International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Structural Equation Modeling using Multivariate Analysis

[a] Manish K. Barve, [b] Amarsinh B. Landage

[a] Research Scholar, Government College of Engineering, Karad, M.S. India
[b] Assistant Professor, Government College of Engineering, Karad, M.S. India

A B S T R A C T

There are so many researches adapted on River water and its parameters for further development. As being of so many researches available, there was a need to have a compatible idea which can show that the values of your study is validating with the actual values. Structural Equation Modeling is a concept which performs factor analysis like Good Fit, IFI, RMSEA and other valuable factors which are proven to be helpful to the studies. In this research we have perform structural modelling on the river water parameters of Krishna River, India. After analyzing this method with the actual observed values, it was found that the 74 % of values are corelating with each other and as the model has alco given result which proves the factor analysis to be correct. So, we can go for Structural Equation Modeling for knowing that the observed value on-situ is perfectly fit for the model or not, so this study will be useful for analysis of water quality parameters.

Keywords: BOD, DO, Krishna River, Good Fir Model, RMSEA

1. Introduction

A group of statistical approaches known as structural equation modelling (SEM) enable the examination of a set of interactions between one or more independent variables (IVs), which may be continuous or discrete, and one or more dependent variables (DVs), which may also be continuous or discrete. Factors or measurable variables can both be IVs and DVs. Other names for structural equation modelling include route analysis, confirmatory factor analysis, causal analysis, simultaneous equation modelling, causal modelling, and analysis of covariance structures. Actually, the last two are specific varieties of SEM.

Multiple regression factor analyses can be used to answer queries using SEM. At its most basic, a study proposes a correlation between one measured variable (like acceptability of dangerous behaviour) and other measured factors (like gender, academic achievement, and institutional ties).

As an illustration, think about the connection between academic success and peer acceptance. The topic is utilised using a single data set to illustrate and compare several structural equation methodologies, and is used with multiple data sets in a number of illustrations, making it a key example for this study.

2. Methodology

For performing SEM, it was important to first create a statistical data in a manner that a software can read. The first analysis performed was Principal Component Analysis (PCA), which provides minimum, maximum and mean value. After obtaining the value we can move further to create a file in SPSS Software which will be useful for the procedure of SEM as shown in figure 1 and Table no. 1. Next to data view we need to develop a variable view with the code language of the software. After completing this procedure, we can import that file into IBM AMOS Graphics 26 which will help us to create a SEM. There also we need to make a regression path of multiple regression model. And this leads to Structural Equation Modeling shown in Figure 2.

Physico Chemical Properties.sav

	Name	Type	Width	Decimals	Label	Values	Missing
1	Month	Numeric	8	2	Month	{1.00, Dece...	None
2	Season	Numeric	8	2	Season	{1.00, Winte...	None
3	Station	Numeric	8	2	Station	{1.00, S1}...	None
4	pH	Numeric	8	2	pH	None	None
5	EC	Numeric	8	2	Electric Conduc...	None	None
6	Turbidity	Numeric	8	2	Turbidity	None	None
7	TDS	Numeric	8	2	Total Dissolve ...	None	None
8	BOD	Numeric	8	2	Biological Oxyg...	None	None
9	Nitrate	Numeric	8	2	Nitrate	None	None
10	DO	Numeric	8	2	Dissolved Oxyg...	None	None
11	Phosphate	Numeric	8	2	Phosphate	None	None

Fig. 1

	Name	**Type**	**Width**	**Decimals**	**Label**	**Values**	**Missing**
1	**Month**	**Numeric**	**8**	**2**	**Month**	**{1.00, December}**	**None**
2	**Season**	**Numeric**	**8**	**2**	**Season**	**{1.00, Winter}**	**None**
3	**Station**	**Numeric**	**8**	**2**	**Station**	**{1.00, S1} …**	**None**
4	**pH**	**Numeric**	**8**	**2**	**pH**	**None**	**None**
5	**EC**	**Numeric**	**8**	**2**	**EC**	**None**	**None**
6	**Turbidity**	**Numeric**	**8**	**2**	**Turbidity**	**None**	**None**
7	**TDS**	**Numeric**	**8**	**2**	**TDS**	**None**	**None**
8	**BOD**	**Numeric**	**8**	**2**	**BOD**	**None**	**None**
9	**Nitrate**	**Numeric**	**8**	**2**	**Nitrate**	**None**	**None**
10	**DO**	**Numeric**	**8**	**2**	**DO**	**None**	**None**
11	**Phosphate**	**Numeric**	**8**	**2**	**Phosphate**	**None**	**None**

Table No. 1

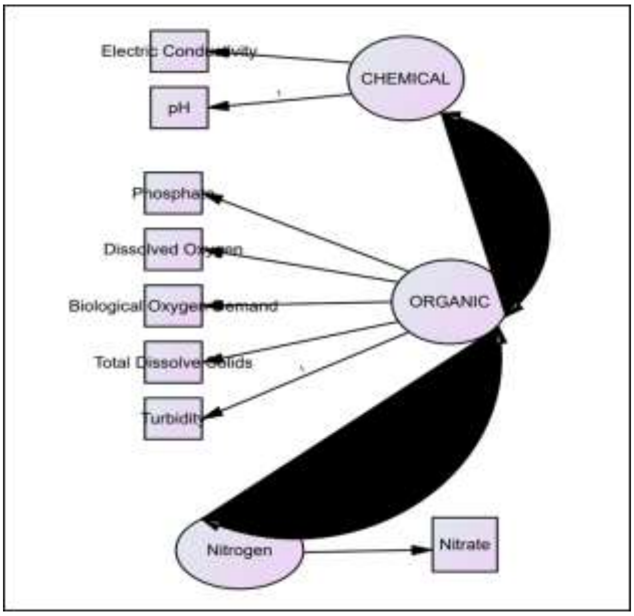

Fig. 2

3. Results and Discussion

As we can observe how the values are inserted and how the SEM model looks the results comes with the tool itself. There are two main methods we can adapt for SEM validation, one if Good Fit Model (GFI/RMR) and second one is Chi-square (CMIN). All these analyses are only possible after PCA of month December, January and February as shown in figure and table no. 3, 4 and 5 respectively. After that we need to insert the observed values in data view and save it as a .sav file which can easily be accessible by IBM Amos tool. The GFI , CMIN output should be below than or equal to 1, if it under the value then it is fitting with the model. The outputs of GFI, IFI , RMSEA and CMIN is shown in figure and table no 6,7,8 and 9 respectively.

	F1	F2	F3	F4
Eigenvalue	3.534	3.110	1.044	0.313
Variability (	44.171	38.875	13.045	3.908
Cumulative	44.171	83.046	96.092	100.000

Eigenvectors:

	F1	F2	F3	F4
pH	-0.087	-0.511	-0.026	0.715
Conductivit	0.381	-0.394	-0.063	-0.072
Turbidity	0.477	0.246	-0.001	-0.167
TDS	0.386	0.264	0.392	0.553
BOD	-0.265	0.151	0.808	-0.044
Nitrate	0.414	-0.293	0.321	-0.259
DO	0.459	0.261	-0.176	0.177
Phosphate	-0.139	0.526	-0.234	0.222

Fig. 3

	F1	**F2**	**F3**	**F4**
Eigenvalue	**3.534**	**3.110**	**1.044**	**0.313**
Variability	**44.171**	**38.875**	**13.045**	**3.908**
Cumulative	**44.171**	**83.046**	**96.092**	**100.000**
Eigenvectors				
	F1	**F2**	**F3**	**F4**
pH	**-0.087**	**-0.511**	**-0.026**	**0.715**
Conductivity	**0.381**	**-0.394**	**-0.063**	**-0.072**
Turbidity	**0.477**	**0.246**	**-0.001**	**-0.167**
TDS	**0.386**	**0.264**	**0.392**	**0.553**
BOD	**-0.265**	**0.151**	**0.808**	**-0.044**
Nitrate	**0.414**	**-0.293**	**0.321**	**-0.259**
DO	**0.459**	**0.261**	**-0.176**	**0.177**
Phosphate	**-0.139**	**0.526**	**-0.234**	**0.222**

Table no. 3

	F1	F2	F3	F4
Eigenvalue	4.317	1.921	1.097	0.665
Variability (%)	53.962	24.019	13.707	8.313
Cumulative %	53.962	77.981	91.687	100.000

Eigenvectors:

	F1	F2	F3	F4
pH	0.456	-0.129	-0.030	0.318
Conductivity	0.350	-0.442	0.274	-0.137
Turbidity	0.421	-0.151	-0.199	0.469
TDS	0.351	0.455	0.084	-0.304
BOD	0.285	0.520	-0.157	0.394
Nitrate	0.061	0.120	0.920	0.205
DO	0.372	-0.420	-0.072	-0.298
Phosphate	-0.381	-0.310	0.033	0.530

Fig. 4

	F1	**F2**	**F3**	**F4**
Eigenvalue	**4.317**	**1.921**	**1.097**	**0.665**
Variability (%)	**53.962**	**24.019**	**13.707**	**8.313**
Cumulative %	**53.962**	**77.981**	**91.687**	**100.000**
Eigenvectors				
	F1	**F2**	**F3**	**F4**
pH	**0.456**	**-0.129**	**-0.030**	**0.318**
Conductivity	**0.350**	**-0.442**	**0.274**	**-0.137**
Turbidity	**0.421**	**-0.151**	**-0.199**	**0.469**
TDS	**0.351**	**0.455**	**0.084**	**-0.304**
BOD	**0.285**	**0.520**	**-0.157**	**0.394**
Nitrate	**0.061**	**0.120**	**0.920**	**0.205**
DO	**0.372**	**-0.420**	**-0.072**	**-0.298**
Phosphate	**-0.381**	**-0.310**	**0.033**	**0.530**

Table no. 4

	F1	F2	F3	F4
Eigenvalue	2.949	2.664	2.280	0.107
Variability (	36.867	33.296	28.503	1.334
Cumulative	36.867	70.163	98.666	100.000

Eigenvectors:

	F1	F2	F3	F4
pH	0.303	0.114	-0.545	-0.407
Conductivit	0.181	-0.469	-0.353	0.552
Turbidity	0.228	0.124	0.594	-0.001
TDS	0.333	0.502	0.009	-0.169
BOD	0.386	0.403	-0.208	0.529
Nitrate	-0.371	0.453	0.103	0.461
DO	0.323	-0.363	0.387	0.084
Phosphate	0.567	-0.027	0.147	0.035

Fig. 5

	F1	**F2**	**F3**	**F4**
Eigenvalue	**2.949**	**2.664**	**2.280**	**0.107**
Variability (%)	**36.867**	**33.296**	**28.503**	**1.334**
Cumulative (%)	**36.867**	**70.163**	**98.666**	**100.000**
Eigenvectors				
	F1	**F2**	**F3**	**F4**
pH	**0.303**	**0.114**	**-0.545**	**-0.407**
Conductivity	**0.181**	**-0.469**	**-0.353**	**0.552**
Turbidity	**0.228**	**0.124**	**0.594**	**-0.001**
TDS	**0.333**	**0.502**	**0.009**	**-0.169**
BOD	**0.386**	**0.403**	**-0.208**	**0.529**
Nitrate	**-0.371**	**0.453**	**0.103**	**0.461**
DO	**0.323**	**-0.363**	**0.387**	**0.084**
Phosphate	**0.567**	**-0.027**	**0.147**	**0.035**

Table no. 5

Following are the results of SEM analysis with Maximum likelihood, Good fit and RMSEA.

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Saturated model	36	.000	0		
Independence model	8	24.475	28	.656	.874
Zero model	0	56.000	36	.018	1.556

Fig. 6

CMIN					
Mode	**NPAR**	**CMIN**	**DF**	**PCMIN/DF**	
Saturated Model	**36**	**0.000**	**0**		
Independence model	**8**	**24.475**	**28**	**.656**	**.874**
Zero model	**0**	**56.000**	**36**	**.018**	**1.556**

Table no. 6

RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Saturated model	.000	1.000		
Independence model	1102.357	.563	.438	.438
Zero model	1475.359	.000	.000	.000

Fig. 7

RMR, GFI				
Model	**RMR**	**GFI**	**AGFI**	**PGFI**
Saturated model	**.000**	**1.000**		
Independence model	**1102.357**	**.563**	**.438**	**.438**
Zero model	**1475.359**	**.000**	**.000**	**.000**

Table no. 7

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Independence model	.000	.000	.172	.700

Fig. 8

RMSEA				
Model	**RMSEA**	**LO 90**	**HI 90**	**PCLOSE**
Independence model	**.000**	**.000**	**.172**	**.700**

Table no. 8

1.1 Absolute fit indices

Absolute fit indices show which proposed model has the best fit to the sample data and determine how well an a priori model matches the data (McDonald and Ho, 2002). The most basic indicator of how well the suggested hypothesis fits the data is provided by these metrics. In contrast to incremental fit indices,

Calculation is based on how well the model fits in comparison to no model at all rather than comparison with a baseline model (Jöreskog and Sörbom, 1993). The Chi-Squared test, RMSEA, GFI, AGFI, the RMR, and the SRMR fall under this group.

1.2 Model Chi Square

According to Hu and Bentler (1999: 2), the Chi-Square value is the conventional metric for assessing overall model fit and "assesses the magnitude of discrepancy between the sample and fitted covariances matrices." The Chi-Square statistic is frequently referred to as a "badness of fit" (Kline, 2005) or "lack of fit" (Mulaik et al, 1989) measure because a decent model fit would produce an insignificant result at a 0.05 threshold (Barrett, 2007). Although the Chi-Squared test continues to be widely used as a fit statistic, there are a number of serious restrictions on its application. First of all, even while the model is adequately defined, large departures from multivariate normality may cause this test to reject a model (McIntosh, 2006). The Chi-Square statistic is sensitive to sample size because it is essentially a statistical significance test, which means that it almost always rejects the model when large samples are utilised (Bentler and Bonnet, 1980; Jöreskog and Sörbom, 1993). On the other hand, the Chi-Square statistic lacks power when using small samples, which may make it difficult to distinguish between models that fit the data well and those that don't (Kenny and McCoach, 2003). Due to the Model Chi-Square's limitations, researchers have looked for alternate metrics to judge model fit. The relative/normed chi-square (2/df) by Wheaton et al. (1977) is an illustration of a statistic that minimises the effect of sample size on the Model Chi-Square. Although there is no agreement on the appropriate ratio for this statistic, recommendations range from 5.0 (Wheaton et al., 1977) to 2.0 (Tabachnick and Fidell, 2007) notwithstanding the lack of a consensus.

1.3 Root mean square error of approximation (RMSEA)

The RMSEA was created by Steiger and Lind (1980, cited in Steiger, 1990) and is the second fit statistic presented in the LISREL programme. The RMSEA indicates how well the model would fit the population's covariance matrix if its parameter estimates were unknown but carefully selected (Byrne, 1998). Due to its sensitivity to the number of estimated parameters in the model, it has recently come to be known as "one of the most informative fit indices" (Diamantopoulos and Siguaw, 2000: 85). To put it another way, the RMSEA supports parsimony by picking the model with the fewest parameters. In the past fifteen years, recommendations for RMSEA cut-off points have been significantly lowered. Prior to the early 1990s, an RMSEA between 0.05 and 0.10 was seen as a sign of fair fit, whereas values more than 0.10 denoted poor fit (MacCallum et al., 1996). According to MacCallum et al. (1996), an RMSEA of 0.08 to 0.10 indicates a decent fit and less than 0.08 indicates a mediocre fit. However, more recently, the general agreement among experts in this field appears to be a cut-off value close to.06 (Hu and Bentler, 1999) or a strict upper limit of 0.07 (Steiger, 2007). The ability to construct a confidence interval around the RMSEA's value is one of its major benefits (MacCallum et al., 1996). This is made possible by the statistic's known distribution values, which enables a more accurate test of the null hypothesis (poor fit) (McQuitty, 2004). It is typically provided along with RMSEA, and in a well-fitting model, the top limit should be less than 0.08 and the lower limit should be near to 0.

1.4 Goodness-of-fit statistics (GFI)

Jöreskog and Sorbom developed the Goodness-of-Fit statistic (GFI) as an alternative to the Chi-Square test, which determines the percentage of variance that is explained by the estimated population covariance (Tabachnick and Fidell, 2007). The variances and covariances that the model accounts for demonstrate how well the observed covariance matrix is replicated by the model (Diamantopoulos and Siguaw, 2000). With larger samples, this statistic's value rises from 0 to 1. The GFI has a downward bias when there are many degrees of freedom relative to the sample size (Sharma et al., 2005). The GFI has also been discovered to rise as the number of parameters rises (MacCallum and Hong, 1997) and and also has an upward bias with large samples (Bollen, 1990; Miles and Shevlin, 1998). A higher cut-off of 0.95 is preferable when factor loadings and sample sizes are low, contrary to the conventional recommendation of an omnibus cut-off point of 0.90 for the GFI (Miles and Shevlin, 1998). This index has lost popularity in recent years due to its sensitivity, and it has even been suggested that it not be employed (Sharma et al., 2005).

The AGFI, which modifies the GFI depending on degrees of freedom, is related to the GFI and reduces fit as more saturated models are used (Tabachnick and Fidell, 2007). As a result, simpler models are preferred while complex ones are penalised. Furthermore, AGFI tends to rise with sample size. Similar to the GFI, the values for the AGFI similarly fall between 0 and 1, and values of 0.90 or higher are typically considered to be indicative of well-fitting models. These two fit indices are not used as a stand-alone index due to the frequently negative effects of sample size, but rather because of their historical significance, they are frequently reported in covariance structure analyses.

1.5 Incremental Fit Indices (IFI)

Incremental fit indices are a class of indices that do not use the chi-square in its raw form but instead compare the chi-square value to a baseline model (Miles and Shevlin, 2007; McDonald and Ho, 2002). According to McDonald and Ho (2002), the null hypothesis for these models is that all of the variables are uncorrelated.

1.6 Comparative Fit Index (CFI)

A modified version of the NFI that takes sample size into consideration (Byrne, 1998) and has good performance even with small sample sizes is the Comparative Fit Index (CFI: Bentler, 1990). Bentler (1990) originally presented this measure, which was later added as one of the fit indices in his EQS programme (Kline, 2005). This statistic, like the NFI, compares the sample covariance matrix with this null model by assuming that all latent variables are uncorrelated (null/independence model). This statistic's values, like those of the NFI, vary from 0.0 to 1.0, with values nearer 1.0 suggesting good fit. Initially, a cut-off criterion of CFI 0.90 was advanced. Recent research, however, indicates that in order to guarantee that incorrectly described models are rejected, a value greater than 0.90 is required (Hu and Bentler, 1999). This led to the current recognition of a value of CFI 0.95 as being indicative of good fit (Hu and Bentler, 1999). Due to being one of the measures least affected by sample size, this index is now a component of all SEM programmes and is one of the most often reported fit indices (Fan et al., 1999).

4. Conclusion

The aim of this study was to evaluate the SEM analysis to prove whether the observed on-situ parameters are fit or not for the model. Before analysing without the factor analysis it was very difficult to determine whether the values are good or not. But, after the operation with soft tools like SPSS Statistics and IBM Amos, we reach out to the conclusion that the GFI, IFI CMIN And RMSEA shows a truthful value on which we can depend for out further research. Hence use of Multivariate Analysis for Structural Equation Modeling is good. Another thing which is associated with it, is the iteration frequency, this one SEM can provide the data of numerous number of iteration required by the user which will save time and efficiency and provide the best results for it.

References

Akaike, H. (1974), "A New Look at the Statistical Model Identification," IEE Transactions on Automatic Control, 19 (6), 716-23.

Bagozzi, R.P., Yi, Y., and Phillips, L.W. (1991), "Assessing Construct Validity in Organizational Research," Adminstrative Science Quarterly, 36 (3), 421-58.

Barrett, P. (2007), "Structural Equation Modelling: Adjudging Model Fit," Personality and Individual Differences, 42 (5), 815-24.

Bentler, P.M. (1990), "Comparative Fit Indexes in Structural Models," Psychological Bulletin, 107 (2), 238-46

Bentler, P.M. and Bonnet, D.C. (1980), "Significance Tests and Goodness of Fit in the Analysis of Covariance Structures," Psychological Bulletin, 88 (3), 588-606.

Bollen, K.A. (1990), "Overall Fit in Covariance Structure Models: Two Types of Sample Size Effects," Psychological Bulletin, 107 (2), 256-59.

Gerbing, D.W. and Anderson, J.C. (1984), "On the Meaning of Within-Factor Correlated Measurement Errors," Journal of Consumer Research, 11 (June), 572-80.

Hu, L.T. and Bentler, P.M. (1999), "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives," Structural Equation Modeling, 6 (1), 1-55.

Kenny, D.A. and McCoach, D.B. (2003), "Effect of the Number of Variables on Measures of Fit in Structural Equation Modeling," Structural Equation Modeling, 10 (3), 333-51.

MacCallum, R.C., Browne, M.W., and Sugawara, H., M. (1996), "Power Analysis and Determination of Sample Size for Covariance Structure Modeling," Psychological Methods, 1 (2), 130-49.

Marsh, H.W., Hau, K.T., and Wen, Z. (2004), "In Search of Golden Rules: Comment on Hypothesis-Testing Approaches to Setting Cutoff Values for Fit Indexes and Dangers in Overgeneralizing Hu and Bentler's Findings " Structural Equation Modeling, 11 (3), 320-41.

McDonald, R.P. and Ho, M.-H.R. (2002), "Principles and Practice in Reporting Statistical Equation Analyses," Psychological Methods, 7 (1), 64-82.

Sharma, S., Mukherjee, S., Kumar, A., and Dillon, W.R. (2005), "A simulation study to investigate the use of cutoff values for assessing model fit in covariance structure models," Journal of Business Research, 58 (1), 935-43.

Steiger, J.H. (1990), "Structural model evaluation and modification," Multivariate Behavioral Research, 25, 214-12.

Steiger, J.H. (2007), "Understanding the limitations of global fit assessment in structural equation modeling," Personality and Individual Differences, 42 (5), 893-98.

Tomarken, A.J. and Waller, N.G. (2003), "Potential Problems With "Well Fitting" Models," Journal of Abnormal Psychology, 112 (4), 578-98.

Yuan, K.H. (2005), "Fit Indices Versus Test Statistics," Multivariate Behavioral Research, 40 (1), 115-48.