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 refute) 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 emphasizes 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 (Gawed, 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 (Giuna and Khari, 2016).

MATERIALS AND METHODS

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

- Output items: electricity (MW) and fresh water (m³) generation

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

7663
Input data: the following key inputs are required in this analysis
- Desalination data: the amount of steam (tons per day) and seawater (m³/day) needed to produce fresh water
- Steam power plant requirements: steam turbine (ton per day) and boiler (m³/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 (m³) and electricity generation (MW)

Input data: the important inputs needed for this analysis include the following
- Desalination data requirements which include seawater (m³/day) and the amount of steam (tons per day) to produce fresh water
- Steam power plant requirements: boiler (m³/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:

\[
\overline{W}_j = \frac{1}{m_j} \sum_{i=1}^{m_j} W_{ij}
\]

The vector of Winsorized estimator is given below:

\[
\overline{W} = \begin{bmatrix}
\overline{W}_1 \\
\vdots \\
\overline{W}_p
\end{bmatrix}
\quad \text{and} \quad
\overline{S}_{\text{w},ij}(W_i, W_j) = \frac{1}{(n-1)} \sum_{k=1}^{n} W_{ik} W_{jk} - \overline{W}_i \overline{W}_j
\]

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.

\( s_{\text{w},ij} \) 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 \( \beta, \beta_i \) are the parameter estimates that are written as follows:

\[
\sum_{w_{ij}}(X_i) = (\bar{X}_i - \overline{W})^T \overline{S}_{\text{w},ij}(\bar{X}_i - \overline{W})
\]

Where, \( s_{\text{w},ij} : \) 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(\( \eta_p \)) signifying steam turbine and boiler; DW(\( \eta_d \)) denoting the desalination units (steam and seawater), chemical additive CA (\( \eta_c \)) referring to morphine, sodium triphosphate, anti-scale and hydrazine, the exogenous latent variables denoted by the symbol OP (\( \xi \)) which also, represents chemical treatment and fuel costs, output \( (\eta_o) \) is represented by electricity and fresh water. The two indicator variables \( y_1 \) and \( y_2 \) combine to form DW and \( y_w, y_s \) for SPP. Four indicator variables \( y_3, y_4, y_5 \) and \( y_6 \) (sodium triphosphate, morphine, hydrazine and anti-scale) form CA. Two indicator variables \( x_1 \) and \( x_2 \) are included for the exogenous latent variable OP Model. \( \delta_p, \delta_d \) signify the measurement error for DW, SPP and CA, respectively. The \( \omega_p \) and \( \omega_o \) characterize the measurement error for OP. In the final set, output deals with the relationship between an endogenous latent variables (\( \eta_o \)) and the indicator variables \( \gamma_3, \gamma_4, \gamma_5, \gamma_6 \) \( \delta_p \) and \( \delta_o \) 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 \( \eta, \eta_o, y_i (i = 1, 2, \ldots, 10) \). Seawater, boiler, steam, steam
turbine, morphine, sodium triphosphate, anti-scale, hydrazine and the exogenous variables that are described as $\xi_i, x_i \ (i = 1, 2)$ (chemical and fuel treatment) fall under measured variables, $\lambda_i \ (i = 1, \ldots, 10)$ represent the correlation coefficients between endogenous latent variables, $\omega_i \ (i = 1, 2)$ and the indicator variables and the correlation coefficients between the exogenous latent variable and indicator variables. Also, according to researchers, $\xi_i$ signifies the error of estimate for the endogenous latent variables, $i = 1, 2, 3, 4$ while $\delta_i$ for the regression errors and $\Lambda_{s1}, \Lambda_{s2}, \Omega_{s1}, \Omega_{s2}, \Lambda_{s3}, \Lambda_{s4}, \Omega_{s3}, \Omega_{s4}$ 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 utilized 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:

$$\eta_i = \Lambda_{s1} \eta_1 + \Lambda_{s2} \eta_2 + \Lambda_{s3} \eta_3 + \Omega_{s1} \xi_1 + \xi_i$$

where the coefficient $\Omega_i$ 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):

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \delta_3 \\ \delta_4 \end{bmatrix}$$

The measurement model of endogenous latent variable (SPP):

$$\begin{bmatrix} y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} \lambda_3 \\ \lambda_4 \end{bmatrix} \begin{bmatrix} \eta_3 \\ \eta_4 \end{bmatrix} + \begin{bmatrix} \delta_5 \\ \delta_6 \end{bmatrix}$$

The measurement model of endogenous latent variable (CA):

$$\begin{bmatrix} y_5 \\ y_6 \\ y_7 \\ y_8 \end{bmatrix} = \begin{bmatrix} \lambda_5 \\ \lambda_6 \\ \lambda_7 \\ \lambda_8 \end{bmatrix} \begin{bmatrix} \eta_5 \\ \eta_6 \\ \eta_7 \\ \eta_8 \end{bmatrix} + \begin{bmatrix} \delta_9 \\ \delta_10 \\ \delta_11 \\ \delta_12 \end{bmatrix}$$

The measurement model of exogenous latent variable (OP):

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}$$
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:

\[ \eta_b = \Lambda_b \eta_i + \Lambda_{b1} \eta_y + \Omega_{b1} \zeta_1 + \zeta_1 \]  

(10)

The SPP structural model is expressed as:

\[ \eta_s = \Omega_{s1} \xi_1 + \xi_1 \]  

(11)

The CA structural model is expressed as:

\[ \eta_c = \Lambda_{ca} \eta_i + \Omega_{ca} \zeta_2 + \zeta_2 \]  

(12)

The output structural model was stated as:

\[ \eta_o = \Lambda_o \eta_i + \Lambda_{o1} \eta_y + \Lambda_{o2} \eta_y + \Omega_{o1} \zeta_1 + \zeta_1 \]  

(13)

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