Tittle of Manuscript: **Data Analysis Procedures with Structural Equation Modelling (SEM): Narrative Literature Review**

1. Proofread document received (November 20th, 2022)
   -Document from proofreading service
2. Submitted to the journal "Open Access Indonesia Journal of Social Sciences (November 25th, 2022)
3. Peer Reviewer results: Revision Required (December 9th, 2022)
4. Revised version received by journal (December 16th, 2022)
5. Paper Accepted for publication (December 19th, 2022)
6. Galley proof (December 21th, 2022)
7. Paper published (December 23th, 2022)

-Document from proofreading service

November 19th, 2022

HM Publisher

Jl Sirnaraga No 99, 8 Ilir, Ilir Timur 3, Palembang, South Sumatra, Indonesia

CONFIDENTIAL

November 19th, 2022

Certificate Service Confirmation

To whom it may concern,

HM Publisher provided comprehensive editing services for manuscript entitled Data Analysis Procedures with Structural Equation Modelling (SEM): Narrative Literature Review. The edit has achieved Grade A: priority publishing; no language polishing required after editing. Should you require any additional information, please do not hesitate to contact me.

Regards,

Khrishna Murti, PhD
Head of Language Institute-HM Publisher
Email: khrishnamurti@gmail.com

Submitted to the journal "Open Access Indonesia Journal of Social Sciences (November 25th, 2022)

Data Analysis Procedures with Structural Equation Modelling (SEM): Narrative Literature Review

Rachmat Hidayat[1*], Patricia Wulandari[2]

[1]Department of Biology, Faculty of Medicine, Universitas Sriwijaya, Palembang, Indonesia

[2]Cattleya Mental Health Center, Palembang, Indonesia

*email: rachmathidayat@fk.unsri.ac.id

Abstract

Metode structural equation modelling (SEM) memiliki kemampuan analisis dan prediksi yang lebih hebat (stronger predicting power) dibandingkan analisis jalur dan regresi berganda karena SEM mampu menganalisis sampai pada level terdalam terhadap variabel atau konstruk yang diteliti. Literature review ini bertujuan mendeskripsikan penggunaan structural equation modelling pada penelitian. Secara umum, SEM dapat digunakan untuk menganalisis model penelitian yang memiliki beberapa variabel independen (exogen) dan dependen (endogen) serta variabel moderating atau intervening. SEM memberikan beberapa manfaat dan keuntungan bagi para peneliti, diantaranya membangun model penelitian dengan banyak variabel, meneliti variabel atau konstruk yang tidak dapat teramati atau tidak dapat diukur secara langsung (unobserved), menguji kesalahan pengukuran (measurement error) untuk variabel atau konstruk yang teramati (observed) dan confirmatory factor analysis. Secara garis besar metode SEM dapat digolongkan menjadi dua jenis, yaitu covariance based structural equation modeling (CB-SEM) dan variance atau component based SEM (VB-SEM) yang meliputi partial least square (PLS) dan generalized structural component analysis (GSCA). Literature review ini bertujuan mendeskripsikan penggunaan structural equation modelling pada penelitian.

Keywords: data analysis, predicting power, social research, structural equation modelling, variable.

1.Introduction

Sifat dasar manusia ingin terus maju dan berkembang guna mencapai kualitas kehidupan yang lebih baik. Hal ini juga terjadi dalam dunia penelitian. Para ahli ilmu-ilmu sosial atau behavioral termasuk manajemen secara konsisten terus mengembangkan metode penelitian yang dapat digunakan untuk mendapatkan kualitas hasil penelitian yang lebih baik, sempurna, cepat, akurat, efektif dan efisien. Para ahli dalam bidang ilmu sosial atau behavioral termasuk manajemen telah mengembangkan sebuah metode penelitian yang disebut Structural Equation Modeling (SEM). Pada awalnya, metode SEM hanya bagus pada tataran konsepsi. Metode SEM pada saat itu masih belum bisa dioperasionalisasikan karena keterbatasan teknologi. Dengan pesatnya perkembangan teknologi komputer, metode SEM saat ini menjadi semakin dikenal dan banyak digunakan dalam penelitian behavioral dan manajemen. Metode SEM merupakan perkembangan dari analisis jalur (path analysis) dan regresi berganda (multiple regression) yang sama-sama merupakan bentuk model analisis multivariat (multivariate analysis). Dalam analisis yang bersifat asosiatif, multivariate-korelasional atau kausal-efek, metode SEM seakan mematahkan dominasi penggunaan analisis jalur dan regresi berganda yang telah digunakan selama beberapa dekade. Dibandingkan dengan analisis jalur dan regresi berganda, metode SEM lebih unggul karena dapat menganalisis data secara lebih komporehensif. Analisis data pada analisis jalur dan regresi berganda hanya dilakukan terhadap data total score variabel yang merupakan jumlah dari butir-butir instrumen penelitian. Dengan demikian, analisis jalur dan regresi berganda sebenarnya hanya dilakukan pada tingkat variabel laten (unobserved). Sedangkan analisis data pada metode SEM bisa menusuk lebih dalam karena dilakukan terhadap setiap score butir pertanyaan sebuah instrumen variabel penelitian. Butir-butir instrumen dalam analisis SEM disebut sebagai variabel manifes (observed) atau indikator dari sebuah konstruk atau variabel laten.

Metode SEM memiliki kemampuan analisis dan prediksi yang lebih hebat (stronger predicting power) dibandingkan analisis jalur dan regresi berganda karena SEM mampu menganalisis sampai pada level terdalam terhadap variabel atau konstruk yang diteliti. Metode SEM lebih koprehensif dalam menjelaskan fenomena penelitian. Sementara analisis jalur dan regresi berganda hanya mampu menjangkau level variabel laten sehingga mengalami jalan buntu untuk mengurai dan menganalisis fenomena empiris yang terjadi pada level butir-butir atau indicator indikator dari variabel laten. Dilihat dari data yang digunakan, analisis jalur dan regresi berganda sejatinya hanya menjangkau kulit luar sebuah model penelitian. Sedangkan metode SEM dapat diibaratkan mampu menjangkau sekaligus mengurai dan menganalisis isi perut terdalam sebuah model penelitian. Metode SEM diharapkan mampu menjawab kelemahan dan kebuntuan yang dihadapi metode multivariat generasi sebelumnya, yaitu analisis jalur dan regresi berganda. Perkembangan metode SEM menjadi semakin signifikan dalam praktek penelitian sosial, behavioral dan manajemen seiring dengan kemajuan teknologi informasi. Banyak metode statistik multivariat yang pada tahun 1950-an sulit dioperasionalisasikan secara manual, seperti analisis faktor, regresi berganda yang lebih dari tiga variabel bebas, analisis jalur dan analisis diskriminan berangsur angsur menjadi niscaya karena ditemukannya program-program komputer seperti : SPSS (Statistical Package for Social Science), Minitab, Prostat, QSB, SAZAM, dll. Metode SEM saat ini diperkirakan sebagai metode multivariate yang paling dominan. Program komputer yang saat ini dapat digunakan untuk mengolah data pada penelitian metode SEM diantaranya AMOS, LISREL, PLS, GSCA, dan TETRAD. Literature review ini bertujuan mendeskripsikan penggunaan structural equation modelling pada penelitian.

Manfaat SEM dalam penelitian

Secara umum, SEM dapat digunakan untuk menganalisis model penelitian yang memiliki beberapa variabel independen (exogen) dan dependen (endogen) serta variabel moderating atau intervening. SEM memberikan beberapa manfaat dan keuntungan bagi para peneliti, diantaranya membangun model penelitian dengan banyak variabel, meneliti variabel atau konstruk yang tidak dapat teramati atau tidak dapat diukur secara langsung (unobserved), menguji kesalahan pengukuran (measurement error) untuk variabel atau konstruk yang teramati (observed), mengkonfirmasi teori sesuai dengan data penelitian (confirmatory factor analysis), dapat menjawab berbagai masalah riset dalam suatu set analisis secara lebih sistematis dan komprehensif; lebih ilustratif, kokoh dan handal dibandingkan model regresi ketika memodelkan interaksi, non-linieritas, pengukuran error, korelasi error terms, dan korelasi antar variabel laten independen berganda; digunakan sebagai alternatif analisis jalur dan analisis data runtut waktu (time series) yang berbasis kovariat; analisis faktor, jalur dan regresi; menjelaskan keterkaitan variabel secara kompleks dan efek langsung maupun tidak langsung dari satu atau beberapa variabel terhadap variabel lainnya; dan memiliki fleksibilitas yang lebih tinggi bagi peneliti untuk menghubungkan antara teori dengan data.

3. Jenis-jenis SEM

Seperti yang telah diungkapkan diatas, secara garis besar metode SEM dapat digolongkan menjadi dua jenis, yaitu covariance based structural equation modeling (CB-SEM) dan variance atau component based SEM (VB-SEM) yang meliputi partial least square (PLS) dan generalized structural component analysis (GSCA). Varian adalah penyimpangan data dari nilai mean (rata-

rata) data sampel. Varian mengukur penyimpangan data dari nilai mean suatu sampel, sehingga merupakan suatu ukuran untuk variabel-variabel metrik. Secara matematik, varians adalah rata-rata perbedaan kuadrat antara tiap-tiap observasi dengan mean, sehingga varians adalah nilai rata-rata kuadrat dari standar deviasi. Suatu variabel pasti memiliki varians yang selalu bernilai positif, jika nol maka bukan variabel tapi konstanta. Sedangkan covariance menunjukkan hubungan linear yang terjadi antara dua variabel, yaitu X dan Y. Jika suatu variabel memiliki hubungan linear positif, maka kovariannya adalah positif. Jika hubungan antara X dan Y berlawanan, maka kovariannya adalah negatif. Jika tidak terdapat hubungan antara dua variabel X dan Y, maka kovariannya adalah nol.

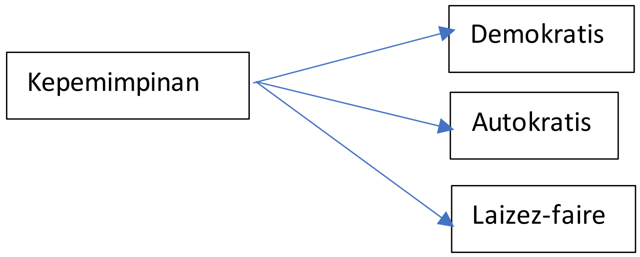
Covariance based structural equation modeling (CB-SEM)

SEM berbasis covariance (CB-SEM) dikembangkan pertama kali oleh Joreskog (1973), Keesling (1972) dan Wiley (1973). CB-SEM mulai populer setelah tersedianya program LISREL III yang dikembangkan oleh Joreskog dan Sorbom pada pertengahan tahun 1970-an. Dengan menggunakan fungsi maximum likelihood (ML), CB-SEM berusaha meminimumkan perbedaan antara covariance matrix sampel dengan covariance matrix prediksi oleh model teoritis sehingga proses estimasi menghasilkan residual covariance matrix yang nilainya kecil mendekati nol. Beberapa hal yang perlu diperhatikan dalam analisis CB-SEM diantaranya :

a. Asumsi penggunaan CB-SEM seperti analisis parametrik. Asumsi yang harus dipenuhi yaitu variabel yang diobservasi harus memiliki multivariate normal distribution serta observasi harus independen satu sama lain. Jika sample kecil dan tidak asimptotik akan memberikan hasil estimasi paramater dan model statistik yang tidak baik atau bahkan menghasilkan varian negatif yang disebut Heywood Case.

b. Jumlah sampel yang kecil secara potensial akan menghasilkan kesalahan Tipe II yaitu model yang jelek masih menghasilkan model yang fit.

c. Analisis CB-SEM mengharuskan bentuk variabel laten yang indikatornya bersifat reflektif. Dalam model reflektif, indikator atau manifest dianggap variabel yang dipengaruhi oleh variabel laten sesuai dengan teori pengukuran klasik. Pada model indikator reflektif, indikator-indikator pada suatu konstruk (variabel laten) dipengaruhi oleh konsep yang sama. Perubahan dalam satu item atau indikator akan mempengaruhi perubahan indikator lainnya dengan arah yang sama. Adapun contoh yang dimaksud sebagai variable reflektif ialah :

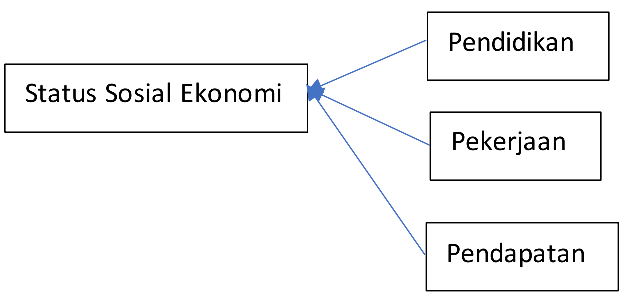


Gambar 1. Contoh variable refletif dari variable laten (konstruk). Demokratis, autokratis dan Laizez-faire merupakan variable reflektif dari kepemimpinan. Variabel reflektif merupakan variable yang menjauhi variable laten (konstruk) sebagaiman terlihat dalam panah biru diatas.

Variance based SEM (VB-SEM)

PLS-SEM

PLS-SEM bertujuan untuk menguji hubungan prediktif antar konstruk dengan melihat apakah ada hubungan atau pengaruh antar konstruk tersebut. Konsekuensi logis penggunaan PLS-SEM adalah pengujian dapat dilakukan tanpa dasar teori yang kuat, mengabaikan beberapa asumsi (non-parametrik) dan parameter ketepatan model prediksi dilihat dari nilai koefisien determinasi (R^2). PLS-SEM sangat tepat digunakan pada penelitian yang bertujuan mengembangkan teori. PLS-SEM dikembangkan untuk mengatasi pengujian yang tidak bisa dilakukan dengan CB-SEM. Misalnya pada pengujian variable formatif, Adapun contoh variable formatif dibawah ini :



Gambar 1. Contoh variable formatif dari variable laten (konstruk). Pendidikan, Pekerjaan dan Pendapatan merupakan variable formatif dari status social ekonomi. Variabel formatif merupakan variable yang menuju atau mempengaruhi atau membentuk variable laten (konstruk) sebagaiman terlihat dalam panah biru diatas.

GSCA

GSCA menggabungkan karakteristik yang terdapat pada CB-SEM dan PLS-SEM. GSCA dapat meng-handle variabel laten dengan banyak indikator sama seperti PLS-SEM, mensyaratkan kriteria goodness of fit model serta indikator dan konstruknya harus berkorelasi seperti CB-SEM. Metode GSCA sampai saat ini jarang digunakan secara luas oleh para peneliti karena metode ini relatif masih baru. GSCA memiliki tujuan yang sama dengan PLS-SEM, tidak mensyaratkan asumsi multivariate normality data, dan bisa dilakukan pengujian tanpa dasar teori yang kuat dengan jumlah sampel yang kecil.

Model covariance-based SEM (CB-SEM) sering disebut hard-modeling, sedangkan component-based atau variance-based SEM (VB-SEM) modeling disebut soft-modeling. Hard modeling bertujuan memberikan pernyataan tetang hubungan kausalitas atau memberikan deskripsi mekanisme hubungan kausalitas (sebab-akibat). Hal ini memberikan gambaran yang ideal secara ilmiah dalam analisis data. Namun demikian, data yang akan dianalisis tidak selalu memenuhi kriteria ideal sehingga tidak dapat dianalisis dengan hard modeling. Sebagai solusinya, soft modeling mencoba menganalisis data yang tidak ideal. Secara harafiah, soft sebenarnya memiliki arti lunak atau lembut, namun dalam kontek penelitian soft diartikan sebagai tidak mendasarkan pada asumsi skala pengukuran, distribusi data dan jumlah sampel. Tujuan utama analisis dengan hard modeling adalah menguji hubungan kausalitas antar yang sudah dibangun berdasarkan teori, apakah model dapat dikonfirmasi dengan data empirisnya. Sedangkan tujuan utama analisis soft modeling bertujuan mencari hubungan linear prediktif antar konstruk laten. Perlu dipahami bahwa hubungan kausalitas atau estimasi tidak sama dengan hubungan prediktif. Pada hubungan kausalitas, CB-SEM mencari invariant parameter yang secara struktural atau fungsional menggambarkan bagaimana sistem di dunia ini bekerja. Invariant parameter menggambarkan hubungan kausalitas antar variabel dalam sistem tertutup (closed system) sehingga kejadian yang ada dapat dikendalikan secara penuh. Sedangkan pada Partial Least Square, Variance atau Component-Based SEM, hubungan linear yang optimal antar laten dihitung dan diinterpretasikan sebagai hubungan prediktif terbaik yang tersedia dengan segala keterbatasan yang ada. Sehingga kejadian yang ada tidak dapat dikendalikan secara penuh. Jika data yang akan dianalisis memenuhi semua asumsi yang dipersyaratkan oleh CB-SEM, maka sebaiknya peneliti menganalisis data dengan hard modeling menggunakan Software yang sesuai, seperti AMOS, LISREL.

Jika data tidak memenuhi semua asumsi yang dipersyaratkan namun peneliti tetap menggunakan analisis hard modeling atau CB-SEM, maka beberapa masalah yang mungkin akan dihadapi adalah : terjadi im-proper solution atau solusi yang tidak sempurana, karena adanya Heywood Case, yaitu gejala nilai varian yang negatif; model menjadi un-identified karena terjadi faktor indeterminacy; dan non-convergence algorithm. Bila kondisi di atas terjadi dan kita masih ingin menganalisis data, maka tujuan kita ubah bukan mencari hubungan kausalitas antar variabel, tapi mencari hubungan linear prediktif optimal dengan menggunakan component atau variance based-SEM.

Berdasarkan tujuannya riset empiris paradigma kuantitatif dapat dibagi menjadi dua, yaitu estimasi dan prediksi. Riset estimasi adalah riset yang bertujuan untuk menguji suatu model empiris dengan pengukur-pengukur yang valid dan reliabel. Pengujian dan pengukuran dilakukan pada level indikator. Hipotesis yang diuji adalah hipotesis model. Kriteria pengukuran untuk menguji kelayakan model disebut goodness of fit test. Untuk tujuan riset estimasi, CB-SEM adalah teknik yang tepat untuk digunakan. Riset prediksi adalah riset yang bertujuan untuk menguji pengaruh antar konstruk untuk memprediksi hubungan sebab akibat. Pengujian dan pengukuran dilakukan pada level konstruk atau variabel laten. Hipotesis yang dilakukan pada umumnya hipotesis parsial. Kriteria pengujian parsial dengan uji signifikansi prediksi hubungan antar variabel dengan menggunakan uji t-statistik. Teknik PLS-SEM dan regresi adalah pilihan teknik statistik yang tepat untuk digunakan. Jadi component atau variance based SEM (PLS dan GSCA) hanya digunakan jika data yang kita miliki tidak dapat diselesaikan dengan covariance based SEM (CB-SEM).

Conclusion

SEM dapat digunakan untuk menganalisis model penelitian yang memiliki beberapa variabel independen dan dependen serta variabel moderating atau intervening.

References

Burnham KP, Anderson DR, Huyvaert KP (2011) AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. Behav Ecol Sociobiol 65(1):23–35

Byrne BM (2013) Structural equation modeling with AMOS: basic concepts, applications, and programming. Routledge, New York

Capmourteres V, Anand M (2016) Assessing ecological integrity: a multi-scale structural and functional approach using structural equation modeling. Ecol Indic http://dx.doi.org/10.1016/j.ecolind.2016.07.006

Chang WY (1981) Path analysis and factors affecting primary productivity. J Freshwater Ecol 1(1):113–120 Chapin FS, Conway AJ, Johnstone JF, Hollingsworth TN, Hollingsworth J (2016) Absence of net long-term successional facilitation by alder in a boreal Alaska floodplain. Ecology, doi: http://dx.doi.org/10.1002/ecy.1529

Chaudhary VB, Bowker MA, O'Dell TE, Grace JB, Redman AE, Rillig MC, Johnson NC (2009) Untangling the biological contributions to soil stability in semiarid shrublands. Ecol Appl 19(1):110–122

Chen Y, Lin L (2010) Structural equation-based latent growth curve modeling of watershed attribute-regulated stream sensitivity to reduced acidic deposition. Ecol Model 221(17):2086–2094

Chen F, Curran PJ, Bollen KA, Kirby J, Paxton P (2008) An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. Socio Meth Res 36(4):462–494

Chen J, John R, Shao C, Fan Y, Zhang Y, Amarjargal A, Brown DG, Qi J, Han J, Lafortezza R, Dong G (2015) Policy shifts influence the functional changes of the CNH systems on the Mongolian plateau. Environ Res Lett 10(8):085003

Cohen J (2013) Statistical power analysis for the behavioral sciences. Academic Press, New York

Cover TM, Thomas JA (2012) Elements of information theory. John Wiley & Sons, Hoboken, New Jersey

Cudeck R, Odell LL (1994) Applications of standard error estimates in unrestricted factor analysis: significance tests for factor loadings and correlations. Psychol Bull 115(3):475–487

Curran PJ (2003) Have multilevel models been structural equation models all along? Multivar Behav Res 38(4):529–569

Curran PJ, West SG, Finch JF (1996) The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. Psychol Methods 1(1):16–29

Curran PJ, Bollen KA, Paxton P, Kirby J, Chen F (2002) The noncentral chi-square distribution in misspecified structural equation models: finite sample results from a Monte Carlo simulation. Multivar Behav Res 37(1):1–36

Duncan TE, Duncan SC, Strycker LA (2013) An introduction to latent variable growth curve modeling: concepts, issues, and applications, 2nd edn. Psychology Press, New York

Eisenhauer N, Bowker M, Grace J, Powell J (2015) From patterns to causal understanding: structural equation modeling (SEM) in soil ecology. Pedobiologia 58(2):65–72

Fan X, Sivo SA (2005) Sensitivity of fit indexes to misspecified structural or measurement model components: rationale of two-index strategy revisited. Struct Equ Modeling 12(3):343–367

Fan X, Thompson B, Wang L (1999) Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. Struct Equ Modeling 6(1):56–83

Fritz MS, MacKinnon DP (2007) Required sample size to detect the mediated effect. Psychol Sci 18(3):233–239

Galton F (1888) Personal identification and description. Nature 38:173–177

Grace JB (2006) Structural equation modeling and natural systems. Cambridge University Press, New York

Grace JB, Bollen KA (2008) Representing general theoretical concepts in structural equation models: the role of composite variables. Environ Ecol Stat 15(2):191–213

Grace JB, Anderson TM, Olff H, Scheiner SM (2010) On the specification of structural equation models for ecological systems. Ecol Monogr 80(1):67–87

Haavelmo T (1943) The statistical implications of a system of simultaneous equations. Econometrica 11(1):1–12

Hair JF, Hult GT, Ringle C, Sarstedt M (2013) A primer on partial least squares structural equation modeling (PLS-SEM). Sage, Thousand Oak

Harrington D (2009) Confirmatory factor analysis. Oxford University Press, New York Heise DR (1975) Causal analysis. John Wiley & Sons, Oxford

Hinton GE, Osindero S, Teh YW (2006) A fast learning algorithm for deep belief nets. Neural Comput 18(7):1527–1554

Hoyle RH (2011) Structural equation modeling for social and personality psychology. Sage, London

Hoyle RH, Isherwood JC (2013) Reporting results from structural equation modeling analyses in Archives of Scientific Psychology. Arch Sci Psychol 1:14–22

Hu LT, Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Modeling 6(1):1–55

Iacobucci D (2010) Structural equations modeling: fit indices, sample size, and advanced topics. J Consum Psychol doi: http://dx.doi.org/10.1016/j.jcps.2009.09.003

Jackson DL, Gillaspy JA, Purc-Stephenson R (2009) Reporting practices in confirmatory factor analysis: an overview and some recommendations. Psychol Methods 14(1):6–23

Johnson JB, Omland KS (2004) Model selection in ecology and evolution. Trends Ecol Evol 19(2):101–108

Jones CM, Spor A, Brennan FP, Breuil MC, Bru D, Lemanceau P, Griffiths B, Hallin S, Philippot L (2014) Recently identified microbial guild mediates soil N2O sink capacity. Nat Clim Change 4(9):801–805

Joreskog K, Sorbom D (1993) LISREL 8: structural equation modeling with the SIMPLIS command language. Scientific Software International, Chicago

Kaplan D, Depaoli S (2012) Bayesian structural equation modeling. In: Hoyle RH (ed) Handbook of structural equation modeling. Guilford, New York

Kim KH (2005) The relation among fit indexes, power, and sample size in structural equation modeling. Struct Equ Modeling 12(3):368–390

Kline RB (2006) Reverse arrow dynamics. Formative measurement and feedback loops. In: Hancock GR, Mueller RO (eds) Structural equation modeling: A second course. Information Age Publishing, Greenwich

Kline RB (2010) Principles and practice of structural equation modeling. Guilford Press, New York

Lamb EG, Mengersen KL, Stewart KJ, Attanayake U, Siciliano SD (2014) Spatially explicit structural equation modeling. Ecology 95(9):2434–2442

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444

Lee SY (2007) Structural equation modeling: a Bayesian approach. John Wiley & Sons, Hong Kong

Lefcheck J (2015) piecewiseSEM: Piecewise structural equation modelling in r for ecology, evolution, and systematics. Methods Ecol Evol, doi: http://dx.doi.org/10.1111/2041-210X.12512

Liu X, Swenson NG, Lin D, Mi X, Umaña MN, Schmid B, Ma K (2016) Linking individual-level functional traits to tree growth in a subtropical forest. Ecology, doi: http://dx.doi.org/10.1002/ecy.1445

MacCallum RC, Hong S (1997) Power analysis in covariance structure modeling using GFI and AGFI. Multivar Behav Res 32(2):193–210

Maddox GD, Antonovics J (1983) Experimental ecological genetics in Plantago: a structural equation approach to fitness components in P. aristata and P. patagonica. Ecology 64(5):1092–1099

McDonald RP, Ho MH (2002) Principles and practice in reporting structural equation analyses. Psychol Methods 7(1):64–82

Mehta PD, Neale MC (2005) People are variables too: multilevel structural equations modeling. Psychol Methods 10(3):259–284

Monecke A, Leisch F (2012) semPLS: structural equation modeling using partial least squares. J Stat Softw 48(3):1–32

Mulaik SA, James LR, Van Alstine J, Bennett N, Lind S, Stilwell CD (1989) Evaluation of goodness-of-fit indices for structural equation models. Psychol Bull 105(3):430–445

Murtaugh PA (2009) Performance of several variable-selection methods applied to real ecological data. Ecol Lett 12(10):1061–1068

Submitted to the journal “Open Access Indonesia Journal of Social Sciences
(November 25th, 2022)

Open Access Indonesia Journal of Social Sciences

HM Publisher

Submission acknowledgement

Dear author(s),

Rachmat Hidayat*, Patricia Wulandari has submitted the manuscript “Data Analysis Procedures with Structural Equation Modelling (SEM): Narrative Literature Review“ to Open Access Indonesia Journal of Social Sciences. The paper will be screened by editor and reviewed by peer review.

Cordially,

Prof. Paula Magnano, PhD
Editor HM Publisher

(*) Corresponding author

Bioscientia Medicina

Journal of Biomedicine and Translational Research

HM Publisher

<u>Peer Review Results</u>

Dear author(s),

Rachmat Hidayat*, Patricia Wulandari has submitted the manuscript "Data Analysis Procedures with Structural Equation Modelling (SEM): Narrative Literature Review" to Open Access Indonesia Journal of Social Sciences. The decision : Revision Required.

Cordially,

<u>Prof. Paula Magnano, PhD</u>
Editor HM Publisher

(*) Corresponding author

Reviewer 1: Revision required

Data Analysis Procedure with Structural Equation Modeling (SEM): Narrative Literature Review →1

Rachmat Hidayat[1*], Patricia Wulandari[2]

[1]Department of Biology, Faculty of Medicine, Universitas Sriwijaya, Palembang, Indonesia

[2]Cattleya Mental Health Center, Palembang, Indonesia

Email: rachmathidayat@fk.unsri.ac.id

Abstract→3

Relationships between variables in structural equation modeling (SEM) form a structural model. This structural model can be explained through structural equations as in regression analysis. This structural equation describes the prediction of latent (exogenous) independent variables against latent (endogenous) dependent variables. Researchers who use analysis with structural equation models need to know whether models built with empirical data have unique values or not so that the model can be estimated. If a model does not have a unique value then it is unidentified. The reason a model is categorized as unidentified is because the information contained in the empirical data is not enough to produce a unique solution in calculating the model's estimation parameters. This literature review aims to describe the process of data analysis using SEM.

Keywords: Data analysis, Structural equation modelling, Variable. →2

1.Introduction→4

There is a principle difference between regression and path analysis with SEM in terms of variable measurement. In the regression analysis, dependent and independent variables are variables that can be measured directly (observable), while in SEM dependent and independent variables are variables that cannot be measured directly (unobservable). Unobserved variables are also often called latent variables. The structural equation model or SEM is a model that explains the relationship between latent variables so that the SEM model is often referred to as latent analysis or linear structural relationship. The relationship between variables in SEM is the same as the relationship in path analysis. However, in explaining the relationship between latent variables, the SEM model is different from path analysis where path analysis uses observable variables while SEM uses unobservable variables. Relationships between variables in SEM form a structural model. This structural model can be explained through structural equations as in regression analysis. This structural equation describes the prediction of latent (exogenous) independent variables against latent (endogenous) dependent variables.

2. Model Specifications

SEM begins by specifying a research model. The analysis will not begin until the researcher has specified a model that shows the relationships between the variables to be analyzed. Through the steps below, the researcher can obtain the desired model;

- Define the latent variables of the study.
- Define observed variables.
- Define the relationship between latent variables and observed variables. Pay attention to the constructive aspect of the variable unidimensional or multidimensional. A unidimensional construct (first order construct) is a construct that directly describes the relationship between a latent variable and a reflectively observed variable (the direction of the arrow away from the latent variable) or formatively (the direction of the arrow towards the latent variable). A multidimensional construct (second order construct) is a construct formed from several unidimensional constructs.

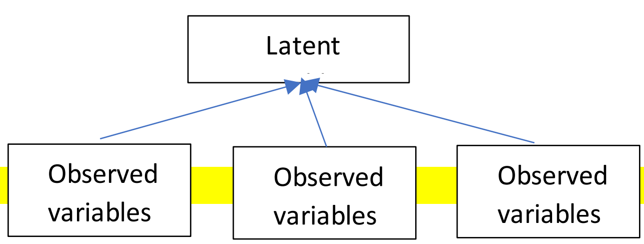


Figure 1. Unidimensional construct (First order construct).

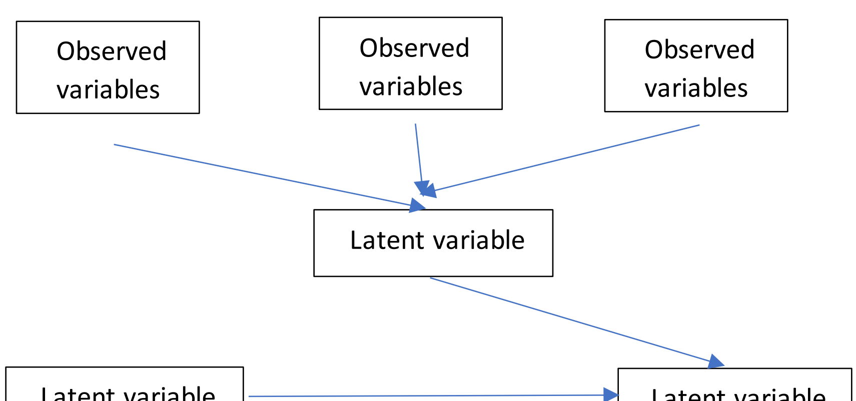


Figure 2. Multi-dimensional construct (Second order construct).

3. Model Identification

Researchers who use analysis with structural equation models need to know whether models built with empirical data have unique values or not so that the model can be estimated. If a model does not have a unique value then it is unidentified. The reason a model is categorized as unidentified is because the information contained in the empirical data is not enough to produce a unique solution in calculating the model's estimation parameters. An example of an under-identified case is A x B = 1000. The question is, what is the value of A or B? To determine the magnitude of the value of A or B, of course, the answer varies greatly, it can be 100 x 10, 500 x 2, 250 x 4, 200 x 5 or 1 x 1000. To ensure an answer, we must select the most appropriate (unique) answer called model identification. This example also occurs in SEM, where constructed theoretical models and empirical data are not enough to produce one unique solution in calculating the estimated parameters of the model. However, if we specify the value of A = 100, then automatically the value of B = 10. This can also be done in SEM analysis, to overcome unidentified models by constraining models with:

- Adding indicators or manifest variables of latent constructs,
- Determining the value of additional fixed parameters so as to produce a positive calculation of the degree of freedom (this method is most often used by researchers),
- Assume that one parameter has the same value.

It should be noted that the use of the above three methods to change an under-identified model must be theoretical, not merely done so that the model can be identified.

There are three possibilities for model identification in SEM:

- Under-identified Model, where the value of $t \geq s / 2$; that is, a model with the number of parameters estimated to be greater than the number of known data (the data is the variance and covariance of the observed variables). For example, there is the Equation X + Y = 10, representing 1 piece of known data and 2 parameters to be estimated, namely X and Y, then the number of df = 1-2 = -1. From the understanding of unidentified models in SEM has df = number of known data – the number of parameters estimated < 0. So it can be concluded that under-identified models have a negative df.
- Just-Identified Model, where t = s/2; i.e. a model with the number of parameters estimated equal to the known data. For example, there are two equations X + Y = 10 and X + 2Y = 16, which are 2 pieces of known data and 2 parameters to be estimated, namely X and Y, then the number of df = 2-2 = 0. So it can be concluded that just-identified models have zero df.
- Over-Identified Model, where $t \leq s/2$; i.e. a model with an estimated number of parameters smaller than the number of known data. For example, there are three equations X + Y = 10, X + 2Y = 16 and 3X + 2Y = 22, which are 3 pieces of known data and 2 parameters to be estimated, namely X and Y, then the number of df = 3-2 = 1. So it can be concluded that over-identified models have a positive *df*.

Information:

t = number of estimated parameters,

s = number of variants and covariance between indicators.

5. Model Fit Test (Assessment **of Fit)**

a. Overall Fit of the Model.

The first stage of the match test is shown to evaluate in general terms the degree of goodness of fit (GOF) between the data and the model. Assessing the overall fit of the model in SEM cannot be done directly as in other multivariate techniques (multiple regression, discriminant analysis, MANOVA, etc.). SEM does not have one of the best statistical tests that can explain the predictive power of the model. Instead, researchers have developed several sizes of GOF or

Goodness of Fit Indices (GOFI) that can be used in tandem or in combination. This circumstance makes the overall fit test stage a move that invites a lot of debate and controversy. The controversy and debate surrounding the GOF arises when the question of size arises in the discussion, i.e. how much of a fit can be said to be accepted? Apart from the controversy, there was finally a consesus among researchers, some of whom were:

- The best clue in assessing the suitability of a model is a strong theory of substance. If the model only shows or represents a substantive theory that is not strong, and although the model has an excellent fit, it is rather difficult for us to judge the model.
- The Chi-square (X2) statistical test should not be the only basis for determining the suitability of the data to the model.
- None of the GOF or GOF Indices (GOFI) measures can exclusively be used as a basis for evaluating the overall fit of the model. GOFI is grouped into three parts, namely absolute fit measures, incremental fit measures and parsimonious fit measures.

Absolute Fit Size

The absolute fit measure determines the degree of prediction of the overall model (structural and measuring models) against the correlation matrix and covariance. This measure contains measures that represent the overall fit point of view mentioned earlier. Of the various absolute fit measures, the measures that are usually used to evaluate SEM are:

1. Chi-square (X2). A low Chi-square (X 2) value that results in significance level > 0.05 or ($p > 0.05$) indicating a null hypothesis is accepted. This means that the predicted input matrix with the actual one is not statistically different. Chi-square (X 2) cannot be used the only size fits the entire model, one of the reasons is because Chi-square (X2) is sensitive to the sample size.
2. Non-Centrality Parameter (NCP). Like X2, NCP is also a measure of badness of fit where the greater the difference between Σ and $\Sigma(\Theta)$ the greater the NCP value. So, we need to look for NCPs that are small or low in value.
3. Goodness of Fit Index (GFI). GFI values range from 0 (poor fit) to 1 (perfect fit), and GFI values > 0.90 are good fit, while $0.80 < GFI < 0.90$ are often referred to as marginal fit.

4. Root Mean Square Residual (RMR). RMR represents the average value of all standardized residuals, and has about from 0 to 1. Models that have a good fit will have an RMR value smaller than 0.05.
5. Root Mean Square Error of Approximation (RMSEA). RMSEA values between 0.08 and 0.10 indicate marginal fit, and RMSEA values > 0.10 indicate poor fit.
6. Single sample cross-validation index/expected cross-validation index (ECVI). ECVI is used for model comparison and the smaller the ECVI of a model the better the degree of fit.

Incremental Match Size

Incremental match measures compare the proposed model with the baseline model often referred to as the null model or independence model and saturated model. A null model is a model with the worst fit. Saturated models are the ones with the best fit. The concept of incremental match would put the data-model match rate between the null model and the saturated model. The data-model match rate that resides between the null model and the saturated model is called a nested model. Of the various incremental match sizes, the measures typically used to evaluate SEM are:

1. Adjusted Goodness of Fit Index (AGFI). NAGFI values range from 0 to 1and AGFI values > 0.90 indicate good fit. While 0.80 < GFI < 0.90 is often referred to as marginal fit.
2. Tucker-Lewis Index / Non Normed Fit Index (TLI/NNFI). TLI values range from 0 to 1.0, with TLI values > 0.90 indicating good fit and 0.80 < TLI< 0.90 is marginal fit.
3. Normed Fit Index (NFI). This NFI has values that range from 0 to 1. NFI values > 0.90 indicate good fit, while 0.80 < NFI < 0.90 are often referred to as marginal fit.
4. Relative Fit Index (RFI). NRFI values will range from 0 to 1. An RFI value > 0.90 indicates good fit, while 0.80 < RFI < 0.90 is often referred to as marginal fit.
5. Incremental Fit Index (IFI). The value of the IFI will range from 0 to 1. An IFI value > 0.90 indicates good fit, while 0.80 < an IFI < 0.90 is often referred to as marginal fit.

6. Comparative Fit Index (CFI). CFI values will range from 0 to 1. A CFI value > 0.90 indicates a good fit, while a 0.80 < a CFI < 0.90 is often referred to as a marginal fit.

Parsimoni match size

Models with relatively few parameters (and relatively many degrees of freedom) are often known as models that have parsimony or high efficiency. Whereas a model with many parameters (and a little degree of freedom) can be said to be a complex and less parsimony model. Of the various sizes of parsimony fit, the measures that are usually used to evaluate SEM are:

1. Parsimonious Normed Fit Index (PNFI). Higher PNFI values are better. The use of PNFI is mainly for the comparison of two or more models that have different degrees of freedom. PNFI is used to compare alternative models, and there is no acceptable match rate recommendation. However, when comparing 2 models, the difference in PNFI values of 0.06 to 0.09 indicates a considerable difference in models.
2. Parsimonious goodness of fit index (PGFI). PGFI values range between 0 and 1, with higher values indicating a better parsimoni model.
3. Normed Chi-square. Recommended values: lower limit: 1.0, upper limit: 2.0 or 3.0 and looser 5.0.
4. Akaike information criterion (AIC). A small, near-zero AIC value indicates a better match, as well as a higher parsimony.

Measurement model fit (measurement model analysis)

The suitability of the measurement model is carried out by conducting validity and reliability tests. Validity relates to whether a variable measures what it should measure. Although validity can never be proved, but support towards such proof can be developed. A variable is said to have good validity to its latent construct or variable, if: the value of t load *factor* is greater than the critical value or >1.96 or practically >2), and its standard load factor (standardized loading factors) >0.70. Reliability is the consistency of a measurement. High reliability indicates that the indicator has a high consistency in measuring its latent construct. To measure reliability in SEM will be used: composite reliability measure and variance extracted measure (variant extract size).

A construct has good reliability, if its Construct Reliability (CR) value is > 0.70 and its Variance Extracted (VE) value is > 0.50.

Structural Model Fit (Structural model analysis)

Analysis of structural models includes an examination of the significance of the estimated coefficients. The SEM method and its Software not only provide the estimated values of the coefficients but also the t-calculated values for each coefficient. By specifying a significant degree (usually □□= 0.50), then each coefficient representing a hypothesized causal relationship can be statistically tested for significance. In addition to this, it is also necessary to evaluate the standard solution where all beta coefficients on multiple regressions, that is, coefficient values close to zero indicate a smaller influence. The increase in the value of this coefficient corresponds to an increase in the importance of the variable in question in the causal relationship. As a comprehensive measure of structural equations, the overall coefficient of determination (R2) is calculated as in multiple regression.

Respecification/modification and modeling strategy

Respecification is the next step after the match test is carried out. The implementation of the respecification is highly dependent on the modeling strategy to be used. There are 3 modeling strategies to choose from in SEM, namely:

1. Confirmatory modeling strategy. In this modeling strategy, a single model is formulated or specified, then empirical data collection is carried out to be tested for significance. This test will result in an acceptance or rejection of the model. This strategy does not require respecification.
2. Competing modeling strategy. In this modeling strategy several alternative models are specified and based on the analysis of a group of empirical data one of the most suitable models is selected. In this strategy, respecification is only necessary if alternative models are developed from several existing models.
3. Model development strategy. In this modeling strategy an initial model is specified and empirical data is collected. If the initial model does not match the existing data, then the model is modified and retested with the same data. Several models can be tested in this process with the aim of finding a single model that not only matches the data well, but also

has the property that each of its parameters can be interpreted properly. Respecification of the model can be done on a theory-driven or data-driven basis, however, respecation based on theory-driven is recommended

Confirmatory modeling (SC) strategies are rarely encountered, as generally researchers are not satisfied enough with simply rejecting a model without proposing an alternative model. Currently the most used in research is the model development strategy. After the model estimation is carried out, the researcher can still make modifications to the developed model if it turns out that the resulting estimate has a large residual. However, modifications can only be made if the researcher has a strong enough theoretical justification, because SEM is not intended to produce a theory, but to test a model that has a correct theoretical footing, therefore to provide an interplay as to whether the theory-based model being tested is directly acceptable or needs modification, the researcher must focus his attention on the predictive power of the model, namely by observing the residual magnitude that Generated. If in the standardized residual covariances matrix there is a value outside the range of -2.58 < the residual < 2.58 and probability (P) if < 0.05 then the estimated model needs to be further modified based on the modification index by choosing the largest modification index (MI) and has a theoretical basis. The largest MI will indicate that if the coefficient is estimated, there will be a significant reduction in the value of chi square (X2). In SEM software, the modification index is listed in the output so that the researcher only needs to choose which coefficient to estimate. If the value of chi square (X2) is still not significant, the next largest MI value is sought and so on.

Conclusion→5

The data analysis process using SEM starts from model specifications, model identification, model match tests and specifications/modifications and modeling strategies. modifications can only be made if the researcher has a strong enough theoretical justification, because SEM is not intended to produce a theory, but to test a model that has a correct theoretical footing, therefore to provide an interplay as to whether the theory-based model being tested is directly acceptable or needs modification, the researcher must focus his attention on the predictive power of the model, namely by observing the residual magnitude that Generated.

References→6

Astrachan, C.B., Patel V. K., &Wanzenried G. (2014). A Comparative Studyof CB-SEM and PLS-SEM for Theory development in Family Firm Research. Journal of Family Business Strategy, 5, 116-128.

Babin, B. J. , Boles, J. S., & Robin, D. P. (2000). Representing the perceived ethical work climate among marketing employees. Journal of the Academy of Marketing Science. 28(3), 345-358.

Babin, B.J., Hair, J.F., & Boles, J.S. (2008). Publishing Research in Marketing Journals Using Structural Equation Modeling. Journal of Marketing Theory & Practice, 16(4), 279-285.

Bagozzi, R., & Yi, Y. 2012. Specification, evaluation, and interpretation of structural equation models. Journal of the Academy of Marketing Science, 40(1): 8-34.

Blocker, C., Flint, D., Myers, M., & Slater, S. (2011). Proactive customer orientation and its role for creating customer value in global markets. Academy of Marketing Science Journal, 39(2), 216-233.

Baker, W., E., & Syncula, J., M. (1999). Learning orientation, market orientation, and innovation: Integrating and extending models of organizational performance. Journal of Market - Focused Management, 4(4): 295-308.

Bido, D. D. S., Souza, C. A. D., Silva, D. D., Godoy, A. S., & Torres, R. R. (2012). Quality of reporting methodological procedures in national publications in the area of business administration: the case of structural equation modelling. Organizações & Sociedade, 19(60), 125-144.

Brei, V. A., & Liberali Neto, G. (2006). O uso da técnica de modelagem em equações estruturais na área de marketing: um estudo comparativo entre publicações no Brasil e no exterior. Revista de Administração Contemporânea, 10(4), 131-151.

Byrne, B. M. (2010). Structural equation modeling with AMOS: Basic concepts, applications, and programming (2nd. Edition) New York: Routledge.

Chin, W.W., Peterson, R.A., & Brown, S.P. (2008). Structural Equation Modeling in Marketing: Some Practical Reminders. Journal of Marketing Theory & Practice, 16(4), 287-289.

Churchill, G. A., Jr. (1979). A paradigm for developing better measures of marketing constructs. JMR, Journal of Marketing Research, 16(1), 64-73.

DeConinck, J. B. (2010b). The influence of ethical climate on marketing employees' job attitudes and behaviors. Journal of Business Research, 63(4): 384-391.

Delaney, J. T., & Huselid, M. A. (1996). The impact of human resource management practices on perceptions of organizational performance. Academy of Management Journal, 39(4): 949-969.

DeVellis, R. F. (2011). Scale development: Theory and applications (Vol. 26): Sage Publications, Inc.

Deshpande, R., & Farley, J. (1998). Measuring market orientation: Generalization and synthesis. Journal of Market-Focused Management, 2(3), 213-232.

Deshpande, R., & Webster Jr, F. E. (1989). Organizational culture and marketing: defining the research agenda. The Journal of Marketing, 3-15.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). Internet, mail, and mixed-mode surveys: The tailored design method (3rd ed.). Hoboken: John Wiley & Sons Inc.

Fabrigar, L. R., Porter, R. D., & Norris, M. E. 2010. Some things you should know about structural equation modeling but never thought to ask. Journal of Consumer Psychology, 20(2): 221-225.

Ferrell, O., Gonzalez-Padron, T., Hult, G., & Maignan, I. (2010). From market orientation to stakeholder orientation. Journal of Public Policy & Marketing, 2010, 29(1), 93-96.

Freeman, R. (1984). Strategic management: A stakeholder approach: Pitman Boston, MA.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research, 18(1), 39–50.

Gallagher, D., Ting, L., & Palmer, A. (2008). A journey into the unknown; taking the fear out of structural equation modeling with AMOS for the first-time user. The Marketing Review, 8(3), 255-275.

Grinstein, A. (2008). The effect of market orientation and its components on innovation consequences: A meta-analysis. Journal of the Academy of Marketing Science, 36(2), 166-173.

Reviewer Comment:

1→ Title of Manuscripts should be explained main review and declared type of literature review: narrative or systematic review.

2→ Keywords should be showed the main words of the study, the authors can use MeSH to develop keywords.

3→ Abstract should be showed the main of background, main of review and conclusion of study.

4→Introduction should be showed the urgency of study (epidemiology data), biological plausibility concept, and lack of knowledge in the study.

5→ Conclusion should more specific and not more showed more review.

6→ Authors must check the references for make update references. References should no more than 10 years.

Reviewer 2: Revision required

Data Analysis Procedure with Structural Equation Modeling (SEM): Narrative Literature Review

Rachmat Hidayat[1*], Patricia Wulandari[2]

[1]Department of Biology, Faculty of Medicine, Universitas Sriwijaya, Palembang, Indonesia

[2]Cattleya Mental Health Center, Palembang, Indonesia

Email: rachmathidayat@fk.unsri.ac.id

Abstract **A**

Relationships between variables in structural equation modeling (SEM) form a structural model. This structural model can be explained through structural equations as in regression analysis. This structural equation describes the prediction of latent (exogenous) independent variables against latent (endogenous) dependent variables. Researchers who use analysis with structural equation models need to know whether models built with empirical data have unique values or not so that the model can be estimated. If a model does not have a unique value then it is unidentified. The reason a model is categorized as unidentified is because the information contained in the empirical data is not enough to produce a unique solution in calculating the model's estimation parameters. This literature review aims to describe the process of data analysis using SEM.

Keywords: Data analysis, Structural equation modelling, Variable.

1.Introduction B

There is a principle difference between regression and path analysis with SEM in terms of variable measurement. In the regression analysis, dependent and independent variables are variables that can be measured directly (observable), while in SEM dependent and independent variables are variables that cannot be measured directly (unobservable). Unobserved variables are also often called latent variables. The structural equation model or SEM is a model that explains the relationship between latent variables so that the SEM model is often referred to as latent analysis or linear structural relationship. The relationship between variables in SEM is the same as the relationship in path analysis. However, in explaining the relationship between latent variables, the SEM model is different from path analysis where path analysis uses observable variables while SEM uses unobservable variables. Relationships between variables in SEM form a structural model. This structural model can be explained through structural equations as in regression analysis. This structural equation describes the prediction of latent (exogenous) independent variables against latent (endogenous) dependent variables.

2. Model Specifications

SEM begins by specifying a research model. The analysis will not begin until the researcher has specified a model that shows the relationships between the variables to be analyzed. Through the steps below, the researcher can obtain the desired model;

- Define the latent variables of the study.
- Define observed variables.
- Define the relationship between latent variables and observed variables. Pay attention to the constructive aspect of the variable unidimensional or multidimensional. A unidimensional construct (first order construct) is a construct that directly describes the relationship between a latent variable and a reflectively observed variable (the direction of the arrow away from the latent variable) or formatively (the direction of the arrow towards the latent variable). A multidimensional construct (second order construct) is a construct formed from several unidimensional constructs.

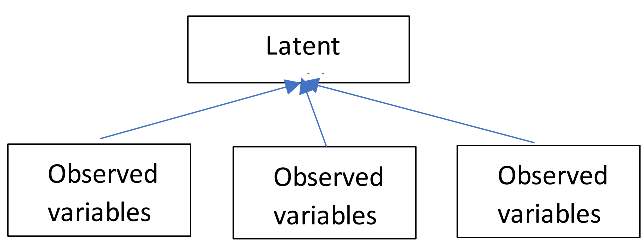


Figure 1. Unidimensional construct (First order construct).

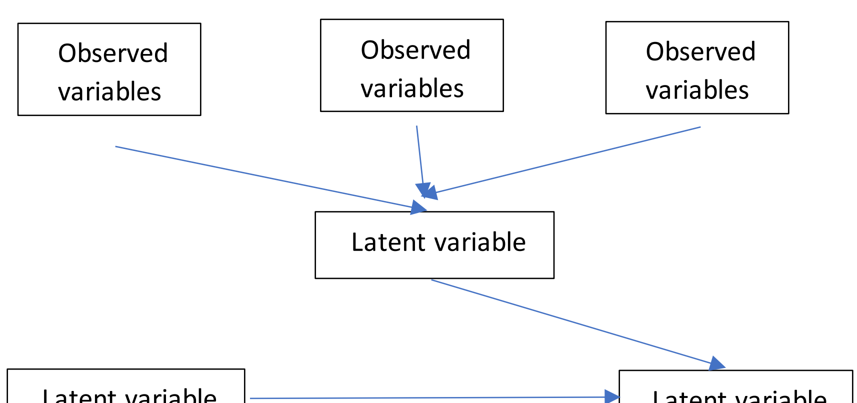


Figure 2. Multi-dimensional construct (Second order construct).

3. Model Identification

Researchers who use analysis with structural equation models need to know whether models built with empirical data have unique values or not so that the model can be estimated. If a model does not have a unique value then it is unidentified. The reason a model is categorized as unidentified is because the information contained in the empirical data is not enough to produce a unique solution in calculating the model's estimation parameters. An example of an under-identified case is A x B = 1000. The question is, what is the value of A or B? To determine the magnitude of the value of A or B, of course, the answer varies greatly, it can be 100 x 10, 500 x 2, 250 x 4, 200 x 5 or 1 x 1000. To ensure an answer, we must select the most appropriate (unique) answer called model identification. This example also occurs in SEM, where constructed theoretical models and empirical data are not enough to produce one unique solution in calculating the estimated parameters of the model. However, if we specify the value of A = 100, then automatically the value of B = 10. This can also be done in SEM analysis, to overcome unidentified models by constraining models with:

- Adding indicators or manifest variables of latent constructs,
- Determining the value of additional fixed parameters so as to produce a positive calculation of the degree of freedom (this method is most often used by researchers),
- Assume that one parameter has the same value.

It should be noted that the use of the above three methods to change an under-identified model must be theoretical, not merely done so that the model can be identified.

There are three possibilities for model identification in SEM:

- Under-identified Model, where the value of $t \geq s / 2$; that is, a model with the number of parameters estimated to be greater than the number of known data (the data is the variance and covariance of the observed variables). For example, there is the Equation X + Y = 10, representing 1 piece of known data and 2 parameters to be estimated, namely X and Y, then the number of df = 1-2 = -1. From the understanding of unidentified models in SEM has df = number of known data – the number of parameters estimated < 0. So it can be concluded that under-identified models have a negative df.
- Just-Identified Model, where t = s/2; i.e. a model with the number of parameters estimated equal to the known data. For example, there are two equations X + Y = 10 and X + 2Y = 16, which are 2 pieces of known data and 2 parameters to be estimated, namely X and Y, then the number of df = 2-2 = 0. So it can be concluded that just-identified models have zero df.
- Over-Identified Model, where $t \leq s/2$; i.e. a model with an estimated number of parameters smaller than the number of known data. For example, there are three equations X + Y = 10, X + 2Y = 16 and 3X + 2Y = 22, which are 3 pieces of known data and 2 parameters to be estimated, namely X and Y, then the number of df = 3-2 = 1. So it can be concluded that over-identified models have a positive *df*.

Information:

t = number of estimated parameters,

s = number of variants and covariance between indicators.

5. Model Fit Test (Assessment **of Fit)**

a. Overall Fit of the Model.

The first stage of the match test is shown to evaluate in general terms the degree of goodness of fit (GOF) between the data and the model. Assessing the overall fit of the model in SEM cannot be done directly as in other multivariate techniques (multiple regression, discriminant analysis, MANOVA, etc.). SEM does not have one of the best statistical tests that can explain the predictive power of the model. Instead, researchers have developed several sizes of GOF or

Goodness of Fit Indices (GOFI) that can be used in tandem or in combination. This circumstance makes the overall fit test stage a move that invites a lot of debate and controversy. The controversy and debate surrounding the GOF arises when the question of size arises in the discussion, i.e. how much of a fit can be said to be accepted? Apart from the controversy, there was finally a consesus among researchers, some of whom were:

- The best clue in assessing the suitability of a model is a strong theory of substance. If the model only shows or represents a substantive theory that is not strong, and although the model has an excellent fit, it is rather difficult for us to judge the model.
- The Chi-square (X2) statistical test should not be the only basis for determining the suitability of the data to the model.
- None of the GOF or GOF Indices (GOFI) measures can exclusively be used as a basis for evaluating the overall fit of the model. GOFI is grouped into three parts, namely absolute fit measures, incremental fit measures and parsimonious fit measures.

Absolute Fit Size

The absolute fit measure determines the degree of prediction of the overall model (structural and measuring models) against the correlation matrix and covariance. This measure contains measures that represent the overall fit point of view mentioned earlier. Of the various absolute fit measures, the measures that are usually used to evaluate SEM are:

1. Chi-square (X2). A low Chi-square (X 2) value that results in significance level > 0.05 or ($p > 0.05$) indicating a null hypothesis is accepted. This means that the predicted input matrix with the actual one is not statistically different. Chi-square (X 2) cannot be used the only size fits the entire model, one of the reasons is because Chi-square (X2) is sensitive to the sample size.
2. Non-Centrality Parameter (NCP). Like X2, NCP is also a measure of badness of fit where the greater the difference between Σ and $\Sigma(\Theta)$ the greater the NCP value. So, we need to look for NCPs that are small or low in value.
3. Goodness of Fit Index (GFI). GFI values range from 0 (poor fit) to 1 (perfect fit), and GFI values > 0.90 are good fit, while $0.80 < \text{GFI} < 0.90$ are often referred to as marginal fit.

4. Root Mean Square Residual (RMR). RMR represents the average value of all standardized residuals, and has about from 0 to 1. Models that have a good fit will have an RMR value smaller than 0.05.
5. Root Mean Square Error of Approximation (RMSEA). RMSEA values between 0.08 and 0.10 indicate marginal fit, and RMSEA values > 0.10 indicate poor fit.
6. Single sample cross-validation index/expected cross-validation index (ECVI). ECVI is used for model comparison and the smaller the ECVI of a model the better the degree of fit.

Incremental Match Size

Incremental match measures compare the proposed model with the baseline model often referred to as the null model or independence model and saturated model. A null model is a model with the worst fit. Saturated models are the ones with the best fit. The concept of incremental match would put the data-model match rate between the null model and the saturated model. The data-model match rate that resides between the null model and the saturated model is called a nested model. Of the various incremental match sizes, the measures typically used to evaluate SEM are:

1. Adjusted Goodness of Fit Index (AGFI). NAGFI values range from 0 to 1and AGFI values > 0.90 indicate good fit. While 0.80 < GFI < 0.90 is often referred to as marginal fit.
2. Tucker-Lewis Index / Non Normed Fit Index (TLI/NNFI). TLI values range from 0 to 1.0, with TLI values > 0.90 indicating good fit and 0.80 < TLI< 0.90 is marginal fit.
3. Normed Fit Index (NFI). This NFI has values that range from 0 to 1. NFI values > 0.90 indicate good fit, while 0.80 < NFI < 0.90 are often referred to as marginal fit.
4. Relative Fit Index (RFI). NRFI values will range from 0 to 1. An RFI value > 0.90 indicates good fit, while 0.80 < RFI < 0.90 is often referred to as marginal fit.
5. Incremental Fit Index (IFI). The value of the IFI will range from 0 to 1. An IFI value > 0.90 indicates good fit, while 0.80 < an IFI < 0.90 is often referred to as marginal fit.

6. Comparative Fit Index (CFI). CFI values will range from 0 to 1. A CFI value > 0.90 indicates a good fit, while a 0.80 < a CFI < 0.90 is often referred to as a marginal fit.

Parsimoni match size

Models with relatively few parameters (and relatively many degrees of freedom) are often known as models that have parsimony or high efficiency. Whereas a model with many parameters (and a little degree of freedom) can be said to be a complex and less parsimony model. Of the various sizes of parsimony fit, the measures that are usually used to evaluate SEM are:

1. Parsimonious Normed Fit Index (PNFI). Higher PNFI values are better. The use of PNFI is mainly for the comparison of two or more models that have different degrees of freedom. PNFI is used to compare alternative models, and there is no acceptable match rate recommendation. However, when comparing 2 models, the difference in PNFI values of 0.06 to 0.09 indicates a considerable difference in models.
2. Parsimonious goodness of fit index (PGFI). PGFI values range between 0 and 1, with higher values indicating a better parsimoni model.
3. Normed Chi-square. Recommended values: lower limit: 1.0, upper limit: 2.0 or 3.0 and looser 5.0.
4. Akaike information criterion (AIC). A small, near-zero AIC value indicates a better match, as well as a higher parsimony.

Measurement model fit (measurement model analysis)

The suitability of the measurement model is carried out by conducting validity and reliability tests. Validity relates to whether a variable measures what it should measure. Although validity can never be proved, but support towards such proof can be developed. A variable is said to have good validity to its latent construct or variable, if: the value of t load *factor* is greater than the critical value or >1.96 or practically >2), and its standard load factor (standardized loading factors) >0.70. Reliability is the consistency of a measurement. High reliability indicates that the indicator has a high consistency in measuring its latent construct. To measure reliability in SEM will be used: composite reliability measure and variance extracted measure (variant extract size).

A construct has good reliability, if its Construct Reliability (CR) value is > 0.70 and its Variance Extracted (VE) value is > 0.50.

Structural Model Fit (Structural model analysis)

Analysis of structural models includes an examination of the significance of the estimated coefficients. The SEM method and its Software not only provide the estimated values of the coefficients but also the t-calculated values for each coefficient. By specifying a significant degree (usually □□= 0.50), then each coefficient representing a hypothesized causal relationship can be statistically tested for significance. In addition to this, it is also necessary to evaluate the standard solution where all beta coefficients on multiple regressions, that is, coefficient values close to zero indicate a smaller influence. The increase in the value of this coefficient corresponds to an increase in the importance of the variable in question in the causal relationship. As a comprehensive measure of structural equations, the overall coefficient of determination (R2) is calculated as in multiple regression.

Respecification/modification and modeling strategy

Respecification is the next step after the match test is carried out. The implementation of the respecification is highly dependent on the modeling strategy to be used. There are 3 modeling strategies to choose from in SEM, namely:

1. Confirmatory modeling strategy. In this modeling strategy, a single model is formulated or specified, then empirical data collection is carried out to be tested for significance. This test will result in an acceptance or rejection of the model. This strategy does not require respecification.
2. Competing modeling strategy. In this modeling strategy several alternative models are specified and based on the analysis of a group of empirical data one of the most suitable models is selected. In this strategy, respecification is only necessary if alternative models are developed from several existing models.
3. Model development strategy. In this modeling strategy an initial model is specified and empirical data is collected. If the initial model does not match the existing data, then the model is modified and retested with the same data. Several models can be tested in this process with the aim of finding a single model that not only matches the data well, but also

has the property that each of its parameters can be interpreted properly. Respecification of the model can be done on a theory-driven or data-driven basis, however, respecation based on theory-driven is recommended

Confirmatory modeling (SC) strategies are rarely encountered, as generally researchers are not satisfied enough with simply rejecting a model without proposing an alternative model. Currently the most used in research is the model development strategy. After the model estimation is carried out, the researcher can still make modifications to the developed model if it turns out that the resulting estimate has a large residual. However, modifications can only be made if the researcher has a strong enough theoretical justification, because SEM is not intended to produce a theory, but to test a model that has a correct theoretical footing, therefore to provide an interplay as to whether the theory-based model being tested is directly acceptable or needs modification, the researcher must focus his attention on the predictive power of the model, namely by observing the residual magnitude that Generated. If in the standardized residual covariances matrix there is a value outside the range of $-2.58 <$ the residual < 2.58 and probability (P) if < 0.05 then the estimated model needs to be further modified based on the modification index by choosing the largest modification index (MI) and has a theoretical basis. The largest MI will indicate that if the coefficient is estimated, there will be a significant reduction in the value of chi square (X2). In SEM software, the modification index is listed in the output so that the researcher only needs to choose which coefficient to estimate. If the value of chi square (X2) is still not significant, the next largest MI value is sought and so on.

Conclusion C

The data analysis process using SEM starts from model specifications, model identification, model match tests and specifications/modifications and modeling strategies. modifications can only be made if the researcher has a strong enough theoretical justification, because SEM is not intended to produce a theory, but to test a model that has a correct theoretical footing, therefore to provide an interplay as to whether the theory-based model being tested is directly acceptable or needs modification, the researcher must focus his attention on the predictive power of the model, namely by observing the residual magnitude that Generated.

References D

Astrachan, C.B., Patel V. K., &Wanzenried G. (2014). A Comparative Studyof CB-SEM and PLS-SEM for Theory development in Family Firm Research. Journal of Family Business Strategy, 5, 116-128.

Babin, B. J. , Boles, J. S., & Robin, D. P. (2000). Representing the perceived ethical work climate among marketing employees. Journal of the Academy of Marketing Science. 28(3), 345-358.

Babin, B.J., Hair, J.F., & Boles, J.S. (2008). Publishing Research in Marketing Journals Using Structural Equation Modeling. Journal of Marketing Theory & Practice, 16(4), 279-285.

Bagozzi, R., & Yi, Y. 2012. Specification, evaluation, and interpretation of structural equation models. Journal of the Academy of Marketing Science, 40(1): 8-34.

Blocker, C., Flint, D., Myers, M., & Slater, S. (2011). Proactive customer orientation and its role for creating customer value in global markets. Academy of Marketing Science Journal, 39(2), 216-233.

Baker, W., E., & Syncula, J., M. (1999). Learning orientation, market orientation, and innovation: Integrating and extending models of organizational performance. Journal of Market - Focused Management, 4(4): 295-308.

Bido, D. D. S., Souza, C. A. D., Silva, D. D., Godoy, A. S., & Torres, R. R. (2012). Quality of reporting methodological procedures in national publications in the area of business administration: the case of structural equation modelling. Organizações & Sociedade, 19(60), 125-144.

Brei, V. A., & Liberali Neto, G. (2006). O uso da técnica de modelagem em equações estruturais na área de marketing: um estudo comparativo entre publicações no Brasil e no exterior. Revista de Administração Contemporânea, 10(4), 131-151.

Byrne, B. M. (2010). Structural equation modeling with AMOS: Basic concepts, applications, and programming (2nd. Edition) New York: Routledge.

Chin, W.W., Peterson, R.A., & Brown, S.P. (2008). Structural Equation Modeling in Marketing: Some Practical Reminders. Journal of Marketing Theory & Practice, 16(4), 287-289.

Churchill, G. A., Jr. (1979). A paradigm for developing better measures of marketing constructs. JMR, Journal of Marketing Research, 16(1), 64-73.

DeConinck, J. B. (2010b). The influence of ethical climate on marketing employees' job attitudes and behaviors. Journal of Business Research, 63(4): 384-391.

Delaney, J. T., & Huselid, M. A. (1996). The impact of human resource management practices on perceptions of organizational performance. Academy of Management Journal, 39(4): 949-969.

DeVellis, R. F. (2011). Scale development: Theory and applications (Vol. 26): Sage Publications, Inc.

Deshpande, R., & Farley, J. (1998). Measuring market orientation: Generalization and synthesis. Journal of Market-Focused Management, 2(3), 213-232.

Deshpande, R., & Webster Jr, F. E. (1989). Organizational culture and marketing: defining the research agenda. The Journal of Marketing, 3-15.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). Internet, mail, and mixed-mode surveys: The tailored design method (3rd ed.). Hoboken: John Wiley & Sons Inc.

Fabrigar, L. R., Porter, R. D., & Norris, M. E. 2010. Some things you should know about structural equation modeling but never thought to ask. Journal of Consumer Psychology, 20(2): 221-225.

Ferrell, O., Gonzalez-Padron, T., Hult, G., & Maignan, I. (2010). From market orientation to stakeholder orientation. Journal of Public Policy & Marketing, 2010, 29(1), 93-96.

Freeman, R. (1984). Strategic management: A stakeholder approach: Pitman Boston, MA.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research, 18(1), 39–50.

Gallagher, D., Ting, L., & Palmer, A. (2008). A journey into the unknown; taking the fear out of structural equation modeling with AMOS for the first-time user. The Marketing Review, 8(3), 255-275.

Grinstein, A. (2008). The effect of market orientation and its components on innovation consequences: A meta-analysis. Journal of the Academy of Marketing Science, 36(2), 166-173.

Reviewer Comment:

a. Authors should make more specific and attractive. Authors also should be declared the type of study clearly.
b. Authors should make abstract more specific and try to show the main concept of the review.
c. Authors should develop introduction more attractive. Authors should be tried to show the urgency by epidemiology data. Authors also should be developed main sub tittle of review in systematic, constructive and specific. Authors not only showed the standar review and no focus, try to focused in your review.
d. Authors should develop references by Vancouver style. Authors should be used the references not more than 10 years.

Revised version received by journal (December 16th, 2022)

Open Access Indonesia Journal of Social Sciences Vol 5 Issue 6 2022

Open Access Indonesia Journal of Social Sciences

Journal Homepage: https://journalsocialsciences.com/index.php/OAIJSS

Data Analysis Procedures with Structural Equation Modelling (SEM): Narrative Literature Review

Rachmat Hidayat[1]*, Patricia Wulandari[2]

[1]Department of Biology, Faculty of Medicine, Universitas Sriwijaya, Palembang, Indonesia

[2]Cattleya Mental Health Center, Palembang, Indonesia

ARTICLE INFO

Keywords:
Data analysis
Structural equation modelling
Variable

***Corresponding author:**
Rachmat Hidayat

E-mail address:
rachmathidayat@fk.unsri.ac.id

All authors have reviewed and approved the final version of the manuscript.

https://doi.org/10.37275/oaijss.v5i6.142

ABSTRACT

The relationship between variables in structural equation modeling (SEM) forms a structural model. This structural model can be explained through structural equations, such as in regression analysis. This structural equation describes the prediction of the latent (exogenous) independent variable on the latent (endogenous) dependent variable. Researchers who use analysis with structural equation models need to know whether the model built with empirical data has a unique value or not so that the model can be estimated. If the model does not have a unique value, then the model cannot be identified (unidentified). The cause of a model is categorized as unidentified because the information contained in empirical data is not sufficient to produce a unique solution in calculating model estimation parameters. This literature review aims to describe the process of data analysis using SEM.

1. Introduction

There is a principal difference between regression analysis and path (path analysis) and SEM in terms of measuring variables. In the regression analysis, the dependent and independent variables are variables that can be measured directly (observable), whereas, in SEM, the dependent and independent variables are variables that cannot be measured directly (unobservable). Unobserved variables are also often called latent variables. The structural equation model or SEM is a model that explains the relationship between latent variables, so the SEM model is often referred to as latent variable analysis or linear structural relationship. The relationship between variables in SEM is the same as the relationship in path analysis (Astrachan, 2014). However, in explaining the relationship between latent variables, the SEM model differs from path analysis, where path analysis uses observable variables while SEM uses unobservable variables (Babin et al., 2008). The relationship between variables in SEM forms a structural model. This structural model can be explained through structural equations, such as in regression analysis. This structural equation describes the prediction of the latent (exogenous) independent variable on the latent (endogenous) dependent variable.

Model specifications

SEM begins by specifying the research model. The analysis will not begin until the researcher specifies a model that shows the relationship between the variables to be analyzed. Through the steps below, researchers can obtain the desired model: 1) Define the latent variables of the study, 2) Define the observed variable, and 3) Define the relationship between latent variables and observed variables. (Bagozzi et al., 2012) Pay attention to aspects of the unidimensional or multidimensional construct variable. Unidimensional constructs (first-order constructs) are constructs that directly describe the relationship between latent variables and observed variables reflectively (arrows away from latent variables) or formatively (arrows towards latent variables). Multidimensional constructs (second-order constructs) are constructs formed from several unidimensional constructs.

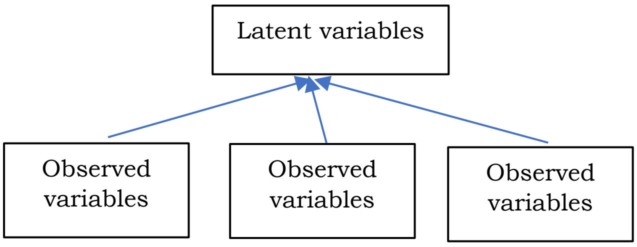


Figure 1. Unidimensional construct (First order construct).

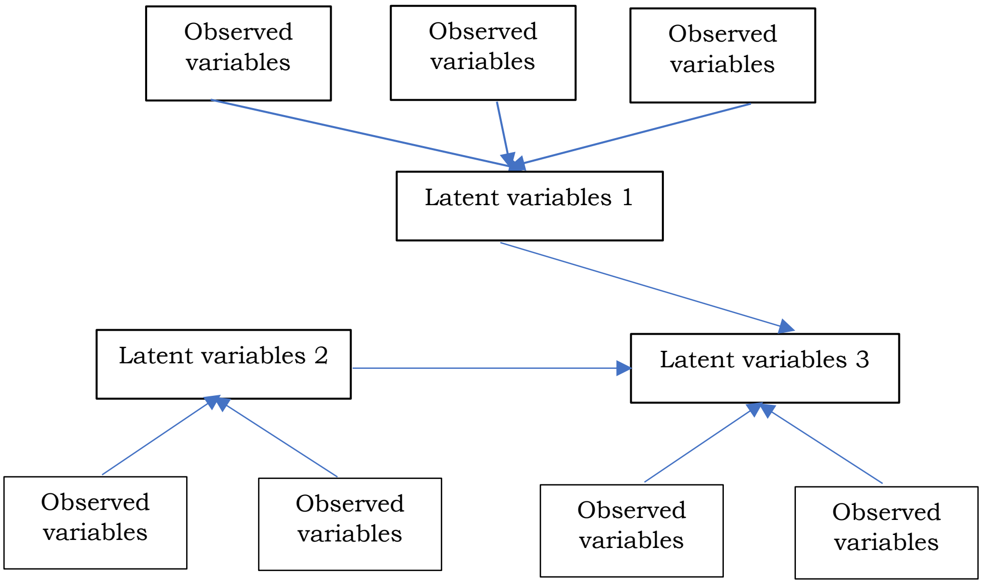


Figure 2. Multidimensional construct (Second order construct).

Model identification

Researchers who use analysis with structural equation models need to know whether the model built with empirical data has a unique value or not so that the model can be estimated. If the model does not have a unique value, then the model cannot be identified (unidentified). The cause of a model is categorized as unidentified because the information contained in empirical data is not sufficient to produce a unique solution in calculating model estimation parameters (Baker, 1999; Bido et al., 2012). An example of an under-identified case is A x B = 1000. The question is, what is the value of A or B? To determine the value of A or B, of course, the answers vary widely. It can be 100 x 10, 500 x 2, 250 x 4, 200 x 5 or 1 x 1000. To ensure an answer, we must choose the most appropriate (unique) answer, called model identification. This example also occurs in SEM, where the theoretical model built and empirical data are not sufficient to produce a unique solution in calculating model parameter estimates (Byrne, 2010). However, if we determine the value of A = 100, then the value of B = 10 will be automatic. This can also be done in SEM analysis to overcome the unidentified model by constraining the model by 1) Adding indicators or manifest variables from latent constructs, 2) Determining the value of additional fixed parameters so that the calculation of the degree of freedom becomes positive (this method is most often used by researchers), 3) Assuming that the parameters with each other have the same value.

It should be noted that the use of the three methods above to change an under-identified model must be, in theory, not merely done so that the model can be identified. There are three possible identification models in SEM: 1) Under-identified model, where the value of $t \geq s/2$; namely, a model with a greater number of estimated parameters than the known amount of data (the data is the variance and covariance of the observed variables). For example, there is the equation X + Y = 10, representing 1 known piece of data and 2 parameters to be estimated, namely X and Y, so the number of df = 1-2 = -1. From the understanding of the unidentified model in SEM, it has df = the amount of data that is known - the number of parameters estimated < 0. So it can be concluded that the under-identified model has a negative df, 2) Just-Identified model, where $t = s/2$; i.e., a model with the same number of estimated parameters as the known data. For example, there are two equations X + Y = 10 and X + 2Y = 16, which are 2 known pieces of data and 2 parameters to be estimated, namely X and Y, then the number of df = 2-2 = 0. So it can be concluded that the model just - identified has a zero df, 3) Over-Identified model, where $t \leq s/2$; namely, a model with a smaller number of estimated parameters than the known amount of data (Churchill, 1979). For example, there are three equations X + Y = 10, X + 2Y = 16 and 3X + 2Y = 22, which are 3 known pieces of data and 2 parameters to be estimated, namely X and Y, so the number of df = 3-2 = 1. So it can be concluded that the model is over-identified *df* positive.

Information:
t = number of parameters estimated,
s = total variance and covariance between indicators.

The goodness of fit (Assessment of fit)

Overall model fit

The first stage of the fit test is intended to evaluate, in general, the degree of fit or goodness of fit (GOF) between the data and the model. Assessing the overall fit of the model in SEM cannot be done directly as in other multivariate techniques (multiple regression, discriminant analysis, MANOVA, etc.). SEM does not have a single best statistical test that can explain the predictive power of the model (Delaney, 1996). Instead, researchers have developed several measures of GOF or goodness of fit indices (GOFI) that can be used together or in combination. This situation causes the thorough compatibility test stage, which is a step that invites a lot of debate and controversy. The controversy and debate around GOF arise when the question of size arises in the discussion, i.e., what is the acceptable level of fit? Despite the controversy, there is finally a consensus among researchers, some of which are: 1) The best guide in assessing model fit is strong substantive theory. If the model only shows or represents a substantive theory that is not strong, and even though the model has a very good fit, it is rather difficult for us to judge the model, 2) The Chi-square (X2) statistical test should not be the only basis for determining the fit of the data to the model, 3) None of the GOF or GOF Indices (GOFI) measures can be used exclusively as a basis for evaluating the fit of the entire model. GOFI is grouped into three parts, namely absolute fit measures, incremental fit measures, and parsimonious fit measures (parsimony fit size).

Absolute fit measures

The absolute fit measure determines the degree of prediction of the overall model (structural and measurement models) on the correlation and covariance matrices Deshpande, 1989; Deshpande, 1998). This measure contains measures that represent the overall fit point of view mentioned earlier. Of the various absolute fit measures, the measures usually used to evaluate SEM are 1) Chi-square (X2). A low Chi-square value (X2) results in a significance level > 0.05 or (p > 0.05), which indicates the null hypothesis is accepted. This means that the predicted input matrix is not statistically different from the actual one. Chi-square (X2) cannot be used as the sole measure of the overall fit of the model. One of the reasons is that Chi-square (X2) is sensitive to sample size. 2) Non-Centrality Parameters (NCP). Like X2, NCP is also a measure of the badness of fit where the greater the difference between Σ and $\Sigma(\theta)$ the more**,** the greater the NCP value. So, we need to find an NCP whose value is small or low. 3) The goodness of fit index (GFI). The GFI value ranges from 0 (poor fit) to 1 (perfect fit), and a GFI value > 0.90 is a good fit, while 0.80 < GFI < 0.90 is often referred to as marginal fit. 4) Root Mean Square Residual (RMR). RMR represents the average value of all standardized residuals and ranges from 0 to 1. A model with a good fit will have an RMR value smaller than 0.05. 5) Root Mean Square Error of Approximation (RMSEA). An RMSEA value between 0.08 and 0.10 indicates marginal fit, and an RMSEA value > 0.10 indicates poor fit. 6) Single Sample Cross-Validation Index/Expected Cross-Validation Index (ECVI). ECVI is used for model comparison, and the smaller the ECVI of a model, the better the level of fit.

Incremental fit measures

The incremental fit measure compares the proposed model to the baseline model, which is often referred to as the null model or the independence model, and the saturated model. The null model is the model with the worst fit of the model data (worst fit). A saturated model is the best fit for the model data (best fit). The concept of incremental fit will place the model-data match level between the null model and the saturated model. The level of model-data compatibility that is between the null model and the saturated model is called a nested model. Of the various incremental fit measures, the measures usually used to evaluate SEM are 1) Adjusted Goodness of Fit Index (AGFI). AGFI values range from 0 to 1, and AGFI values > 0.90 indicate a good fit. Whereas 0.80 < GFI < 0.90 is often referred to as marginal fit. 2) tucker-lewis index/non-normal fit index (TLI/NNFI). TLI values ranged from 0 to 1.0, with TLI values > 0.90 indicating good fit and 0.80 < TLI < 0.90 indicating marginal fit. 3) normed fit index (NFI). This NFI has values ranging from 0 to 1. An NFI value > 0.90 indicates a good fit, while 0.80 < NFI < 0.90 is often referred to as a marginal fit. 4) relative fit index (RFI). The RFI value will range from 0 to 1. An RFI value > 0.90 indicates a good fit, while 0.80 < RFI < 0.90 is often referred to as a marginal fit. 5) incremental fit index (IFI). IFI values will range from 0 to 1. IFI values > 0.90 indicate a good fit, while 0.80 < IFI < 0.90 is often referred to as a marginal fit. 6) comparative fit index (CFI). CFI values will range from 0 to 1. CFI values > 0.90 indicate good fit, while 0.80 < CFI < 0.90 are often referred to as marginal fit.

Parsimony fit measures

Models with relatively few parameters (and relatively many degrees of freedom) are often known as models that have high parsimony or frugality. Meanwhile, a model with many parameters (and few degrees of freedom) can be said to be a model that is complex and lacks parsimony. Of the various parsimony fit measures, the measures that are usually used to evaluate SEM are: 1) Parsimonious Normal Fit Index (PNFI). The higher the PNFI value, the better. The use of PNFI is primarily for comparisons of two or more models that have different degrees of freedom. PNFI was used to compare alternative models, and no acceptable match level is recommended. However, when comparing the 2 models, the difference in the PNFI value of 0.06 to 0.09 indicates a fairly large model difference. 2) Parsimonious goodness of fit index (PGFI). PGFI values range between 0 and 1, with higher values indicating a better parsimony model. 3) Normed Chi-square. Recommended values: lower limit: 1.0, upper limit: 2.0 or 3.0, and looser 5.0. 4) Akaike information criterion (AIC). A small AIC value close to zero indicates a better fit and higher parsimony.

Measurement model fit (measurement model analysis)

Measurement model fit is carried out by testing its validity and reliability. Validity relates to whether variable measures what it is supposed to measure. Although validity can never be proven, support for such evidence can be developed. A variable is said to have good validity against its construct or latent variable if the factor loading t value (loading factors) is greater than the critical value or >1.96 or for practice >2), and standardized loading factors >0.70. Reliability is the consistency of measurement. High reliability indicates that the indicator has high consistency in measuring its latent constructs. To measure the reliability in SEM will be used: the composite reliability measure (composite reliability measure) and variance extracted measure (variance extract size). A construct has good reliability if the Construct Reliability (CR) value is > 0.70 and the variance Extracted value is (VE)> 0.50.

Structural model fit (Structural model analysis)

Analysis of the structural model includes examining the significance of the estimated coefficients. The SEM method and its software provide not only the estimated coefficients but also the t-count values for each coefficient. By specifying a significant level (usually α= 0.50), then each coefficient representing the hypothesized causal relationship can be tested for statistical significance. In addition to this, it is also necessary to evaluate the standard solution where all the beta coefficients are in multiple regression, that is, the coefficient value that is close to zero indicates a smaller effect. An increase in the value of this coefficient is associated with an increase in the importance of the variable in question in a causal relationship. As an overall measure of the structural equation, the overall coefficient of determination (R2) is calculated as in multiple regression.

Respecification/modification and modeling strategy

Re-specification is the next step after the compatibility test is carried out. The implementation of respecification is highly dependent on the modeling strategy to be used. There are 3 modeling strategies that can be chosen in SEM, namely: 1) Confirmatory modeling strategy or confirmatory modeling strategy. In this modeling strategy, a single model is formulated or specified, and then empirical data is collected to test its significance. This test will result in an acceptance or rejection of the model. This strategy does not require respecification. 2) Model competition strategy or competing for modeling strategy. In this modeling strategy, several alternative models are specified, and based on an analysis of a group of empirical data, one of the most suitable models is selected. In this strategy, respecification is only needed if alternative models are developed from several existing models. 3) Model development strategy or model development strategy. In this modeling strategy, an initial model is specified, and empirical data is collected. If the initial model does not match the existing data, then the model is modified and tested again with the same data. Several models can be tested in this process with the aim of finding a model that not only fits the data well but also has the property that each parameter can be interpreted properly. Respecification of the model can be done based on theory-driven or data-driven.

However, respecification based on theory-driven is more recommended (Freeman, 1984).

A confirmatory modeling strategy (CS) is rarely encountered because, generally, researchers are not satisfied with simply rejecting a model without proposing an alternative model. Currently, the most widely used in research is the model development strategy. After the model estimation is done, the researcher can still modify the developed model if it turns out that the resulting estimate has a large residual. However, modifications can only be made if the researcher has sufficiently strong theoretical justification because SEM is not intended to generate theory but to test models that have a correct theoretical basis, therefore providing an interpretation of whether the theory-based model being tested can be accepted directly or not (Fornell, 1981). If modifications are needed, researchers must direct their attention to the predictive power of the model by observing the number of residuals produced. If in the standardized residual covariances matrix, there are values outside the range of -2.58 < residual < 2.58 and probability (P) if < 0.05, then the estimated model needs to be further modified based on the modified index by selecting the modification index (MI) is the largest and has a theoretical basis. The largest MI will give an indication that if the coefficient is estimated, there will be a significant reduction in the chi-square (X2) value (Grinstein, 2008). In SEM software, the modification index is included in the output so that the researcher only has to choose which coefficient to estimate. If the chi-square value (X2) is still not significant, look for the next largest MI value and so on.

2. Conclusion

The process of data analysis using SEM starts with model specification, model identification, model suitability testing, and respecification/modification and modeling strategy. modifications can only be made if the researcher has a sufficiently strong theoretical justification because SEM is not intended to generate theory but to test models that have a correct theoretical basis, therefore providing an interpretation of whether the theory-based model being tested can be directly accepted or needs modification. Then the researcher must direct his attention to the predictive power of the model by observing the amount of residual generated.

3. References

Astrachan CB, Patel VK, Wanzenried G. 2014. A comparative study of CB-SEM and PLS-SEM for theory development in family firm research. Journal of Family Business Strategy. 5; 116-28.

Babin BJ, Boles JS, Robin DP. 2000. Representing the perceived ethical work climate among marketing employees. Journal of the Academy of Marketing Science. 28(3): 345-58.

Babin BJ, Hair JF, Boles JS. 2008. Publishing research in marketing journals using structural equation modeling. Journal of Marketing Theory & Practice. 16(4): 279-285.

Bagozzi R, Yi Y. 2012. Specification, evaluation, and interpretation of structural equation models. Journal of the Academy of Marketing Science, 40(1): 8-34.

Blocker C, Flint D, Myers M, Slater S. 2011. Proactive customer orientation and its role for creating customer value in global markets. Academy of Marketing Science Journal. 39(2): 216-33.

Baker WE, Sinkula JM. 1999. Learning orientation, market orientation, and innovation: Integrating and extending models of organizational performance. Journal of Market-Focused Management. 4(4): 295-308.

Bido DDS, Souza CAD, Silva DD, Godoy AS, Torres RR. 2012. Quality of reporting methodological procedures in national publications in the area of business administration: the case of structural equation modelling. Organizações & Sociedade, 19(60): 125-44.

Brei VA, Liberali NG. 2006. O uso da técnica de modelagem em equações estruturais na área de marketing: um estudo comparativo entre publicações no Brasil e no exterior. Revista de Administração Contemporânea. 10(4): 131-51.

Byrne BM. 2010. Structural equation modeling with AMOS: Basic concepts, applications, and programming, 2nd ed. New York: Routledge.

Chin WW, Peterson RA, Brown SP. 2008. Structural equation modeling in marketing: some practical reminders. Journal of Marketing Theory & Practice, 16(4): 287-9.

Churchill GAJr. 1979. A paradigm for developing better measures of marketing constructs. JMR, Journal of Marketing Research, 16(1): 64-73.

DeConinck JB. 2010b. The influence of ethical climate on marketing employees' job attitudes and behaviors. Journal of Business Research, 63(4): 384-91.

Delaney JT, Huselid MA. 1996. The impact of human resource management practices on perceptions of organizational performance. Academy of Management Journal, 39(4): 949-69.

DeVellis RF. 2011. Scale development: Theory and applications. Sage Publications, Inc. 26.

Deshpande R, Farley J. 1998. Measuring market orientation: Generalization and synthesis. Journal of Market-Focused Management, 2(3): 213-32.

Deshpande R, Webster JrFE. 1989. Organizational culture and marketing: defining the research agenda. The Journal of Marketing. 3-15.

Dillman DA, Smyth JD, Christian LM. 2009. Internet, mail, and mixed-mode surveys: The tailored design method. 3rd ed. Hoboken: John Wiley & Sons Inc.

Fabrigar LR, Porter RD, Norris ME. 2010. Some things you should know about structural equation modeling but never thought to ask. Journal of Consumer Psychology. 20(2): 221-5.

Ferrell O, Gonzalez-Padron T, Hult G, Maignan I. 2010. From market orientation to stakeholder orientation. Journal of Public Policy & Marketing, 29(1): 93-6.

Freeman R. 1984. Strategic management: A stakeholder approach: Pitman Boston, MA.

Fornell C, Larcker DF. 1981. Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research. 18(1): 39–50.

Gallagher D, Ting L, Palmer A. 2008. A journey into the unknown; taking the fear out of structural equation modeling with AMOS for the first-time user. The Marketing Review. 8(3): 255-75.

Grinstein A. 2008. The effect of market orientation and its components on innovation consequences: A meta-analysis. Journal of the Academy of Marketing Science. 36(2): 166-73.

Paper Accepted for publication (December 19th, 2022)

Open Access Indonesia Journal of Social Sciences

HM Publisher

Letter of Acceptance

Manuscript “Data Analysis Procedures with Structural Equation Modelling (SEM): Narrative Literature Review“ by Rachmat Hidayat*, Patricia Wulandari, has been accepted to publish in Open Access Indonesia Journal of Social Sciences Vol 5 issue 6 in December 2022.

Cordially,

Prof. Paula Magnano, PhD
Editor

HM Publisher

(*) Corresponding author
The Coresponding Author can access the acount in website :
https://www.journalsocialsciences.com/index.php/oaijss/login
User: rachmathidayat
Password: 210587

Galley proof (December 21th, 2022)

Open Access Indonesia Journal of Social Sciences Vol 5 Issue 6 2022

Open Access Indonesia Journal of Social Sciences

Journal Homepage: https://journalsocialsciences.com/index.php/OAIJSS

Data Analysis Procedures with Structural Equation Modelling (SEM): Narrative Literature Review

Rachmat Hidayat[1]*, Patricia Wulandari[2]

[1]Department of Biology, Faculty of Medicine, Universitas Sriwijaya, Palembang, Indonesia

[2]Cattleya Mental Health Center, Palembang, Indonesia

ARTICLE INFO

Keywords:
Data analysis
Structural equation modelling
Variable

***Corresponding author:**
Rachmat Hidayat

E-mail address:
rachmathidayat@fk.unsri.ac.id

All authors have reviewed and approved the final version of the manuscript.

https://doi.org/10.37275/oaijss.v5i6.142

A B S T R A C T

The relationship between variables in structural equation modeling (SEM) forms a structural model. This structural model can be explained through structural equations, such as in regression analysis. This structural equation describes the prediction of the latent (exogenous) independent variable on the latent (endogenous) dependent variable. Researchers who use analysis with structural equation models need to know whether the model built with empirical data has a unique value or not so that the model can be estimated. If the model does not have a unique value, then the model cannot be identified (unidentified). The cause of a model is categorized as unidentified because the information contained in empirical data is not sufficient to produce a unique solution in calculating model estimation parameters. This literature review aims to describe the process of data analysis using SEM.

1. Introduction

There is a principal difference between regression analysis and path (path analysis) and SEM in terms of measuring variables. In the regression analysis, the dependent and independent variables are variables that can be measured directly (observable), whereas, in SEM, the dependent and independent variables are variables that cannot be measured directly (unobservable). Unobserved variables are also often called latent variables. The structural equation model or SEM is a model that explains the relationship between latent variables, so the SEM model is often referred to as latent variable analysis or linear structural relationship. The relationship between variables in SEM is the same as the relationship in path analysis (Astrachan, 2014). However, in explaining the relationship between latent variables, the SEM model differs from path analysis, where path analysis uses observable variables while SEM uses unobservable variables (Babin et al., 2008). The relationship between variables in SEM forms a structural model. This structural model can be explained through structural equations, such as in regression analysis. This structural equation describes the prediction of the latent (exogenous) independent variable on the latent (endogenous) dependent variable.

Model specifications

SEM begins by specifying the research model. The analysis will not begin until the researcher specifies a model that shows the relationship between the variables to be analyzed. Through the steps below, researchers can obtain the desired model: 1) Define the latent variables of the study, 2) Define the observed variable, and 3) Define the relationship between latent variables and observed variables. (Bagozzi et al., 2012) Pay attention to aspects of the unidimensional or multidimensional construct variable. Unidimensional constructs (first-order constructs) are constructs that directly describe the relationship between latent variables and observed variables reflectively (arrows away from latent variables) or formatively (arrows towards latent variables). Multidimensional constructs (second-order constructs) are constructs formed from several unidimensional constructs.

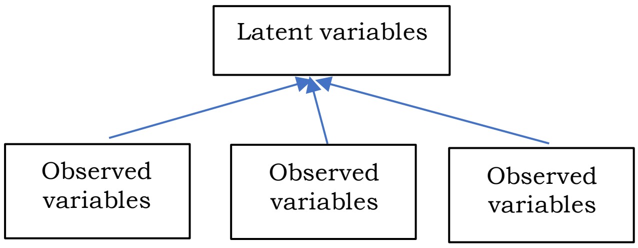


Figure 1. Unidimensional construct (First order construct).

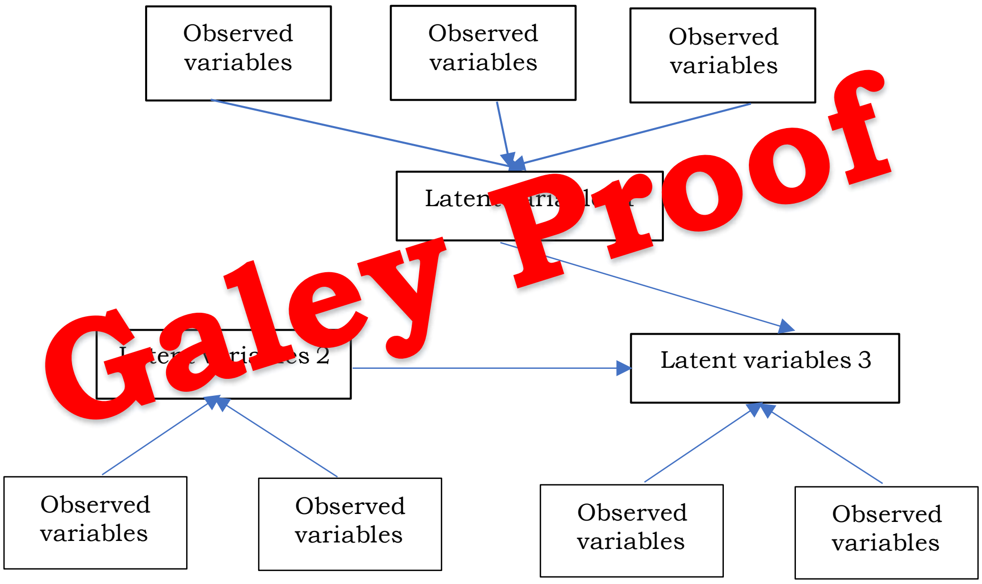


Figure 2. Multidimensional construct (Second order construct).

Model identification

Researchers who use analysis with structural equation models need to know whether the model built with empirical data has a unique value or not so that the model can be estimated. If the model does not have a unique value, then the model cannot be identified (unidentified). The cause of a model is categorized as unidentified because the information contained in empirical data is not sufficient to produce a unique solution in calculating model estimation parameters (Baker, 1999; Bido et al., 2012). An example of an under-identified case is A x B = 1000. The question is, what is the value of A or B? To determine the value of A or B, of course, the answers vary widely. It can be 100 x 10, 500 x 2, 250 x 4, 200 x 5 or 1 x 1000. To ensure an answer, we must choose the most appropriate (unique) answer, called model identification. This example also occurs in SEM, where the theoretical model built and empirical data are not sufficient to produce a unique solution in calculating model parameter estimates (Byrne, 2010). However, if we determine the value of A = 100, then the value of B = 10 will be automatic. This can also be done in SEM analysis to overcome the unidentified model by constraining the model by 1) Adding indicators or manifest variables from latent constructs, 2) Determining the value of additional fixed parameters so that the calculation of the degree of freedom becomes positive (this method is most often used by researchers), 3) Assuming that the parameters with each other have the same value.

It should be noted that the use of the three methods above to change an under-identified model must be, in theory, not merely done so that the model can be identified. There are three possible identification models in SEM: 1) Under-identified model, where the value of t ≥ s/2; namely, a model with a greater number of estimated parameters than the known amount of data (the data is the variance and covariance of the observed variables. For example, there is the equation X + Y = 10, representing 1 known piece of data and 2 parameters to be estimated, namely X and Y, so the number of df = 1-2 = -1. From the understanding of the unidentified model in SEM, it has df = the amount of data that is known - the number of parameters estimated < 0. So it can be concluded that the under-identified model has a negative df, 2) Just-Identified model, where t = s/2; i.e., a model with the same number of estimated parameters as the known data. For example, there are two equations X + Y = 10 and X + 2Y = 16, which are 2 known pieces of data and 2 parameters to be estimated, namely X and Y, then the number of df = 2-2 = 0. So it can be concluded that the model just -identified has a zero df, 3) Over-Identified model, where t ≤ s/2; namely, a model with a smaller number of estimated parameters than the known amount of data (Churchill, 1979). For example, there are three equations X + Y = 10, X + 2Y = 16 and 3X + 2Y = 22, which are 3 known pieces of data and 2 parameters to be estimated, namely X and Y, so the number of df = 3-2 = 1. So it can be concluded that the model is over-identified *df* positive.

Information:

t = number of parameters estimated,

s = total variance and covariance between indicators.

The goodness of fit (Assessment of fit)

Overall model fit

The first stage of the fit test is intended to evaluate, in general, the degree of fit or goodness of fit (GOF) between the data and the model. Assessing the overall fit of the model in SEM cannot be done directly as in other multivariate techniques (multiple regression, discriminant analysis, MANOVA, etc.). SEM does not have a single best statistical test that can explain the predictive power of the model (Delaney, 1996). Instead, researchers have developed several measures of GOF or goodness of fit indices (GOFI) that can be used together or in combination. This situation causes the thorough compatibility test stage, which is a step that invites a lot of debate and controversy. The controversy and debate around GOF arise when the question of size arises in the discussion, i.e., what is the acceptable level of fit? Despite the controversy, there is finally a consensus among researchers, some of which are: 1) The best guide in assessing model fit is strong substantive theory. If the model only shows or represents a substantive theory that is not strong, and even though the model has a very good fit, it is rather difficult for us to judge the model, 2) The Chi-square (X2) statistical test should not be the only basis for determining the fit of the data to the model, 3) None of the GOF or GOF Indices (GOFI) measures can be used exclusively as a basis for evaluating the fit of the entire model. GOFI is grouped into three parts, namely absolute fit measures, incremental fit measures, and parsimonious fit measures (parsimony fit size).

Absolute fit measures

The absolute fit measure determines the degree of prediction of the overall model (structural and measurement models) on the correlation and covariance matrices Deshpande, 1989; Deshpande, 1998). This measure contains measures that represent the overall fit point of view mentioned earlier. Of the various absolute fit measures, the measures usually used to evaluate SEM are 1) Chi-square (X2). A low Chi-square value (X2) results in a significance level > 0.05 or (p > 0.05), which indicates the null hypothesis is accepted. This means that the predicted input matrix is not statistically different from the actual one. Chi-square (X2) cannot be used as the sole measure of the overall fit of the model. One of the reasons is that Chi-square (X2) is sensitive to sample size. 2) Non-Centrality Parameters (NCP). Like X2, NCP is also a measure of the badness of fit where the greater the difference between Σ and Σ(θ) the more, the greater the NCP value. So, we need to find an NCP whose value is small or low. 3) The goodness of fit index (GFI). The GFI value ranges from 0 (poor fit) to 1 (perfect fit), and a GFI value > 0.90 is a good fit, while 0.80 < GFI < 0.90 is often referred to as marginal fit. 4) Root Mean Square Residual (RMR). RMR represents the average value of all standardized residuals and ranges from 0 to 1. A model with a good fit will have an RMR value smaller than 0.05. 5) Root Mean Square Error of Approximation (RMSEA). An RMSEA value between 0.08 and 0.10 indicates marginal fit, and an RMSEA value > 0.10 indicates poor fit. 6) Single Sample Cross-Validation Index/Expected Cross-Validation Index (ECVI). ECVI is used for model comparison, and the smaller the ECVI of a model, the better the level of fit.

Incremental fit measures

The incremental fit measure compares the proposed model to the baseline model, which is often referred to as the null model or the independence model, and the saturated model. The null model is the model with the worst fit of the model data (worst fit). A saturated model is the best fit for the model data (best fit). The concept of incremental fit will place the model-data match level between the null model and the saturated model. The level of model-data compatibility that is between the null model and the saturated model is called a nested model. Of the various incremental fit measures, the measures usually used to evaluate SEM are 1) Adjusted Goodness of Fit Index (AGFI). AGFI values range from 0 to 1, and AGFI values > 0.90 indicate a good fit. Whereas 0.80 < GFI < 0.90 is often referred to as marginal fit. 2) tucker-lewis index/non-normal fit index (TLI/NNFI). TLI values ranged from 0 to 1.0, with TLI values > 0.90 indicating good fit and 0.80 < TLI < 0.90 indicating marginal fit. 3) normed fit index (NFI). This NFI has values ranging from 0 to 1. An NFI value > 0.90 indicates a good fit, while 0.80 < NFI < 0.90 is often referred to as a marginal fit. 4) relative fit index (RFI). The RFI value will range from 0 to 1. An RFI value > 0.90 indicates a good fit, while 0.80 < RFI < 0.90 is often referred to as a marginal fit. 5) incremental fit index (IFI). IFI values will range from 0 to 1. IFI values > 0.90 indicate a good fit, while 0.80 < IFI < 0.90 is often referred to as a marginal fit. 6) comparative fit index (CFI). CFI values will range from 0 to 1. CFI values > 0.90 indicate good fit, while 0.80 < CFI < 0.90 are often referred to as marginal fit.

Parsimony fit measures

Models with relatively few parameters (and relatively many degrees of freedom) are often known as models that have high parsimony or frugality. Meanwhile, a model with many parameters (and few degrees of freedom) can be said to be a model that is complex and lacks parsimony. Of the various parsimony fit measures, the measures that are usually used to evaluate SEM are: 1) Parsimonious Normal Fit Index (PNFI). The higher the PNFI value, the better. The use of PNFI is primarily for comparisons of two or more models that have different degrees of freedom. PNFI was used to compare alternative models, and no acceptable match level is recommended. However, when comparing the 2 models, the difference in the PNFI value of 0.06 to 0.09 indicates a fairly large model difference. 2) Parsimonious goodness of fit index (PGFI). PGFI values range between 0 and 1, with higher values indicating a better parsimony model. 3) Normed Chi-square. Recommended values: lower limit: 1.0, upper limit: 2.0 or 3.0, and looser 5.0. 4) Akaike information criterion (AIC). A small AIC value close to zero indicates a better fit and higher parsimony.

Measurement model fit (measurement model analysis)

Measurement model fit is carried out by testing its validity and reliability. Validity relates to whether variable measures what it is supposed to measure. Although validity can never be proven, support for such evidence can be developed. A variable is said to have good validity against its construct or latent variable if the factor loading t value (loading factors) is greater than the critical value or >1.96 or for practice >2), and standardized loading factor > 0.70. Reliability is the consistency of measurement. High reliability indicates that the indicator has high consistency in measuring its latent constructs. To measure the reliability in SEM will be used: the composite reliability measure (composite reliability measure) and variance extracted measure (variance extract size). A construct has good reliability if the Construct Reliability (CR) value is > 0.70 and the variance Extracted value is (VE)> 0.50.

Structural model fit (Structural model analysis)

Analysis of the structural model includes examining the significance of the estimated coefficients. The SEM method and its software provide not only the estimated coefficients but also the t-count values for each coefficient. By specifying a significant level (usually α= 0.50), then each coefficient representing the hypothesized causal relationship can be tested for statistical significance. In addition to this, it is also necessary to evaluate the standard solution where all the beta coefficients are in multiple regression, that is, the coefficient value that is close to zero indicates a smaller effect. An increase in the value of this coefficient is associated with an increase in the importance of the variable in question in a causal relationship. As an overall measure of the structural equation, the overall coefficient of determination (R2) is calculated as in multiple regression.

Respecification/modification and modeling strategy

Re-specification is the next step after the compatibility test is carried out. The implementation of respecification is highly dependent on the modeling strategy to be used. There are 3 modeling strategies that can be chosen in SEM, namely: 1) Confirmatory modeling strategy or confirmatory modeling strategy. In this modeling strategy, a single model is formulated or specified, and then empirical data is collected to test its significance. This test will result in an acceptance or rejection of the model. This strategy does not require respecification. 2) Model competition strategy or competing for modeling strategy. In this modeling strategy, several alternative models are specified, and based on an analysis of a group of empirical data, one of the most suitable models is selected. In this strategy, respecification is only needed if alternative models are developed from several existing models. 3) Model development strategy or model development strategy. In this modeling strategy, an initial model is specified, and empirical data is collected. If the initial model does not match the existing data, then the model is modified and tested again with the same data. Several models can be tested in this process with the aim of finding a model that not only fits the data well but also has the property that each parameter can be interpreted properly. Respecification of the model can be done based on theory-driven or data-driven.

However, respecification based on theory-driven is more recommended (Freeman, 1984).

A confirmatory modeling strategy (CS) is rarely encountered because, generally, researchers are not satisfied with simply rejecting a model without proposing an alternative model. Currently, the most widely used in research is the model development strategy. After the model estimation is done, the researcher can still modify the developed model if it turns out that the resulting estimate has a large residual. However, modifications can only be made if the researcher has sufficiently strong theoretical justification because SEM is not intended to generate theory but to test models that have a correct theoretical basis, therefore providing an interpretation of whether the theory-based model being tested can be accepted directly or not (Fornell, 1981). If modifications are needed, researchers must direct their attention to the predictive power of the model by observing the number of residuals produced. If the standardized residual covariance matrix there are values outside the range of -2.58 < residual < 2.58 and probability (P) if < 0.05, then the estimated model needs to be further modified based on the modified index by selecting the modification index (MI) is the largest and has a theoretical basis. The largest MI will give an indication that if the coefficient is estimated, there will be a significant reduction in the chi-square (X2) value (Grinstein, 2008). In SEM software, the modification index is included in the output so that the researcher only has to choose which coefficient to estimate. If the chi-square value (X2) is still not significant, look for the next largest MI value and so on.

2. Conclusion

The process of data analysis using SEM starts with model specification, model identification, model suitability testing, and respecification/modification and modeling strategy. modifications can only be made if the researcher has a sufficiently strong theoretical justification because SEM is not intended to generate theory but to test models that have a correct theoretical basis, therefore providing an interpretation of whether the theory-based model being tested can be directly accepted or needs modification. Then the researcher must direct his attention to the predictive power of the model by observing the amount of residual generated.

3. References

Astrachan CB, Patel VK, Wanzenried G. 2014. A comparative study of CB-SEM and PLS-SEM for theory development in family firm research. Journal of Family Business Strategy. 5; 116-28.

Babin BJ, Boles JS, Robin DP. 2000. Representing the perceived ethical work climate among marketing employees. Journal of the Academy of Marketing Science. 28(3): 345-58.

Babin BJ, Hair JF, Boles JS. 2008. Publishing research in marketing journals using structural equation modeling. Journal of Marketing Theory & Practice. 16(4): 279-285.

Bagozzi R, Yi Y. 2012. Specification, evaluation, and interpretation of structural equation models. Journal of the Academy of Marketing Science, 40(1): 8-34.

Blocker C, Flint D, Myers M, Slater S. 2011. Proactive customer orientation and its role for creating customer value in global markets. Academy of Marketing Science Journal. 39(2): 216-33.

Baker WE, Sinkula JM. 1999. Learning orientation, market orientation, and innovation: Integrating and extending models of organizational performance. Journal of Market-Focused Management. 4(4): 295-308.

Bido DDS, Souza CAD, Silva DD, Godoy AS, Torres RR. 2012. Quality of reporting methodological procedures in national publications in the area of business administration: the case of structural equation modelling. Organizações & Sociedade, 19(60): 125-44.

Brei VA, Liberali NG. 2006. O uso da técnica de modelagem em equações estruturais na área de marketing: um estudo comparativo entre publicações no Brasil e no exterior. Revista de Administração Contemporânea. 10(4): 131-51.

Byrne BM. 2010. Structural equation modeling with AMOS: Basic concepts, applications, and programming, 2nd ed. New York: Routledge.

Chin WW, Peterson RA, Brown SP. 2008. Structural equation modeling in marketing: some practical reminders. Journal of Marketing Theory & Practice, 16(4): 287-9.

Churchill GAJr. 1979. A paradigm for developing better measures of marketing constructs. JMR, Journal of Marketing Research, 16(1): 64-73.

DeConinck JB. 2010b. The influence of ethical climate on marketing employees' job attitudes and behaviors. Journal of Business Research, 63(4): 384-91.

Delaney JT, Huselid MA. 1996. The impact of human resource management practices on perceptions of organizational performance. Academy of Management Journal. 39(4): 949-69.

DeVellis RF. 2011. Scale development: Theory and applications. Sage Publications, Inc. 26.

Deshpande R, Farley J. 1998. Measuring market orientation: Generalization and synthesis. Journal of Market-Focused Management, 2(3): 213-32.

Deshpande R, Webster JrFE. 1989. Organizational culture and marketing: defining the research agenda. The Journal of Marketing. 3-15.

Dillman DA, Smyth JD, Christian LM. 2009. Internet, mail, and mixed-mode surveys: The tailored design method. 3rd ed. Hoboken: John Wiley & Sons Inc.

Fabrigar LR, Porter RD, Norris ME. 2010. Some things you should know about structural equation modeling but never thought to ask. Journal of Consumer Psychology. 20(2): 221-5.

Ferrell O, Gonzalez-Padron T, Hult G, Maignan I. 2010. From market orientation to stakeholder orientation. Journal of Public Policy & Marketing, 29(1): 93-6.

Freeman R. 1984. Strategic management: A stakeholder approach: Pitman Boston, MA.

Fornell C, Larcker DF. 1981. Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research. 18(1): 39–50.

Gallagher D, Ting L, Palmer A. 2008. A journey into the unknown; taking the fear out of structural equation modeling with AMOS for the first-time user. The Marketing Review. 8(3): 255-75.

Grinstein A. 2008. The effect of market orientation and its components on innovation consequences: A meta-analysis. Journal of the Academy of Marketing Science. 36(2): 166-73.

CERTIFICATE

OF PUBLICATION

For the article titled:

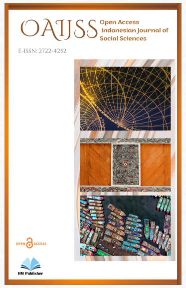
Data Analysis Procedures with Structural Equation Modelling (SEM): Narrative Literature Review

Authored by;

Rachmat Hidayat, Patricia Wulandari

Published in

Open Access Indonesia Journal of Social Sciences Volume 5 Issue 6 2022



Indexed in:

doi Crossref GARUDA GARBA RUJUKAN DIGITAL INDEX COPERNICUS INTERNATIONAL Dimensions