

SỬ DỤNG PHƯƠNG PHÁP PHÂN TÍCH DỮ LIỆU PLS – SEM TRONG KIỂM ĐỊNH CÁC MÔ HÌNH NGHIÊN CỨU KHOA HỌC

MAI ANH VŨ

PLS-SEM là mô hình phương trình cấu trúc bình phương tối thiểu riêng phần. PLS-SEM ước lượng các tham số của mô hình dựa trên ma trận phương sai. Đây là phương pháp được ưa thích đối với nhà nghiên cứu khi mục đích nghiên cứu của họ là phát triển lý thuyết và giải thích sự thay đổi của biến phụ thuộc. Hiện nay, kỹ thuật phân tích thể hệ thứ hai (PLS-SEM) đã có những phần mềm được phát triển và sử dụng một cách hiệu quả. Rất nhiều nhà nghiên cứu lựa chọn PLS-SEM khi mô hình nghiên cứu có chứa biến tiềm ẩn được đo lường qua các biến quan sát phản ánh nguyên nhân, cỡ mẫu nhỏ và dữ liệu phân phối không chuẩn.

Từ khóa: PLS-SEM, Partial Least Square, cấu trúc, kiểm định, mô hình, nghiên cứu

USING PLS – SEM DATA ANALYSIS METHOD IN TESTING THE SCIENTIFIC RESEARCH MODELS

Mai Anh Vu

PLS-SEM is a partial least squares structure equation model. PLS-SEM estimates the parameters of the model based on the variance matrix. This is the preferred method for researchers when the purpose of their research is to develop the theory and explain the change of the dependent variable. Currently, the second generation analysis technique (PLS-SEM) has already been developed and used effectively. Many researchers choose PLS-SEM when the research model contains latent variables that are measured through observational variables that reflect the cause, small sample size, and non-standard distribution data.

Keywords: PLS-SEM, Partial Least Square, structure, testing, model, research

Ngày nhận bài: 2/01/2020

Ngày hoàn thiện biên tập: 15/01/2020

Ngày duyệt đăng: 29/1/2020

Giới thiệu

PLS-SEM là một phương pháp phân tích dữ liệu đa biến thể hệ thứ hai. Phân tích dữ liệu đa biến liên quan đến ứng dụng của phương pháp thống kê, đồng thời phân tích đa biến liên quan đến các cá nhân, các tổ chức, các sự kiện, các hoạt động, các tình huống... SEM được

sử dụng để khám phá hoặc khẳng định lý thuyết. Mô hình khám phá liên quan đến phát triển lý thuyết trong khi mô hình khẳng định kiểm định lý thuyết. Có 2 loại SEM, một là dựa vào hiệp phương sai và còn lại dựa vào phương sai. CB-SEM được sử dụng để khẳng định (hoặc bác bỏ) lý thuyết. Mô hình cấu trúc tuyến tính dựa vào phương sai (PLS-SEM) thường được sử dụng cho những nghiên cứu khám phá và để phát triển lý thuyết.

Như một thay thế cho CB-SEM, PLS-SEM nhấn mạnh các mục tiêu dự báo trong khi tương đồng về các nhu cầu cần thiết dữ liệu và đặc điểm của các mối quan hệ. PLS-SEM tối đa phương sai được giải thích của biến tiềm ẩn nội sinh bằng việc ước lượng các mối quan hệ mô hình riêng phần trong chuỗi lặp lại của hồi quy OLS. Ngược lại, CB-SEM ước lượng các tham số mô hình mà không có sự nhất quán giữa ước lượng và ma trận hiệp phương sai là nhỏ nhất. Thay vì tuân theo logic mô hình các yếu tố chung như CB-SEM, PLS-SEM tính toán các hợp nhất của các biến quan sát dùng làm đại diện cho các khái niệm trong nghiên cứu. PLS-SEM không bị hạn chế bởi vấn đề nhận dạng, ngay cả khi mô hình trở nên phức tạp - một tình huống mà thông thường hạn chế sử dụng CB-SEM - và không yêu cầu trong hầu hết các giả định phân phối. Hơn nữa, PLS-SEM có thể xử lý tốt hơn mô hình đo lường nguyên nhân và có lợi thế khi mẫu tương đối nhỏ.

Dữ liệu yêu cầu của PLS-SEM và các đặc điểm chính

Đặc điểm dữ liệu yêu cầu của PLS-SEM

Đặc điểm dữ liệu như là kích thước mẫu tối thiểu, dữ liệu không chuẩn và thang đo lường (tức là sử dụng các loại thang đo khác nhau) được xem như là lý do bất



BẢNG 1: ĐẶC ĐIỂM CHÍNH CỦA PLS-SEM

Những đặc điểm về mô hình

Số lượng biến quan sát trong từng mô hình đo lường khái niệm nghiên cứu	Đo lường khái niệm với đo lường đơn biến và đa biến
Các mối quan hệ giữa các khái niệm nghiên cứu và các biến quan sát của chúng	Dễ dàng kết hợp các mô hình đo lường nguyên nhân và kết quả.
Độ phức tạp mô hình	Xử lý các mô hình phức tạp với nhiều mối quan hệ mô hình cấu trúc
Thiết lập mô hình	Không vòng lặp nhân quả nào được cho phép trong mô hình cấu trúc
Số lượng biến quan sát trong từng mô hình đo lường khái niệm nghiên cứu	Đo lường khái niệm với đo lường đơn biến và đa biến

Đặc tính thuật toán của PLS-SEM

Mục tiêu	Giảm thiểu phương sai không giải thích được (tức là, tối đa giá trị R ²)
Hiệu quả	Hội tụ sau khi lặp đi lặp lại (thậm chí trong các tình huống với mô hình phức tạp và/ hoặc tập hợp dữ liệu lớn) để đưa ra giải pháp tối ưu; thuật toán hiệu quả
Bản chất của khái niệm nghiên cứu	Được xem như là các đại diện của khái niệm tiềm ẩn được điều tra, được đại diện bởi các biến tổng hợp Ước lượng các tổ hợp tuyến tính của các biến quan sát Được xác định
Điểm số của khái niệm	Sử dụng cho mục đích dự đoán Có thể sử dụng như đầu vào cho phân tích tiếp theo Không ảnh hưởng bởi sự thiếu dữ liệu Các quan hệ của mô hình cấu trúc nhìn chung đều bị đánh giá thấp và các mối quan hệ của mô hình đo lường nhìn chung được đánh giá cao khi ước lượng dữ liệu từ những mô hình yếu tố chung
Ước lượng tham số	Tính nhất quán cao Độ nhạy thống kê cao

Vấn đề đánh giá mô hình

Sự đánh giá mô hình tổng thể	Không có tiêu chí độ phù hợp mô hình Mô hình đo lường kết quả: độ tin cậy và giá trị được đánh giá thông qua nhiều tiêu chí
Sự đánh giá mô hình đo lường	Mô hình đo lường nguyên nhân: đánh giá giá trị, mức ý nghĩa và sự liên quan của các trọng số quan sát, sự đa cộng tuyến
Sự đánh giá mô hình cấu trúc	Sự đa cộng tuyến giữa các tập khái niệm, mức ý nghĩa của hệ số đường dẫn, tiêu chí để đánh giá khả năng dự đoán của mô hình

Nguồn: Joe F. Hair, Christian M. Ringle & Marko Sarstedt (2011) PLS-SEM: Indeed a Silver Bullet, *Journal of Marketing Theory and Practice*, 19:2, 139152, DOI: 10.2753/JMP10696679190202

đầu thông thường nhất trong việc áp dụng PLS-SEM (Hair, Sarstedt, Ringle và cộng sự, 2012; Henseler và cộng sự, 2009). Trong khi một vài lập luận nhất quán với khả năng của phương pháp, phần còn lại thì không. Ví dụ: Kích thước mẫu nhỏ có lẽ thường bị lạm dụng nhất trong tranh luận với các nhà nghiên cứu sử dụng PLS-SEM với kích cỡ mẫu nhỏ không thể chấp nhận được (Goodhue, Lewis, và Thompson, 2012; Marcoulides và Saunders, 2006). Các nhà nghiên cứu cho rằng, có một “ma thuật” nào đó trong phương pháp PLS-SEM cho phép họ sử dụng một mẫu rất nhỏ (ví dụ dưới 100)

để có được các kết quả đại diện cho các tác động tồn tại trong một tổng thể của một vài triệu yếu tố hoặc cá nhân.

Một số nhà nghiên cứu tin rằng, kích thước mẫu nhỏ không đóng vai trò đặc biệt trong ứng dụng PLS-SEM. Ý kiến này được thúc đẩy bởi quy tắc 10 lần (Barclay, Higgins, và Thompson, 1995), cho rằng, kích thước mẫu nên bằng hoặc lớn hơn: Mười (10) lần số lớn nhất của các biến quan sát nguyên nhân được sử dụng để đo lường khái niệm đơn; hoặc mười (10) lần số lớn nhất của đường dẫn cấu trúc hướng vào một khái niệm riêng biệt trong mô hình cấu trúc.

Trong khi quy định 10 lần đề xuất chỉ dẫn sơ bộ về yêu cầu kích thước mẫu tối thiểu, PLS-SEM cũng như các kỹ thuật thống kê khác yêu cầu nhà nghiên cứu cân nhắc lại kích thước mẫu với nền tảng mô hình và đặc điểm dữ liệu (Hair, Ringle, và Sarstedt, 2011; Marcoulides và Chin, 2012). Đặc biệt, kích thước mẫu cần thiết nên được xác định bởi phân tích độ nhạy dựa trên một phần mô hình với số lượng lớn nhất của biến dự báo.

Đặc điểm chính của PLS-SEM

Ngoài đặc điểm đã nêu về dữ liệu, PLS-SEM còn có một số đặc điểm chính được tác giả nghiên cứu và tổng hợp (Bảng 1).

Một số chỉ số kiểm định trong PLS-SEM

Việc kiểm định các nghiên cứu khoa học trong PLS-SEM được thực hiện bởi rất nhiều chỉ số. Một số chỉ số kiểm định cần lưu ý khi sử dụng PLS-SEM như sau:

- Hệ số tải mô hình (Model Loading): Về nguyên tắc, hệ số tải càng gần giá trị 1 càng cho thấy, độ tin cậy của



biến tiềm ẩn. Hệ số tải $\geq 0,7$ được coi là chấp nhận được (Henseler, Ringle & Sarstedt, 2012).

- Hệ số Composite Reliability: Hệ số này cho biết, độ tin cậy của thang đo khi sử dụng với kỹ thuật PLS-SEM. Hệ số Composite Reliability biến thiên từ 0 đến 1, giá trị càng gần 1 cho thấy, độ tin cậy trong mô hình PLS-SEM càng cao. Trong một mô hình có tính chất khám phá, nếu hệ số này $\geq 0,6$ là chấp nhận được (Höck & Ringle, 2006) và nếu trong trường hợp mô hình khẳng định, hệ số $\geq 0,7$ là phù hợp (Henseler, Ringle, & Sarstedt, 2012).

- Hệ số Average Variance Extracted (AVE): Hệ số này kiểm tra độ hội tụ và phân tán của mô hình. Một mô hình tốt cần có hệ số AVE $\geq 0,5$ (Höck & Ringle, 2006).

- Chỉ số Standardized Root Mean Square Residual (SRMR): Chỉ số này cho biết, mức độ phù hợp của mô hình nghiên cứu. Theo Hu & Bentler (1998), thông thường một mô hình phù hợp sẽ có giá trị SRMR nhỏ hơn 0,08.

- Chỉ số Cross loading và intended loading: Đây là 2 chỉ số cho biết hệ số tải của nhân tố trong mô hình và tương quan với các nhân tố khác. Theo đó, chỉ số Intended Loading của một nhân tố nên lớn hơn 0,7 và chỉ số Cross loading nên nhỏ hơn 0,3.

- Chỉ số Variance Inflation Factor (VIF): Chỉ số cho biết khả năng xảy ra trường hợp đa cộng tuyến trong mô hình. Chỉ số VIF < 10 có thể chấp nhận được; tuy nhiên, để đảm bảo độ tin cậy chỉ số VIF không được lớn hơn 5 (Hair & Cộng sự (2011)).

- Đo lường hệ số tổng thể xác định (R-square value), là một chỉ số để đo lường mức độ phù hợp với mô hình của dữ liệu (khả năng giải thích của mô hình).

- Đánh giá giá trị phân biệt: Có 2 chỉ số để đánh giá giá trị phân biệt:

+ Vùng điều kiện của Fornell và Larcker (1981) so sánh căn bậc hai của AVE của mỗi khái niệm với tương quan (Pearson) giữa khái niệm hay biến tiềm ẩn. Căn bậc hai của AVE nên dao động hơn tương quan những khái niệm khác.

+ Hay một cách khác để đánh giá giá trị phân biệt được các nhà nghiên cứu đề xuất là heterotrait – monotrait – HTMT. HTMT là trung bình của tất cả các mối tương quan của các biến quan sát của từng biến nghiên cứu với biến nghiên cứu khác. Hệ số HTMT lớn hơn 0,9 chứng tỏ 2 biến nghiên cứu thiếu giá trị phân biệt, ngưỡng chấp nhận phải thấp hơn 0,85.

- PLS Bootstrapping: Phân tích Bootstrapping được sử dụng để loại bỏ sai số chuẩn và kiểm chứng mức độ ý nghĩa của mô hình PLS ở mức ý nghĩa 5%. Ở mức độ khám phá, số lần Bootstrapping có thể ở mức 500 lần. Nhưng trong giai đoạn phân tích hoàn chỉnh, số lần Bootstrapping cần phải được tăng lên.

- Chỉ số Inner Model p-value và Outer Model p-value: Giá trị p-value phải nhỏ hơn 0,05, giá trị T-Value lớn hơn 1,96

Khi sử dụng PLS-SEM để kiểm định thì một giả thuyết nghiên cứu ngoài tính đúng đắn của dữ liệu nghiên cứu từ định tính tới khảo sát còn cần thỏa mãn các chỉ số kiểm định vừa nêu trên. Nếu thỏa mãn thì giả thuyết được chấp nhận và khẳng định trong mô hình nghiên cứu.

Kết luận

Theo Rasoolimanesh và cộng sự (2018), số ứng dụng PLS-SEM trong nghiên cứu đã tăng đáng kể trong vài năm qua. Bằng việc sử dụng PLS - SEM kết quả nghiên cứu đã khẳng định sự hợp lý của phương pháp phân tích PLS - SEM đặc biệt trong nghiên cứu thực nghiệm đòi hỏi liên kết chặt chẽ với nhau trong quá trình chuyển đổi từ một lý thuyết sang một mô hình cấu trúc tuyến tính. PLS-SEM thể hiện nhiều ưu điểm vượt trội hơn so với CB-SEM và các kỹ thuật phân tích thể hệ thứ nhất khác. Các nhà nghiên cứu nên lựa chọn PLS-SEM khi mô hình nghiên cứu có chứa biến tiềm ẩn được đo lường qua các biến quan sát phản ánh nguyên nhân, cỡ mẫu nhỏ và dữ liệu phân phối không chuẩn. Mô hình phân tích phức tạp cần nhiều kỹ thuật phân tích nâng cao với thang đo một hay nhiều cấp thì mô hình PLS-SEM được sử dụng rất phù hợp. Tuy nhiên, PLS-SEM vẫn còn mới đối với nhiều nhà nghiên cứu đặc biệt là đối với các nhà nghiên cứu Việt Nam. Thông qua nghiên cứu này có thể khẳng định, PLS-SEM là một phương pháp dựa trên phương sai ước tính các khái niệm đại diện cho các biến tiềm ẩn trong các mô hình liên kết.

Tài liệu tham khảo:

1. Ali, F., Rasoolimanesh, S. M., Sarstedt, M., Ringle, C. M., Ryu, K., do Valle, P. O., ... & Rasoolimanesh, S. M. (2018), *Soft modeling: The basic design and some extensions. In Applying Partial Least Squares in Tourism and Hospitality Research (Vol. 30, No. 1, pp. i-xviii)*. Mahwah, NJ: Springer International Publishing;
2. Fornell, C., and Larcker, D. F. (1981), *Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. Journal of Marketing Research (18:1), pp. 39-50;*
3. Hair, J. F., Hult, G. T. M., Ringle, C., & Sarstedt, M. (2016), *A primer on partial least squares structural equation modeling (PLS-SEM)*. California: Sage Publications;
4. Hair, J. F., Ringle, C. M., & Sarstedt, M. (2013), *Partial least squares structural equation modeling: Rigorous applications, better results and higher acceptance. Long range planning, 46(1-2), 1-12.*

Thông tin tác giả:

NCS., ThS. Mai Anh Vũ

Trường Đại học Văn hóa Thể thao và Du lịch Thanh Hóa

Email: anhvu.dhvh@gmail.com