Journal of Engineering and Applied Sciences 14 (20): 7663-7668, 2019
ISSN: 1816-949X

Development of Energy Generation Model using Robust Partial Least Squares-Structural Equation Modeling (RPLS-SEM) Through Winsorization Method

Fadya Shakhim and Zulkifley Mohamed
Department of Statistics, Faculty of Science, Al-Zawya University, Al-Zawya, Libya
zulkifley@fsmt.upsi.edu.my

Abstract: Based on the PLS-SEM structure, validation is required for each part of the model for the structural model, measurement model and overall model. This study's objective is to outline the measurement and structural criteria required to develop the model. It also, presents a Robust Partial Least Squares-Path Modeling (RPLS-PM) via. Winsorization which includes the internal and external models. The relationship between an indicator variable and latent variables is depicted by the external model whilst the relationship between exogenous latent variables and endogenous latent variables is recounted by the internal model. The RPLS-PM's outputs and inputs were derived by considering electricity generation data relating to Al-Zawya Steam Power Plant, Libya.

Key words: Partial Least Square-Path Modeling (PLS-PM), Structural Equation Modeling (SEM), Ordinary Last Square (OLS), winsorization, steam power plant, exogenous

INTRODUCTION

Partial Least Square-Path Modeling (PLS-PM): Two approaches can be described when the structural equation model is considered for estimation of the relations: the Covariance-Based SEM approach (CB-SEM) which is widely employed is used chiefly to confirm (or refuse) a theory (Hair *et al*., 2016) and the variance-based SEM (PLS-SEM) approach is used to develop theories while conducting exploratory research. The path models include few elements: the structural model (denoted as inner model in PLS-SEM) which defines relations amongst the Latent Variables (LVs) and the measurement models (signified as outer model in PLS-SEM) which defines the relations amongst LVs along with their measures, i.e., their indicators (Garson, 2016). In this context, two different types of measurement models are considered: one for the exogenous LVs (i.e., the constructs defining model's other constructs) and one for the endogenous LVs (i.e., those constructs explained by the model) (Joseph *et al*., 2014).

In this study, a new model is developed based on the SEM and parameters are determined by employing the PLS-SEM that make use of real data concerning AL-Zawya Steam Power Plant, Libya, taken from the oil sector in Libya. This study emphasises on SmartPLS3 (Ringle *et al*., 2015), since, it includes the structural as well as measurement models for data analysis which are also, freely available to the research community worldwide. To mitigate or eliminate the impacts of outlying data points, robust methods are employed. Therefore, a new robust PLS-SEM Model is proposed in this research which builds on robustification of the covariance matrix employed in the classical PLS algorithm (Cassel *et al*., 1999). In this study, a robust estimator of covariance is selected in which Winsorization (Winsor, 1895) is employed as an estimator to determine covariance matrix in multivariate data set to decrease outlier's pernicious impact (Clark, 1995).

Data sources: Real data on power generation has been used in this study which were gathered and compiled by AL-Zawya Oil Refining Company's Technical Department in Libya (Gawedar and Ramakumar, 2016). These data extend until 2016. Libyan governmental sources (government reports like the Oil Survey of Libya) including General Directorate of Power Plant and the Energy Research Center are the key data collection sources that also, support national capacities for data collection (Giuma and Khiri, 2016).

MATERIALS AND METHODS

In this study, the variables used to develop and evaluate the model include:

- Output items: electricity (MW) and fresh water (m^3) generation

Corresponding Author: Zulkifley Mohamed, Department of Mathematics, Faculty of Science and Mathematics, Universiti Pendidikan Sultan Idris, 35900 Tanjong Malim, Perak, Malaysia, zulkifley@fsmt.upsi.edu.my

- Input data: the following key inputs are required in this analysis
- Desalination data: the amount of steam (tons per day) and seawater (m^3/day) needed to produce fresh water
- Steam power plant requirements, steam turbine (ton per day) and boiler (m^3/day of distilled water)
- Additive chemical: morphine (litre/day), phosphate (kg/day), anti-scale (L/day) and hydrazine (L/day)
- Maintenance and operation: average cost for chemical treatment (LYD/day) and day of fuel (LYD/day)
- Output items: fresh water (m^3) and electricity generation (MW)
- Input data: the important inputs needed for this analysis include the following
- Desalination data requirements which include: seawater (m^3/day) and the amount of steam (tons per day) to produce fresh water
- Steam power plant requirements: boiler (m^3\day of distilled water) and steam turbine (ton/day)
- Additive chemical: phosphate (kg/day), morphine (L/day), Hydrazine (L/day) and anti scale (L/day)
- Operation and maintenance: average cost for day of fuel (LYD/day) and chemical treatment (LYD/day)

A robust and efficient estimator of covariance: The Winsorized mean for jth variable and the estimated Winsorized covariance (Croux and Rousseeuw, 1992) matrix between W_i and W_j variables are defined as follows:

$$\overline{W}_j = \frac{1}{m_j}\sum_{j=1}^{m_j} W_{ij} \tag{1}$$

The vector of Winsorized estimator is given below:

$$\overline{W} = \begin{bmatrix} \overline{W}_1 \\ . \\ . \\ . \\ \overline{W}_p \end{bmatrix} \text{ and } S_{WQ_n}\left(W_i, W_j\right) = \frac{1}{(n\text{-}1)}\sum_{k=1}^{n} W_{ki}W_{kj}\text{-}n\overline{W}_i\overline{W}_j \tag{2}$$

For developing an alternative for the variance-covariance matrix, a robust estimator (the winsorized estimator) represented by W has been employed in this research in place of the usual mean vector. Also, the inverse of Winsorized covariance matrix.

$s^{-1}_{WQ_n}$ is used instead of the covariance matrix's inverse, since, the out liers can impact both the covariance and mean. Thus, the new robust variance-covariance matrix is employed for developing the Robust Partial Least Square (RPLS) and β_0 β_1 are the parameter estimates that are written as follows:

$$\sum\nolimits_{WQ_n}(X_i) = \left(X_i\text{-}\overline{W}\right)^T S^{-1}_{WQ_n}\left(X_i\text{-}\overline{W}\right) \tag{3}$$

Where, $s^{-1}_{WQ_n}$: inverse of the sample covariance matrix.

RESULTS AND DISCUSSION

Description of the model: A two-step model-building approach was put forward by Anderson and Gerbing (1988) which focuses on the analysis of two models that are conceptually distinct, i.e., a measurement model and then the structural model. The relationships amongst measured (observed) variables as a result of the latent variables are specified by the factor model or the measurement model. Similarly, relationships amongst the latent variables are specified by the structural model as suggested by theory. In the arrow diagram presented in Fig. 1, it is assumed by the researchers that a latent variable (unmeasured) summarizes each block of the manifest variables (measured) (Tenenhaus *et al*, 2005). The researchers also recommended the symbols for endogenous latent variables: steam power plant SPP(η_2) signifying steam turbine and boiler; DW(η_1) denoting the desalination units (steam and seawater), chemical additive CA (η_3) referring to morphine, sodium triphosphate, anti-scaleand hydrazine, the exogenous latent variables denoted by the symbol OP (ξ_1) which also, represents chemical treatment and fuel costs, output (η_4) is represented by electricity and fresh water. The two indicator variables y_1 and y_2 combine to form DW and y_3, y_4 for SPP. Four indicator variables y_5, y_6, y_7 and y_8 (sodium triphosphate, morphine, hydrazine and anti-scale) form CA. Two indicator variables x_1 and x_2 x are included for the exogenous latent variable OP Model. δ_1-δ_8 signify the measurement error for DW, SPP and CA, respectively. The ω_1 and ω_2 characterise the measurement error for OP. In the final set, output deals with the relationship between an endogenous latent variables (η_4) and the indicator variables (y_9 and y_{10}) δ_9 and δ_{10} signify the measurement errors for output. The measurement models DW, SPP, CA, OP and output latent variables (electricity and fresh water) are described by the sets including endogenous variable which can be represented as η_1-η_4, y_i (I = 1, 2, ..., 10). Seawater, boiler, steam, steam

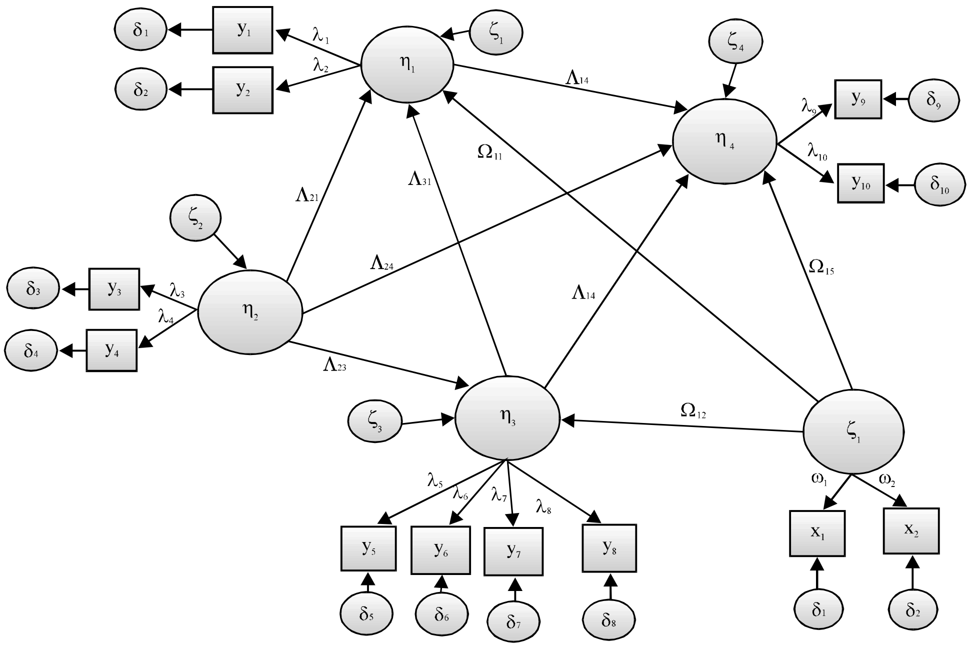

Fig. 1: Structural model and measurement model of DW, SPP, CA, OP and output

turbine, morphine, sodium triphosphate, anti-scale, hydrazine and the exogenous variables that are described as ξ_1, x_i ($i = 1, 2$) (chemical and fuel treatment) fall under measured variables, λ_i ($i = 1, \ldots, 10$) represent the correlat ion coefficients between endogenous latent variables, ω_i ($i = 1, 2$) and the indictor variables and the correlation coefficients between the exogenous latent variable and indictor variables. Also, according to researchers, ξ_i signifies the error of estimate for the endogenous latent variables, $i = 1, 2, 3, 4$ while $\boldsymbol{\delta}_1$-$\boldsymbol{\delta}_{10}$ and $\boldsymbol{\varepsilon}_1$, $\boldsymbol{\varepsilon}_2$ form part of the measurement errors and Λ_{21}, Λ_{31}, Ω_{11}, Ω_{12}, Λ_{23}, Λ_{14}, Ω_{14}, Λ_{24}, Λ_{34} and Ω_{14} for the regression coefficients signifying the endogenous latent and exogenous latent variables (Vinzi *et al*., 2009). The following presents the overall measurement and structural models of SPP, DW, CA, OP and output.

The concept of input and correlation between them can also be utilised for developing a model of electric power generation. The results of research studies conducted by Breeze (2008) and Cardona and Piacentino (2004) were employed to formulate this model which confirmed a significant correlation between DW SPP, CA, OP and the output. The suggested model methodology is described by using it with the real data pertaining to electricity generation in the Libyan oil sector. The following additive model is employed in the theoretical study:

$$\eta_4 = \Lambda_{14}\eta_1 + \Lambda_{24}\eta_2 + \Lambda_{34}\eta_3 + \Omega_{14}\zeta_1 + \zeta \quad (4)$$

where the coefficient Ω_0 denotes the total factor efficiency parameter referring to the composite primary factor inputs in this sector. Elasticity is generated with parameters. The following describes the structure matrix pertaining to the measurement model DW, SPP, CA, OP and output (Ramayah *et al*., 2016). The measurement model of endogenous latent variable (DW):

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} [\eta_1] + \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} \quad (5)$$

The measurement model of endogenous latent variable (SPP):

$$\begin{bmatrix} y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} \lambda_3 \\ \lambda_4 \end{bmatrix} [\eta_2] + \begin{bmatrix} \delta_3 \\ \delta_4 \end{bmatrix} \quad (6)$$

The measurement model of endogenous latent variable (CA):

$$\begin{bmatrix} y_5 \\ y_6 \\ y_7 \\ y_8 \end{bmatrix} = \begin{bmatrix} \lambda_5 \\ \lambda_6 \\ \lambda_7 \\ \lambda_8 \end{bmatrix} [\eta_2] + \begin{bmatrix} \delta_5 \\ \delta_6 \\ \delta_7 \\ \delta_8 \end{bmatrix} \quad (7)$$

The measurement model of exogenous latent variable (OP):

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} [\zeta_1] + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \quad (8)$$

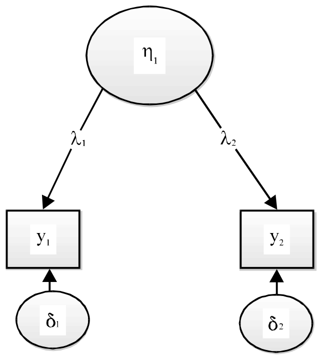

Fig. 2: The measurement model of DW

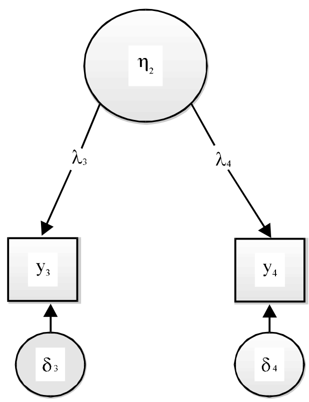

Fig. 3: The measurement model of SPP

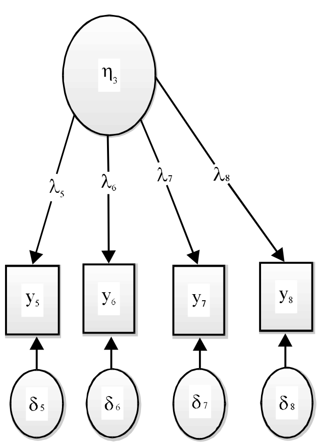

Fig. 4: The measurement model of CA

The measurement model of endogenous latent variable (output):

$$\begin{bmatrix} y_9 \\ y_{10} \end{bmatrix} = \begin{bmatrix} \lambda_9 \\ \lambda_{10} \end{bmatrix} [\eta_4] + \begin{bmatrix} \delta_9 \\ \delta_{10} \end{bmatrix} \quad (9)$$

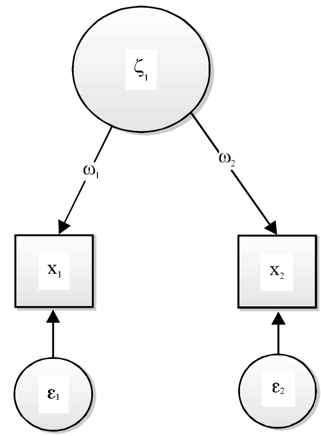

Fig. 5: The measurement model of OP

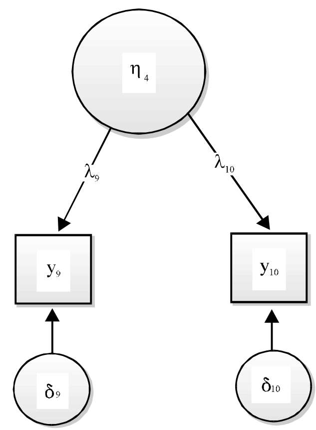

Fig. 6: The measurement model of output

The model includes five structural models, namely the SPP, DW and CA as well as output structural models and can be described as follows: the DW structural model is expressed as:

$$\eta_1 = \Lambda_{21}\eta_2 + \Lambda_{31}\eta_3 + \Omega_{11}\zeta_1 + \zeta_1 \quad (10)$$

The SPP structural model is expressed as:

$$\eta_2 = \Omega_{12}\zeta_1 + \zeta_2 \quad (11)$$

The CA structural model is expressed as:

$$\eta_3 = \Lambda_{23}\eta_2 + \Omega_{13}\zeta + \zeta_3 \quad (12)$$

The output structural model was stated as:

$$\eta_4 = \Lambda_{14}\eta_1 + \Lambda_{24}\eta_2 + \Lambda_{34}\eta_3 + \Omega_{14}\zeta_1 + \zeta_4 \quad (13)$$

Figure 1-9 presents the overall measurement and structural models for DW, SPP, CA, OP as well as the output.

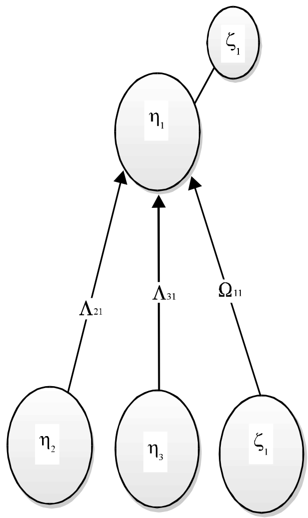

Fig. 7: The structural model of DW

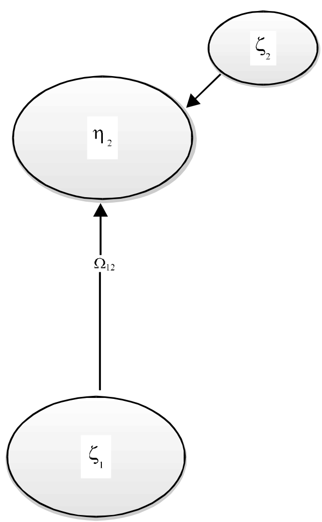

Fig. 8: The structural model of SPP

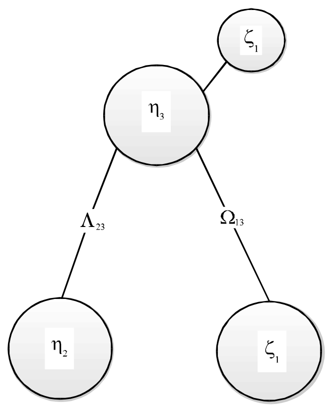

Fig. 9: The structural model of CA

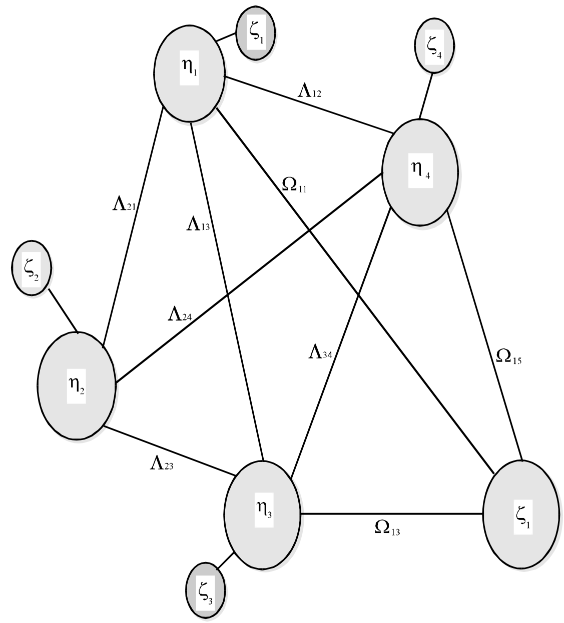

Fig. 10: The structural model of output

CONCLUSION

Figure shows the path diagram for an ideal RPLS-PM Model. In the general factor analysis framework, terms such as unidimensional measurement and simple structure are employed to represent the model meeting two conditions: a subset of indicator variables defines each latent variable which are considered as strong indicators for latent variable and there is a strong relation amongst each indictor variable and other latent variables. This studies prime focus is on a new model development based on the Winsorization augmented with RPLS-PM to determine parameters. The application of real data gathered from the oil sector in Libya is employed for illustrating this developed methodology. This model also, depicts the structural relationships existing with the outputs, based on generation of inputs via. generation of electricity in AL-Zawya Steam Power Plant, Libya.

REFERENCES

Anderson, J.C. and D.W. Gerbing, 1988. Structural equation modeling in practice: A review and recommended two-step approach. Psychol. Bull., 103: 411-423.

Breeze, P., 2008. The Cost of Power Generation: The Current and Future Competitiveness of Renewable and Traditional Technologies. Business Insight Inc., UK., Pages: 122.

Cardona, E. and A. Piacentino, 2004. Optimal design of cogeneration plants for seawater desalination. Desalin., 166: 411-426.

Cassel, C., P. Hackl and A.H. Westlund, 1999. Robustness of partial least-squares method for estimating latent variable quality structures. J. Appl. Stat., 26: 435-446.

Clark, R.G., 1995. Winsorization methods in sample surveys. Masters Thesis, Department of Statistics, The Australian National University, Canberra, Australia.

Garson, G.D., 2016. Partial Least Squares: Regression and Structural Equation Models. Statistical Associates Publishing, Asheboro, North Carolina,.

Gawedar, A. and R. Ramakumar, 2016. Impact of wind energy system integration on the Al-Zawiya refinery electric grid in Libya. J. Power Energy Eng., 4: 11-20.

Hair, J.F., G.T.M. Hult, C. Ringle and M. Sarstedt, 2016. A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). Sage Publications, Thousand Oaks, California, USA., ISBN:9781483377438, Pages: 384.

Joseph, C., C.Y. Tseng, G. Zocchi and T. Tlusty, 2014. Asymmetric effect of mechanical stress on the forward and reverse reaction catalyzed by an enzyme. PloS One, 9: 1-9.

Ramayah, T., J. Cheah, F. Chuah, F. Ting and M.A. Memon, 2016. Partial Least Squares Structural Equation Modeling (PLS-SEM) Using SmartPLS 3.0: An Updated and Practical Guide to Statistical Analysis. 1st Edn., Pearson Malaysia Sdn Bhd, Petaling Jaya, Malaysia, ISBN:978-967-349-739-3,.

Ringle, C.M., S. Wende and J.M. Becker, 2015. SmartPLS 3. SmartPLS GmbH Company, Boenningstedt, Germany.

Tenenhaus, M., V.E. Vinzi, Y.M. Chatelin and C. Lauro, 2005. PLS path modeling. Comput. Stat. Data Anal., 48: 159-205.

Vinzi, V.E., W.W. Chin, J. Henseler and H. Wang, 2009. Handbook of Partial Least Squares: Concepts, Methods and Applications. 1st Edn., Springer, Berlin, ISBN-13: 9783540328254.

Winsor, J., 1895. The New-England Indians: A Bibliographical Survey, 1630-1700. J Wilson & Sons, Boston Spa, England,.