

## AMOS CB-SEM Multicollinearity

John R. Hamilton<sup>1,\*</sup>  
Singwhat Tee<sup>2</sup>

\*Corresponding author

<sup>1</sup> Adj. Chair Professor, James Cook University, Australia, John.Hamilton@jcu.edu.au

<sup>2</sup> Senior Lecturer, James Cook University, Australia, Singwhat.Tee@jcu.edu.au

### ABSTRACT

CB-SEM multicollinearity occurs when a statistical model's exogenous constructs are highly correlated. Multicollinearity typically alters regression path coefficients and potentially leads towards higher standardized errors, making standardized path coefficient estimates less precise, less reliable, and so more difficult to precisely interpret. Multicollinearity detrimentally affects overall fit of the CB-SEM model and typically delivers weaker model fit indices. Thus, particularly in complex CB-SEM models, hypothesized framework model relationships can become more difficult to interpret, as multicollinearity typically reduces the impact of each relevant measurement item(s) on its attached exogenous construct(s). Overall, recommendations to deal with multicollinearity are provided.

*Keywords:* framework modelling, data cleaning, exploratory factor analysis (EFA), confirmatory factor analysis (CFA), quantitative statistics, discriminant validity, collinearity.

### INTRODUCTION

The IBM statistical package AMOS (analysis of moment structures) is a graphical display approach to performing covariance based structural equation modelling (CB-SEM). Bacon and Bacon (2001), Cunningham (2008) and many others advise CB-SEM combines multiple regression and factor analysis, but the embedded statistics also adds further benefits – including effective ways for dealing with multicollinearity, and methods to account for unreliability that sometimes arises across survey respondent data. This powerful statistical approach enables specification and investigatory analysis of the measurement and/or structural relationships between each model's set of exogenous and endogenous latent variables.

### LITERATURE

AMOS models exogenous constructs and their covariance structures with endogenous intermediate and dependent constructs across datasets. It assesses literature and theoretical developed framework models by examining the relationships between multiple constructs (and their observed measurement items). CB-SEM typically defines the theoretical model and its measurement and structural components (refer Figure 1), and ensures the model is statistically identifiable.

### Measurement Versus Structural Model

Figure 1 presents a basic, causal, two construct, structural and measurement model. This measurement model captures both construct data sets of regression equations (McDonald, 2010). This CB-SEM model of two unobserved constructs and their measurement items is literature developed and theoretically framed into a complex probabilistic model (Grace & Bollen, 2008).

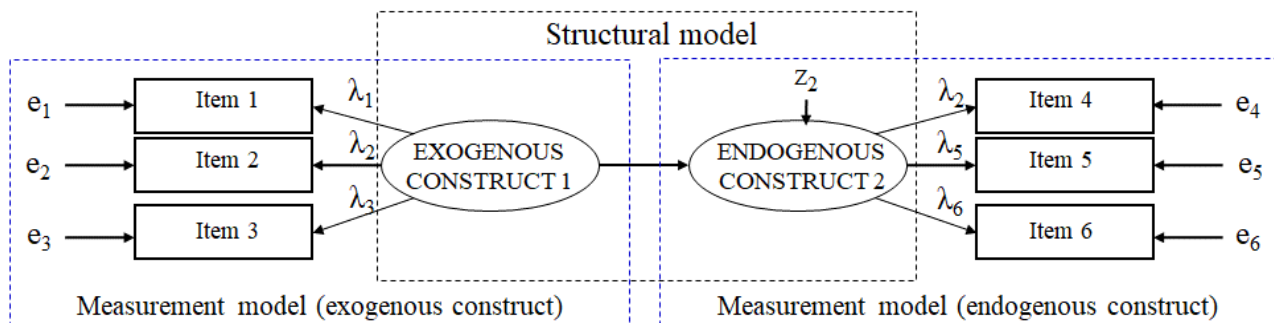


Figure 1. General Structural and Measurement Model (adapted Dandagi *et al.* 2016)

Figure 1's hypothetical two constructs typically represent individual aspects of the measurement model's underlying causes; help explain a measurement item's covariant behavior. In CB-SEM, the construct's measurement items typically encapsulate survey data about these items, and so each likely possess a degree of error ( $e$ ). All endogenous constructs also possess another error due to model misspecification arising from where a proposed model does not accurately represent underlying relationships between constructs ( $z$ ). Each measurement item load ( $\lambda$ ) indicates the strength of relationship between each item measure and its underlying construct. Each of these can be CB-SEM estimated (Joreskog & Sorbom, 1989), but there is a

detailed process to follow prior to instigating a full parameter CB-SEM modelling solution. The Figure 1 CB-SEM model represents two literature developed factors that likely exist within this model, but each factor (often termed a construct) is itself unmeasured (Bollen, 2002).

Figure 1 shows the AMOS CB-SEM factor analysis and regression combination. It specifies a priori, a theoretical model comprising hypothesized (or literature-proposed) causal relationships between literature supported constructs and their measurement items. The exogenous construct(s) provide the independent component, and the endogenous construct(s) provide the intermediate and/or dependent component. These model constructs, defined by their measurement items (loads and errors), plus endogenous regressed error terms (residuals) enable the hypothesized model to be data-tested against the hypothesized theoretical framework model. The AMOS stats allow estimation of parameters across the model, whilst assessing model fit from many perspectives, and allows model modification based against fit indices, estimates, literature and theoretical considerations, and a raft of embedded refinement considerations (Arbuckle, 2021). CB-SEM relates measurement models (relations between constructs and their measurement items) and path analysis (relations between constructs) and allows evaluation of complex interrelated multivariate hypotheses (Cunningham, 2008).

Initially, CB-SEM works with a theoretically developed structural model framework and the constructs involved are developed against literature-support. It draws on survey questionnaire scale (1 to 5 or 1 to 7) (or continuous) questionnaire response measures captured from the survey's quantitative dataset. This dataset enables measure developments necessary to make CB-SEM modelling parameter estimations. Conjointly adding structural model testing and model fit assessments allows model re-specifications and refinements and assists an iterative process culminating in a final acceptable structural model (Bollen & Long, 1993; Coenders, Batista-Foguet & Saris, 2008).

This study highlights multicollinearity considerations as its unique contribution to CB-SEM analysis. Compared to prior literature it highlights the potential of multicollinearity in highly correlated exogenous constructs to alter regression path coefficients and potentially lead towards higher standardized errors, making standardized path coefficient estimates less precise, less reliable, and so more difficult to precisely interpret, and so detrimentally affects overall fit of the CB-SEM model.

## METHODOLOGY

### Establish a Literature and Theoretical Research Framework

In higher level research studies such as in PhD dissertations it is first necessary to clearly define the research problem and its specific research question (or questions). The research question guides the broad, extensive, comprehensive, targeted literature research. It guides collation of detailed journal, conference, book, industry, media, and scholarly resources into an appropriate structured body of stating point knowledge.

Next, perceived change structures, models, and idea frameworks are literature explored, and developed as component change drivers relevant to the research question (or problem). In CB-SEM studies both literature-supported constructs and their likely measurement items are framed individually as either exogenous (independent/input/starting) constructs or as endogenous intermediate (actionable) constructs, or as endogenous dependent (outcome) constructs. Next the theoretically developed framework of interlinked literature-supported constructs creates the CB-SEM structural model component.

This scholarly framework is iteratively model-tested and critically evaluated for quality of fit, additional missing components, for alternatives, for weaknesses, for challenges, for limitations, for modelling outcome potential, and for possible mediators and/or moderators. Each evaluated consideration is included when necessary, or reason-specific eliminated. Thus, a solid research question (or problem) is CB-SEM model solved and likely guides future useful and interpretable understanding.

### Questionnaire Development

This research team prefers mixed methods, online, survey questionnaires as a robust research vehicle. External quantitative Likert scale data capture is best gauged as a five-point (1 = strongly disagree to 5 = strongly agree) scale. Likert scaling often applies to attitudinal, or opinion, or behavioral data capture. Typically, five to six items are literature defined per concept. All questions contain one only brief and simple point or perspective. Demographics is deployed to potentially group or sub-section the resultant data set of individual cases based on age, gender, job, work, qualifications. Nationality, and the like. Qualitative open-ended questions help gain understanding and immediate feedback around the research problem and its possible model linkages, along with specific understanding around the literature concepts within (1) the exogenous domain, (2) the intermediate endogenous actioning domain, and (3) the resultant outcome endogenous domain.

Online surveys like Qualtrics or Survey Monkey are best ordered with quantitative questions first, then, demographics questions and lastly qualitative questions. Research over two decades shows online survey response rates deploying 5 weekly reminders (each varied to further promote the research's importance) can reach 50% to 60% of well targeted respondents if the questions are short (1 to 7 words), simple, and very clear in their request for a decision.

## Steps in CB-SEM Modelling

### *Quantitative data capture and cleaning*

Once the questionnaire is developed from the literature and refined for both clarity and minimal verbiage it is normally pilot tested to assess question clarity and the potential to deliver a CB-SEM model. These preliminary findings are incorporated into the study's final questionnaire.

In CB-SEM studies the data set is the respondent's questionnaire quantitative Likert scale answers. Before CB-SEM analysis the data set is assessed for problems. Respondent missing item values affect the statistical analysis. Hence, a data cleaning process is required. Proper data cleaning can enhance the quality of analysis, ensure accuracy, build greater reliability and enable better validity.

Data entry inconsistencies such as poor item questions, skewing and/or kurtosis, typographical mistakes, incorrect entries, missing value pattern inconsistencies, consistently bias answers, and non-sensical outliers or bimodal distributions can deliver inflated errors and/or poor measurement information. By identifying and removing patterns across missing data, and by choosing an appropriate imputation missing value replacement solution – such as the data set item's mean or median missing value (saved as a whole numerical replacement) a final no missing values full data set for analysis can be delivered, provided the missing values are well below 20%, and are missing completely at random (MCAR) and unrelated to any other measurement item or variable (Hair et al, (2014; 2017b), and the final dataset delivering Little's MCAR p value < 0.05. This dataset solution offers a better future maximum likelihood CB-SEM accuracy of analysis, a better reliability, and a more robust set of findings.

### *Exploratory factor analysis (EFA)*

Cunningham, (2008); Watkins (2018), Hair et al (2014); 2017b, and Kline (2023) provided guidelines to SPSS the varimax approach to exploratory factor analysis (EFA). EFA simplifies the exploration and determination of unrelated factors extracted from the dataset.

Varimax EFA orthogonal rotation delivers an understandable pathway to maximize the squared loadings of a factor against its item variables. Where non-normal data only is available then PAF (not principal components) is used. Where data is normal, principal components under orthogonal rotational extracts data in SPSS tables. The rotated component matrix table is assessed for a KMO > 0.6, then for cross-loading item removal, then for strongest correlations well above 0.3. An iterative process also removes any very weak factors or components. The strongest factors and their embedded items remain as an underlying factor/construct structure that strongly represents its underlying latent constructs. To interpret each factor/construct the remaining measurement items are checked for alignment with literature supported, and theoretically derived, factors/constructs. These remaining constructs and their final measurement items are next CFA investigated.

### *Confirmatory factor analysis (CFA)*

Cunningham, (2008); Watkins (2018), Hair et al (2014;2017b), and Kline (2023) provide guidelines to SPSS CFA. CFA tests and confirms the literature and theoretically proposed or hypothesized relationships between observed item measure variables and their constructs, whilst further to validating the reliability of embedded measurements. It typically reduces the number of observed item measures on a construct. It focuses these into a smaller number of strongest item measures per construct. CFA's maximum likelihood (ML) statistical approach is applied for near normal measurement items.

SPSS CFA ML factor reduction and oblim rotation delivers the best fit observed item measures associated with each proposed latent construct. CFA factor reduction delivers final construct items with high communalities > 0.7, high loadings > 0.7, and each with residuals < 0.05 – indicating model implied correlations remain reliably close to observed item measures, closely represent the data set, likely indicate high construct validity, and correctly specify the factor structure. These remaining observed item measures (residuals <0.05) per construct indicate the proposed or hypothesized relationships between observed item measure variables and their constructs fit against those provided via the literature and its theoretical expectations. Resultant CFA factor reduced constructs typically have construct means (with standard deviations) around normal distribution positioning and construct Cronbach alphas > 0.7. When developed against a literature and theoretical framework structure these constructs and their item measures are ready for further CB-SEM model analysis.

### *Quantitative CFA measurement model development*

The final data set is developed into a Table 1 format. Table 1 is a useful statistical summation of measures related reliability and CB-SEM modelling trials. Cunningham, (2008); Hair et al (2014; 2017b), Byrne, (2016a; 2016b), and Kline (2023) provide guidelines to each Table 1 column measure and its normally acceptable values.

Table 1 is a partial example of how constructs and measurement items can be efficiently presented as one visual array. Column 1 shows separate construct blocks (exogenous construct blocks, followed by intermediate endogenous construct blocks, and the then dependent exogenous construct block) – with each construct showing its own observed item measures (questionnaire items) (including those eliminated during factor reduction). Necessary measures for these constructs are listed as left to right columns. Column 2 lists each measurement item's load. Column 3 calculates the average variance extracted (AVE) per

construct (necessarily > 0.5). Column 4 is the construct mean, and the Column 5 standard deviation (SD) hopefully shows a near normal bell curve. Column 6 provides the construct Cronbach alpha (> 0.7 for suitable internal construct fit consistency of observed construct item measures). Column 7 calculates the construct item measures collapsed into a composite single indicator load. Column 8 calculates composite single indicator degree of measurement error across the underlying construct. Columns 9 to 13 calculate the composite critical reliability (CR) which should best reside between 0.7 and 0.9. Table 1 shows two high composite error terms (0.16 and 0.15) and suggests the possibility of measurement item problems on these two constructs.

Table 1: CB-SEM Measurement Data

MEASUREMENT ITEM (114 respondent cases in data set)	ITEM LOAD (λ)	AVE $(\lambda_{1.1}^2 + \lambda_{1.2}^2 + \lambda_{1.3}^2)/3$	MEAN (σ)	STD DEV (SD)	CROBACH ALPHA (α)	LOAD SD(σ $\alpha$ )	ERROR SD $^2(1-\alpha)$					
								$\lambda^2$	$1 - \lambda^2$	$(\Sigma\lambda)^2$	$\Sigma(1 - \lambda^2)$	$CR = (\Sigma\lambda)^2 / [(\Sigma\lambda)^2 + \Sigma(1 - \lambda^2)]$
<b>EXOGENOUS CONSTRUCT 1:</b>		0.59	2.43	0.91	0.86	0.84	0.12			5.32	1.22	0.81
Observed item measure 1.1	0.70							0.49	0.51			
Observed item measure 1.2	0.79							0.62	0.38			
Observed item measure 1.3	0.82							0.68	0.32			
<b>EXOGENOUS CONSTRUCT 2</b>		0.60	2.89	0.95	0.82	0.86	0.16			5.41	1.19	0.82
Observed item measure 2.1	0.77							0.59	0.41			
Observed item measure 2.2	0.81							0.65	0.35			
Observed item measure 2.3	0.75							0.56	0.44			
<b>EXOGENOUS CONSTRUCT 3:</b>		0.72	2.90	0.92	0.89	0.86	0.10			6.43	0.85	0.88
Observed item measure 3.1	0.84							0.71	0.29			
Observed item measure 3.2	0.87							0.76	0.24			
Observed item measure 3.3	0.82							0.67	0.33			
<b>INTERMEDIATE EDOGENOUS CONSTRUCT 4</b>		0.62	2.89	0.94	0.83	0.86	0.15			5.58	1.14	0.83
Observed item measure 4.1	0.77							0.59	0.41			
Observed item measure 4.2	0.81							0.65	0.35			
Observed item measure 4.3	0.78							0.62	0.38			
....		...	...	...	...	...	...			...	...	...

The above CB-SEM methodology summarizes into the following stepwise Figure 2 approach. This process includes (1) defining research questions, (2) reviewing theory and literature, (3) developing a research framework, (4) designing questionnaire, (5) cleaning data, (6) identifying constructs (EFA), (7) confirming final construct validity (CFA), (8) evaluating constructs, (9) computing AVE, Cronbach alpha, composite reliability, (10) running CB-SEM measurement model and (11) running CB-SEM the structural model.

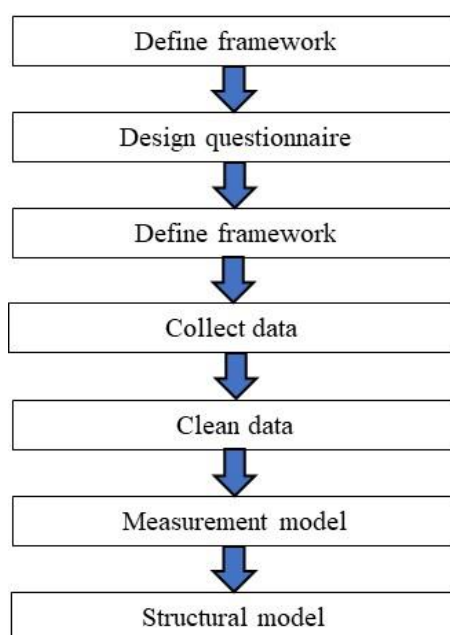


Figure 2 General CB-SEM Modelling Process

## ANALYSIS

Cunningham, (2008); Hair et al (2014; 2017b), Byrne, (2016a; 2016b), and Kline (2023) provide excellent guidelines for CB-SEM analysis. Modelling includes three main approaches – composite models, hybrid models, or full parameter models, along with occasional construct-specific model variations like item parcelling.

### Composite CB-SEM Path Modelling

A simple composite (single indicator) for each construct represents its averaged, or weight summed item measures. This approach attempts to mimic the proposed literature and theoretical framework model pathways. Munck's (1979) single indicator composite load and composite error terms calculate each composite construct's measures. This simple model approach minimizes interaction effects between observed item measures across the model. It helps expose possible significant model pathways. Assuming a suitable array of near excellent to excellent fit measures are composite model established, then a likely reflective relationship between constructs is established, and suggests a hybrid CB-SEM path model with full exogenous indicators and endogenous composite constructs may be achievable. Where no composite model is achieved, then a simple regression model may be attempted.

### Hybrid or Mixed CB-SEM Measurement Model

The hybrid CB-SEM model's exogenous full parameter constructs approach enlists item measures as reflective of their construct. It specifies relationships between constructs and their highly correlated item measures. It offers flexibility and more complexity in modelling relationships. It enables more comprehensive analysis between different types of constructs and can expose improvements to the overall fit across the structural equation model. It is, in effect, a CFA exogenous CB-SEM approach, and compared to the single indicator composites approach, it introduces more model interaction effects and provides the greatest exposure to the extent of possible multicollinearity between the exogenous constructs. Where this modelling is successful then a full structural CB-SEM model is investigated.

### Comprehensive CB-SEM Modelling - With Full Parameter Exogenous and Endogenous Constructs

This comprehensive CB-SEM modelling approach delivers detailed evaluation across complex CB-SEM models. It facilitates detailed analysis of and between model constructs and their embedded relationships. It captures and integrates both measurement model relationships between endogenous constructs and their item measures, along with the structural model causal relationships between the exogenous and endogenous constructs.

The comprehensive CB-SEM modelling offers holistic, simultaneous estimation of multiple relationships between model constructs. Modelling tests, and hopefully validates, the literature and theoretical framework model, and its propositions or hypotheses. It can estimate the strength and direction of causal relationships between constructs, validate theory and capture measurement and other errors.

Thus, CB-SEM modelling assesses model fit across a wide range of approaches and further validates model reliability. Modelling can, where necessary, introduce mediator or moderator effects, whilst capturing direct and indirect effects of exogenous and endogenous constructs

### Implementing CB-SEM Under IBM's AMOS Approach

To implement CB-SEM as an AMOS graphical approach the measurement model provides the relationships between latent variables and their indicators whilst the structural model provides the relationships between the latent variables (Byrne, 2010; Hair *et al.* 2014, 2017b). Data must be clean, with no missing values, and data must be appropriately formatted and be latent variable specific. Where data provides near normal items then CB-SEM analysis should follow maximum likelihood and rotational analysis (Byrne, 2010; Hair et al, 2014; 2017a; 2017b). When data is skewed to considerable degree, a principal axis factoring approach may provide a better analysis approach.

Model fit should deliver a chi-square/degrees of freedom ratio between 1 and 3 (with p value > 0.05). When not achievable a Bollen-Stine bootstrap (200 times) approach may show a  $p(\text{Bollen-Stine}) > 0.05$  (Cunningham, 2008, Hair et al, 2014; 2017b). Many modelling manipulations can help achieve a suitable fit. First to improve fit, assuming the constructs, their relational pathways, along with any mediation or moderation issue are each well understood, and correctly placed, then model paths can be added or removed. Second, latent variable measurement items can be reduced in number to minimize any interaction effects and improve fit. Third, the fixed item value of '1' on one latent variable's specific measurement item per construct can be moved. Fifth, outliers if present, can be removed. Sixth, composite path models can be attempted, Seventh, item parcelling can sometimes add another construct factor framed from dichotomous data. These and other model specific adjustments can improve both a chi-square/degrees of freedom ratio its p value (Cunningham, 2008, Byrne, 2010, Hair et al, 2014; 2017b). Byrne (2010; 2016a; 2016b), Hair et al, (2014; 2017a; 2017b; 2021 and Hamilton and Tee (2015) offer still more model adjustment options.

Next key model fit indices are assessed for goodness-of-fit. These excellent fit measures normally include CFI (> 0.95), TLI (> 0.95), GFI (> 0.90), and AGFI (> 0.90), along with the RMR (< 0.05) and RMSEA (< 0.05). Where these achieve the excellent fit the model presented adequately fits, and represents, the dataset employed (Cunningham, 2008; Byrne, 2010, Hair *et al.* 2014;

2017). When complex models (such as considered below) are under consideration, fit indices excellence measures are somewhat relaxed (Dash & Paul, 2021; Latan, 2023).

**DISCUSSION**

This Table 1 and Figure 3 example captures a 112 data set of valid and near normal distribution complete cases and measurement items, Figure 1 is AMOS 28.0 developed and follows the above processes and checks (including a varimax factor checking process).

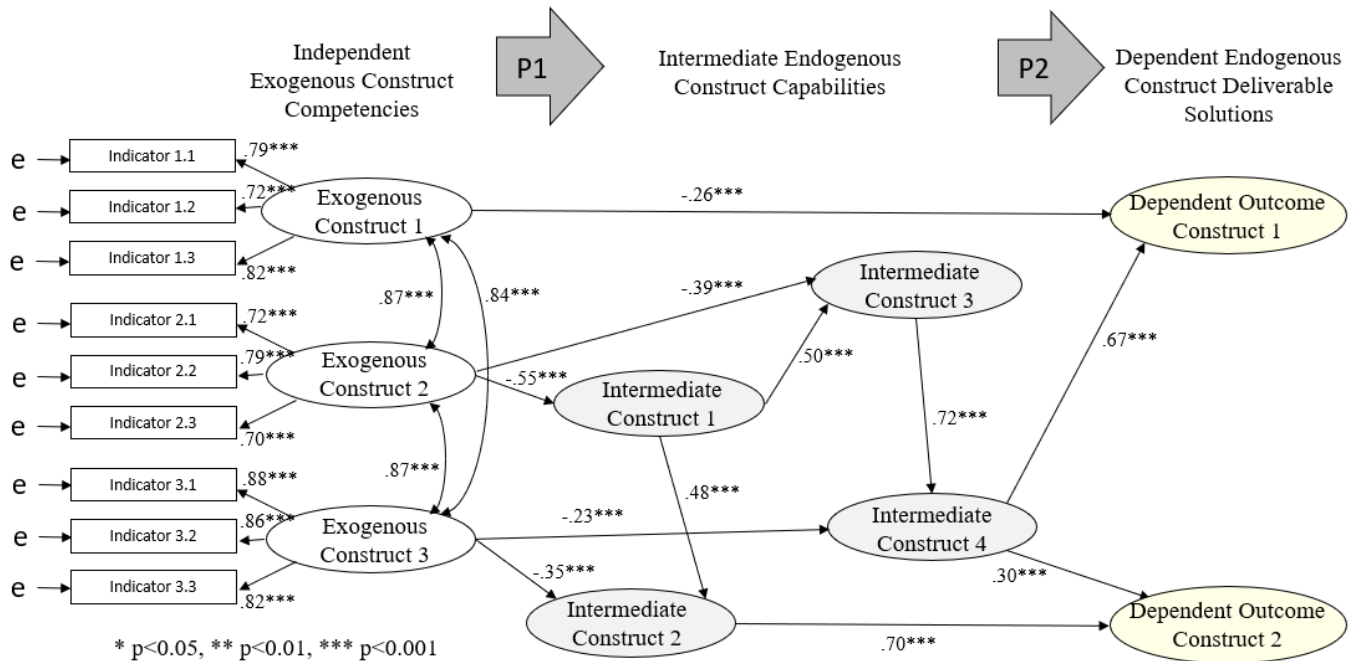


Figure 3: Example of CB-SEM Hybrid Model

The resultant Figure 3 complex CB-SEM hybrid model has three significant standardized high covarying exogenous constructs (Figure 1) and three high correlations (Table 3)– ranging between 0.84 and 0.87. This complex nine construct model (3 exogenous constructs, and six endogenous downstream constructs (four intermediate and two dependent constructs), has a four step longest path, and delivers ten near excellent fit measures (below Figure 1) against all criteria refer Table 1.

Figure 3 presents an example of a complex CB-SEM reflective measurement model analysis with some model problems. This hybrid model was literature and theoretically developed and well supported in the literature as possessing partial overlapping features. Each measurement item is related to one-only, key underlying construct. The exogenous construct items are expected to show high correlations and high covariances (0.84 to 0.87).

Figure 3’s model fit measures (Table 2) show: a p value is below 0.05, but an acceptable p(Bollen-Stine) (> 0.05), moderately fitting RMSEA and RMR values, borderline excellent fit TLI (~ 0.95), excellent CFI fit, weak GFI and AGFI values below 0.090 (Joreskog & Sorbom, 1993). This set of below excellent goodness-of-fit measures is further exacerbated as the data set is small (113 respondent cases).

Table 2: Hybrid CB-SEM model fit measures

Model Approach	Sample Moments	Parameters	Degrees of Freedom (DF)	Chi Square (χ)	χ/DF (> 1 to < 3)	p value (> 0.05)	p(Bollen-Stine) > 0.05	RMSEA (< 0.08)	RMR (< 0.05)	TLI (> 0.95)	CFI (> 0.95)	GFI (> 0.90)	AGFI (> 0.90)	Condition Number	Multi-collinearity
Hybrid	120	38	215	129.912	0.604	0.00	0.194	0.073	0.083	0.945	0.956	0.869	0.794	52.32	x

The high covariances and below excellence fit measures suggests Figure 2’s complex CB-SEM model likely has some degree of multicollinearity between at least one exogenous construct, and the other exogenous constructs, and their item measures. Hence a reinvestigation of each measurement item as being appropriate to the construct is warranted. Depending on circumstances, these exogenous measurement items - as multicollinearity enablers, may be deemed acceptable or unacceptable.

Considering Figure 1’s structural part of the Figure 3 complex CB-SEM model, multicollinearity between exogenous constructs can sometimes result in unstable estimates and error inflation. However, multicollinearity is generally due to literature or theoretical reasons in establishing the exogenous constructs. This material is varimax EFA defined into factor constructs, and CFA factor reduced to the final constructs and their specific item measures. When CB-SEM modeled with

bootstrapping the literature to EFA to CFA to Model process adds further validation of construct measurements and the CB-SEM model structure (Bacon & Bacon, 2001; Hair et al, 2014; 2017b; Mia *et al.* 2019). Hence, consideration of tolerating/accepting or reduction/removing multicollinearity is a research specific decision.

Where tolerance/acceptance is chosen, multicollinearity should be justified and noted as creating possible model instability, plus possible some error inflation, and significant causal and reflective model is accepted as existing. Where multicollinearity reduction/removal is chosen, a loss of model specificity (or a loss of model framework measurement and/or structure) can occur, but this does not necessarily eliminate it or alter the CB-SEM model solution (Bacon & Bacon, 2001; Mia *et al.* 2019).

### Investigating Multicollinearity

To investigate multicollinearity under SPSS (regression>linear>stats>collinearity diagnostics) and for the Figure 3 model the most severe exogenous construct covariance has a correlation with a variance inflation factor [(VIF) =  $1/(1-r^2)$ ] of 3.10 (Table 3) and this occurs between construct 1 and construct 2. This Figure 3 complex hybrid CB-SEM model multicollinearity is considered in at the top of low value range, and likely of little significance. A VIF < 5 or > 3 indicates more moderate correlation with some multicollinearity, and to some degree may affect the AMOS statistics.

Table 3: CB-SEM exogenous correlations and VIF determination

Correlations		Estimate	VIF
Exogenous Construct 1	<--> Exogenous Construct 3	0.843	3.46
Exogenous Construct 2	<--> Exogenous Construct 1	0.868	4.06
Exogenous Construct 3	<--> Exogenous Construct 2	0.870	4.11

Another multicollinearity assessment is via the AMOS CB-SEM sample moments 'condition index.' The condition index indicates small input errors likely cause large output errors and so is targeted across the model to be well below 30 (small models), or well below 100 (complex hybrid models), or well below 200 (complex comprehensive models). This also relies on the processes undertaken in building the model being correct and delivering a near excellent model fit. In this Figure 3 case the condition index is 52 - suggesting multicollinearity exists but is likely acceptable (Hair *et al.* 2014; 2017b). Further, in complex CB-SEM models, the sample moments covariance matrix determinant generally is above zero, and so two (or more) exogenous constructs generally do experience some multicollinearity.

### Multicollinearity Considerations

Grewal *et al.* (2004), Byrne, (2010, 2016a) and Hair et al, (2014) each advise multicollinearity likely lessens the precise estimation of these exogenous latent variables and their measurement items and so affects the downstream endogenous latent variables and the overall model structure by creating larger standard errors, less reliable parameter estimates, and weaker statistical and data interpretation. Such poorer measures likely reduce model outcome specificity and interpretations. Hence, it is best minimized by rigorous pre-model and model development processes.

Successful complex CB-SEM modelling generally necessitates assessment of multicollinearity. Multicollinearity is expected where literature supported exogenous constructs and the theoretically developed framework model place exogenous construct measurement items within similar and related domains, or within different levels of the same domain. Multicollinearity arises when individual effects of an exogenous construct and its measurement items are difficult to isolate from individual effects of other associated exogenous constructs and their measurement items (Hair et al, 2017).

The literature indicates astute preliminary and theoretically supported framework model developments, plus EFA and CFA solutions can partly compensate for multicollinearity and so reduce its effects in complex CB-SEM models. This is achieved by recognizing:

1. CB-SEM is a regression and CFA combined approach and should be developed sequentially (Cunningham, 2008).
2. potential multicollinearity anomalies are compensated for by carefully developing literature constructs and their item measures and theoretically building these exogenous constructs into a rigorous measurement model (Cunningham, 2008).
3. CB-SEM model refinement trials modify the path model towards delivering a well-specified, near excellent fit structural model (Steenkamp & Baumgartner, 2000; Grewal *et al.* 2004; Mia *et al.* 2021).
4. relationships between the latent constructs focus towards observed item measures and their modification directed towards effectively reducing CB-SEM multicollinearity (Grewal *et al.* 2004; Steenkamp & Baumgartner, 2000).
5. reassessment of the data set towards delivering an optimal model involves assessing the CB-SEM outputs via a wide variation of multiple perspective goodness-of-fit indices. These include absolute fit indices (chi square test, root mean square error of approximation, standardized 'root mean square residual'); incremental fit indices (comparative fit index, Tucker-Lewis index, normed fit index), several parsimony fit indices, and overall information criteria (Akaike information criterion).

All five compensations for multicollinearity, collectively (1) lead towards delivering a more reliable and valid modelling solution, and (2) lead towards providing an improved CB-SEM model outcome. These five compensation process approaches help recognize, understand, and partly solve degrees of multicollinearity between exogenous constructs – provided such

standardized exogenous construct correlations are below 0.9 (Hair *et al.* 2014; 2017b; 2021). Where exogenous construct correlations exceed 0.9 then the CB-SEM model exogenous item measure(s) on one or more specific covarying constructs should undergo revision. However, any revision should in general, not contradict the literature measurement item's foundations, or its structural model theoretical pathways.

Thus, (1) where high correlations among complex CB-SEM model exogenous constructs or their item measures are theoretically justified, and measurement items also suitably measurement-related but also map different aspects of one only construct within the literature, and (2) where constructs fit the theoretically developed pathways of the framework model, and (3) where the CB-SEM model overall goodness-of-fit borders on excellent against all measures, then the remaining level of multicollinearity, should be noted as one possibly inflating errors, and as one possibly requiring 'future' attention - such as altering different construct item measures or removal of an apparent overlapping construct, and checking if such steps may reduce multicollinearity. However, developed complex CB-SEM models (such as in Figure 3) generally should be retained as an initial reliable and validated empirical pathway towards expanding, and better understanding the CB-SEM modelling approach across each new research study's literature and the new theoretically developed framework and its modelled complex CB-SEM solution.

### Gauging Multicollinearity

Grewal *et al.* (2004) suggest several methods can combine to indicate if multicollinearity is an issue in AMOS CB-SEM modelling. Cunningham (2008), Byrne (2010; 2016a) and Hair *et al.* (2014; 2017a; 2017b) also provide a combination of several methods that detect multicollinearity. These include construct or factor improvements via:

- (a) literature and theoretically supported framework modelling
- (b) data cleaning
- (c) varimax EFA factor isolation to check literature proposition factors, along with literature factor naming and item determination
- (d) CFA, maximum likelihood factor reduction processes leaving all remaining construct item measure residuals below 0.05
- (e) theoretically developed framework model moved successfully into AMOS CB-SEM modelling

In AMOS CB-SEM multicollinearity applies to assessment of model exogenous, or predictor, or independent latent variables and their directional information flows. Grewal *et al.* (2004); Byrne (2010; 2016a) and Hair *et al.* (2014, 2017a) indicate in complex CB-SEM models, multicollinearity may be significant, and require consideration for adjustment – especially when the model's exogenous constructs and measurement items as individual constructs or factors show:

- (a) model correlation matrix components (exogenous constructs and their measurement items) above 0.8.
- (b) model squared multiple ( $r^2$ ) correlations above 0.9.
- (c) model squared multiple ( $r^2$ ) correlations [expressed as a 'variance inflation factor' ( $1/(1-r^2)$ )] above 10.
- (d) the covariance matrix 'AMOS condition number' (largest correlation (or covariance) eigenvalue divided by smallest correlation (or covariance) eigenvalue) well below 100 for complex CB-SEM models or well below 30 for simple CB-SEM models,

### Fixing Multicollinearity

Bacon and Bacon's (2001) IBM SPSS white paper suggests to deal with multicollinearity (1) ignore it (generally unwise), (2) remove it using the cleaned survey data and varimax exploratory principal components factor analysis, or (3) model it by taking the cleaned survey data and varimax supported literature constructs and then applying CFA reduction (across multiple respondent answers to the same concept (or construct) to deliver construct final item measures with residuals < 0.05. The above process considers (1) and then delivers both (2) and (3) approaches prior to CB-SEM modelling and then uses the fit indices (Dash & Paul, 2021) to further support the validity of the resultant final complex CB-SEM model.

## CONCLUSION

Multicollinearity particularly applies to regression. High correlations between pairs of exogenous constructs infer the presence of high collinearity or multicollinearity, and this presence reduces a construct's predictive power by the extent it is associated with other constructs (Hair *et al.*, 2017b). Regression along with factor analysis contribute towards CB-SEM analysis. Hence, multicollinearity is likely present in CB-SEM analysis. However, in developing complex CB-SEM models there are many construct developments, and many modelling precursor steps, which if correctly followed, can somewhat account for, but generally not eliminate, multicollinearity.

Where a near excellent fit complex CB-SEM model is (1) correctly generated, (2) correctly built from literature-developed constructs and item measures, and (3) theoretically developed into a complex causal SEM framework model, then multicollinearity – provided it is not excessive and is explainable, should not automatically require the instigation of change across the resultant complex CB-SEM model, or its pathways between constructs. For example, Table 3's complex CB-SEM hybrid model exogenous correlations, and their VIF determinations of 3.1 indicate peak low to minimum moderate multicollinearity. This indicates possible low degree of impact on model specificity and significance, plus a likely low reduction in model precision and a likely slight increase in model error. However, it does not mean a near excellent complex

CB-SEM hybrid model is deemed invalid, but rather just considered for modification (refer recommendations below), or even considered via deployment of an alternate investigative approach (refer recommendations below).

In complex CB-SEM models and where very-high degrees of multicollinearity occur (exogenous construct correlations above 0.9, VIF above 10, and condition index above 100), model fixes are required. Hence, considering the Figure 3 example its ongoing empirical research-developed complex hybrid CB-SEM model becomes a framework available for future research study improvements, or for further construct and measurement item change enhancements. In such model fix studies, as exogenous constructs item measures alter, model changes may (or may not) further improve the resultant complex CB-SEM model – but with the proviso that the model’s intended research direction is not adversely altered. Such positive additions (or changes) may (or may not) also lead to reduced construct errors, improved model fit, and greater model specificity.

### RECOMMENDATIONS

Under CB-SEM first understand the data and the constructs. Then consider if multicollinearity is likely, yet probably acceptable for the study undertaken. If multicollinearity is likely present to a high degree, researchers can choose to: (1) ignore it, or (2) remove a relevant exogenous construct, or (3) combine the two or more relevant exogenous constructs, or (4) deploy a larger data set, or (5) remove it against cleaned survey data and varimax principal components exploratory factor analysis (EFA) to remove specific measurement items or their constructs, or (6) model it after EFA and CFA reduction assessments and refinements (residuals < 0.05) down to a few key perspectives of each construct and assess the modelling outcomes.

Alternatively, researchers can use a different statistical technique to CB-SEM, such as the somewhat more flexible PLS-SEM (Latan *et al.* 2023) approach where variance-based assessment helps explain endogenous constructs and works better for small and even non-normal data sets. However, PLS-SEM trades-off these advantages for less sensitive exogenous and endogenous relationship interpretations. A third approach is to engage a Ridge Regression PLS-SEM (Streukens & Leroi-Werelds, 2023) approach. However, Ridge Regression PLS-SEM also introduces model weaknesses by adds estimates bias into the regression. So, no complex modelling approach is perfect, but all three recommendation approaches have their place, and should be considered, and then used appropriately.

Table 4 summarizes the three common approaches to CB-SEM modelling (1) composite (single indicators) for complex path models where no other measurement modelling is achievable, (2) hybrid (mixed constructs) for more detailed independent variable and measurement model assessment, and increased relevance of resultant path model, and (3) comprehensive (full parameter) structural models for full structural and measurement model assessment. But, to achieve any of these CB-SEM models AMOS implementation requires clean data, and modelling with significant  $\chi^2/DF$  ( $p > 0.05$ ), CFI, TLI, GFI, AGFI, RMSEA, and RMR measures. Further, to solve multicollinearity (1) ignore it (not advised), (2) remove/combine constructs, (3) use larger datasets, (4) apply EFA/CFA, then CB-SEM (5) consider PLS-SEM, (6) choose an approach based on the research context and data characteristics.

Table 4: CB-SEM model comparison options

CB-SEM Aspect	Composite Model	Hybrid Model	Comprehensive Structural Model
Exogenous Constructs	Simplified model using single indicators	Combination of full composite constructs	Full parameter modelling
Exogenous Constructs	Composite (single indicator)	Full parameter (multiple indicator)	Full parameter (multiple indicators)
Complexity	Low	Moderate	High
Data Requirements	Moderate	Moderate to high	High
Use Situation	Quick analysis with simplified path structure	Balance between detail and simplicity	Detailed analysis of all constructs
Advantages	Easier to implement	Captures key relationships whilst reducing colinearity	Most accurate representation of theoretical model
Limitations	May oversimplify relationships	Still subject to multicollinearity in exogenous constructs	Requires large sample size and complex analysis

## REFERENCES

- Arbuckle, J. L. (2021). *IBM® SPSS® Amos™ 21 User's Guide* (pp. 1–680). Armonk, NY: IBM Corp.
- Bacon, L. D., & Bacon, L. D. (2001). *Using Amos for Structural Equation Modeling in Market Research*. Chicago, IL: Lynd Bacon & Associates Limited and SPSS Incorporated.
- Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, 53(1), 605-634.
- Byrne, B. M. (2010). *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming* (2<sup>nd</sup> ed.). Routledge.
- Byrne, B. M. (2016). *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming* (3<sup>rd</sup> ed.). Routledge/Taylor & Francis Group.
- Byrne, B. M. (2016). *Structural equation modeling with Mplus: Basic concepts, applications, and Programming* (3<sup>rd</sup> ed.). Routledge.
- Cunningham, E. (2008). *A Practical Guide to Structure Equation Modeling Using AMOS*. Melbourne, Australia: Streams, Statsline.
- Dandagi, S., Bhushi, U., Bagodi, V., & Sinha, D. (2016). Strategic management of technical university: Structural equation modelling approach. *Journal of Modelling in Management*, 11(1), 75-90.
- Dash, G., & Paul, J. (2021). CB-SEM vs PLS-SEM methods for research in social sciences and technology forecasting. *Technological Forecasting and Social Change*, 173, 121092, 1-11.
- Grace, J. B., & Bollen, K. A. (2008). Representing general theoretical concepts in structural equation models: The role of composite variables. *Environmental and Ecological Statistics*, 15(2), 191-213.
- Grewal, R., Cote, J. A., & Baumgartner, H. (2004). Multicollinearity and measurement error in structural equation models: Implications for theory testing. *Marketing Science*, 23(4), 519-529.
- Hair Jr, F., Gabriel, M. L., & Patel, V. K. (2014). AMOS covariance-based structural equation modeling (CB-SEM): Guidelines on its application as a marketing research tool. *REMark: Revista Brasileira de Marketing*, 13(2), 44-55.
- Hair Jr, J. F., Babin, B. J., & Krey, N. (2017). Covariance-based structural equation modeling in the Journal of Advertising: Review and Recommendations. *Journal of Advertising*, 46(1), 163-177.
- Hair, Jr, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2017). *Multivariate Data Analysis* (7<sup>th</sup> Ed.). Harlow, Essex, England: Pearson Ed. Ltd.
- Hair Jr, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Danks, N. P., Ray, S., ... & Ray, S. (2021). An introduction to structural equation modeling. *Partial least squares structural equation modeling (PLS-SEM) using R: A workbook* (pp. 1–29).
- Hamilton, J. R., and Tee, S. (2015). Presenting SEM models: Journal and conference considerations. In *Proceedings of the 20th Asia Pacific Decision Sciences Institute Conference* (Vol. 20, No. 1, pp. 1–8).
- Joreskog, K., & Sorbom, D. (1989). *LISREL VII User's Guide*, Chicago, IL, USA: Scientific Software, Inc.
- Joreskog, K., & Sorbom, D. (1993). *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*. Chicago, IL, USA: Scientific Software, Inc.
- Kline, R. B. (2023). *Principles and Practice of Structural Equation Modeling* (5<sup>th</sup> ed.). New York, NY: Guilford Publications.
- Latan, H., Hair, Jr. J. F., Noonan, R. (2023). *Partial Least Squares Path Modeling: Basic Concepts, Methodological Issues and Applications* (2<sup>nd</sup> ed. pp. 183–207), Nature, Switzerland: Springer.
- McDonald, R. P. (2010). Structural Models and the Art of Approximation. *Perspectives on Psychological Science*, 5(6) 675-686.
- Mia, M. M., Majri, Y., & Rahman, I. K. A. (2019). Covariance based-structural equation modeling (CB-SEM) using AMOS in management research. *Journal of Business and Management*, 21(1), 56-61.
- Munck, I. (1979) *Model Building in Comparative Education: Applications of LISREL Method to Cross-National Survey*. (Doctoral thesis, monograph, pp. 1–199). Stockholm, Sweden: Almqvist & Wiksell International. Retrieved from <https://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-167396> (accessed 15 July 2025).
- Steenkamp, J. B. E., & Baumgartner, H. (2000). On the use of structural equation models for marketing modeling. *International Journal of Research in Marketing*, 17(2-3), 195-202.
- Streukens, S., & Leroi-Werelds, S. (2023). Multicollinearity: An overview and introduction of ridge PLS-SEM estimation. *Partial least squares path modeling: Basic concepts, methodological issues and applications*, 183-207.
- Watkins, M. W. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology*, 44(3), 219-246.