calculation เลือก Bootstrapping



ผลการรัน Bootstrapping



ผลการรัน Bootstrapping



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.