Journal of Information Technology Management

Print ISSN: 2008-5893

Online ISSN: 2423-5059

Comparison between the Conventional Partial Least Squares (Pls) and the Robust Partial Least Squares (Rpls-Sem) Through Winsorization Approach

GholamReza Zandi

Associate Professor, University Kuala Lumpur (UniKL) Business School, Malaysia. E-mail: zandi@unikl.edu.my

Fadya Ramadan Shakhim*

*Corresponding Author, Department of Statistics, Faculty of Science, Al-Zawiya University, Al-Zawiya, Libya. E-mail: f.shakhim@zu.edu.ly

Zulkifley Mohamed

Department of Mathematics, Faculty of Science and Mathematics, University Pendidikan Sultan Idris 35900 Tanjong Malim, Perak, Malaysia. E-mail: zulkifley@fsmt.upsi.edu.my

Amel Saad Alshargawi

Department of Statistics, Faculty of Science, Tripoli University, Tripoli, Libya. E-mail: amelsh@hotmail.com

Abstract

This study compared the performance of the partial least squares-structural equation modelling (PLS-SEM) and the robust partial least squares -structural equation modelling (RPLS-SEM) methods through Winsorisation approach The inputs and the outputs used in this model were based on the electricity generation data, derived from the Al-Zawiya Steam Power Plant, Libya. Furthermore, the researchers compared the novel RPLS-SEM approach with the traditional PLS-SEM approach and noted that the novel RPLS-SEM method was more efficient compared to PLS-SEM.

Keywords: Partial Least Square-Path Modelling (PLS-SEM), Robust Partial Least Squares (RPLS-SEM), Structural Equation Modelling (SEM), Winsorization, Steam Power Plant, SmartPLS3.

Journal of Information Technology Management, 2022, Vol. 14, No.4, pp. 87-94
Published by University of Tehran, Faculty of Management
doi: https://doi.org/ 10.22059/jitm.2022.88291
Article Type: Research Paper

Received: April 13, 2022
Received in revised form: May 25, 2022
Accepted: June 28, 2022
Published online: July 27, 2022

Introduction

PLS-SEM was seen to be the technique which could be applied if the predictor variables displayed high or perfect multicollinearity (Hair et al., 2017). On the other hand, robust methods were developed for decreasing or eliminating the effects of all outliers (Maronna and Zamar, 2002). . In this study, the researchers proposed a novel RPLS-SEM model which was based on the robustification of a covariance matrix that was used in the classical PLS algorithm. This study also chose a robust covariance estimator, which used the Winsorisation estimator for estimating the covariance matrix in the multivariate dataset for decreasing the harmful effect of the outliers. Croux and Rousseeuw (1992) stated that a robust estimator (or a Winsorised estimator, W) could be used instead of the popular mean vector, which could substitute the inverse of the Winsorised covariance matrix. This technique was called the Robust Straightforward Implementation of the statistically- inspired Modification of PLS (RSIMPLS). Thereafter, the researcher compared the novel and the classical PLS-SEM models.

Data

In this study, the researchers collected the secondary data from the Al-Zawiya Steam Power plant in Libya. Real data related to power generation was collected and compiled by the Technical Department of the AL-Zawiya Oil Refining Company the important input parameters for freshwater and power generation, which included:

Desalination unit (DW), i.e., the amount of steam (tons/day) and seawater (m3/ day) needed for freshwater production.

Steam Power Plant (SPP) requirements - steam turbine (tons/day) and boiler (m3/ day of distilled water).

Chemical Additives (CA) - Phosphate (kg/day), Morphine, anti-scale and hydrazine (L/day).

Maintenance and Operation (OP) – mean costs for the chemical treatment and fuel (LYD/day).

Figure 1 presents an arrow diagram, wherein the researcher assumed that every MV (measured variable) block could be summarized by an LV (unmeasured). The following endogenous LV symbols were suggested: DW is desalination units represent steam (D1) and seawater (D2) ; Steam power plant SPP represent steam turbines (S1) and boiler (S2) ; while CA represents chemical additive consists of four indicator variables are quantity of sodium triphosphate,(C1), hydrazine (C2), morphine (C3) and anti-scale (C4) needed; whereas the exogenous latent variables were represented as OP includes chemical treatment (O1) and fuel-related costs (O2); and Output is electricity (P1) and fresh water supply (P2) . The general

structural and measurement models for DW, SPP, CA, OP and Output were as explained in figure1.

Data Analysis

The researchers used a SmartPLS3 software (Ringle et al., 2015) as it offers appropriate techniques for facilitating the fitting of the specific model. This software generated the data processing output, which included the general model fit statistics and all parameter estimates, described in Figure 1 and Figure 2. The causality model presented in this figure summarized the steps involved in a structural regression of an RPLS-SEM model.

The quality of the PLS-SEM model was assessed using two steps: initially, the measurement model was assessed and if it satisfied all criteria, the structural model was evaluated. The measurement model was investigated using parameters like Cronbach's alpha, composite reliability and Average variance extracted (AVE). Tables 1 and 2 presented the RPLS-SEM and PLS-SEM model indices. The results indicated that the Cronbach's alpha values for both the models were greater than 0.7, which showed the indicator homogeneity. Furthermore, the cut-off values for the composite reliability were larger than 0.8, while the AVE was greater than 0.5, which indicated that more than 50% of the variance of the indicators could be explained (Chin, 2010).

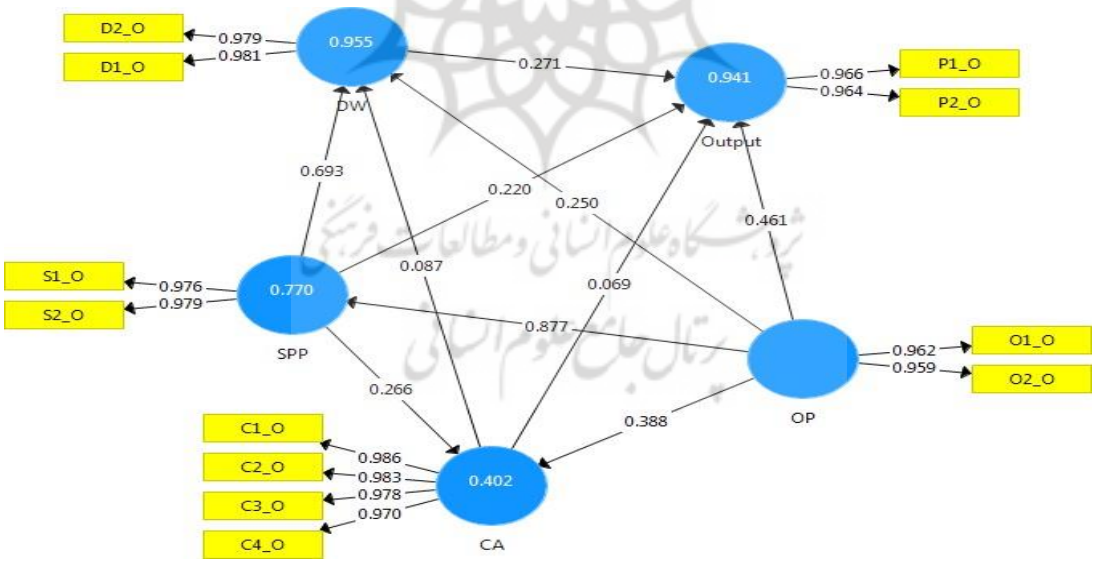

Figure 1. Partial Least Square-Path Modelling

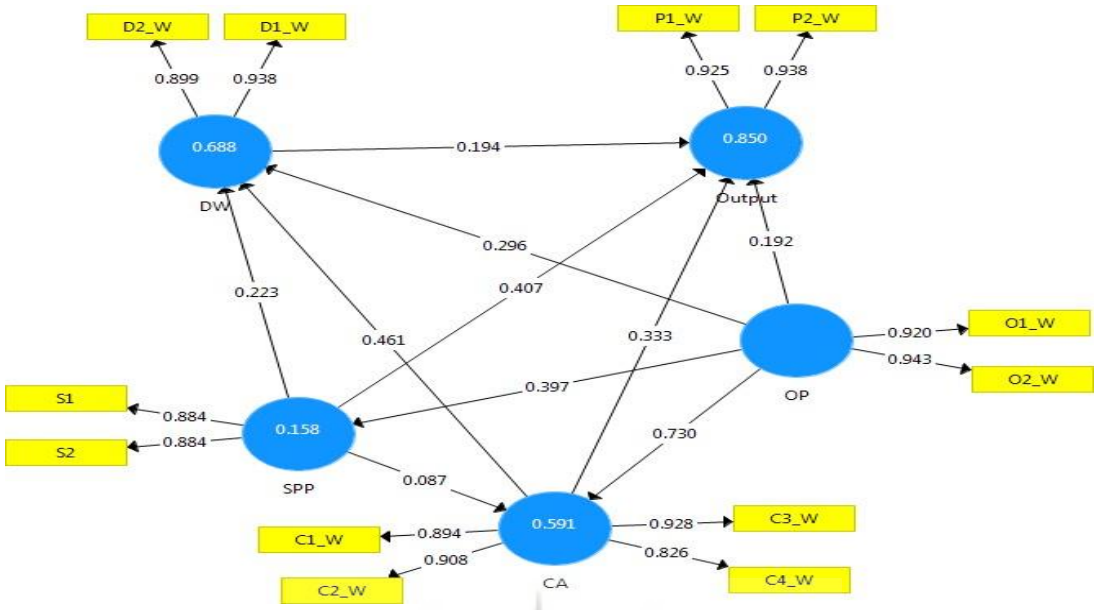

Figure 2. Robust Partial Least Square-Path Modelling

Table 1. Reliability Assessment for the RPLS-SEM

Construct	Composite Reliability	AVE	Squared Root of AVE	Cronbach's Alpha
DW	0.916	0.844	0.919	0.818
SPP	0.877	0.781	0.884	0.720
CA	0.939	0.792	0.890	0.912
OP	0.929	0.868	0.932	0.849
Output	0.929	0.867	0.931	0.847

Table 2. Reliability Assessment of the PLS-SEM

Construct	Composite Reliability	AVE	Cronbach's Alpha
DW	0.980	0.960	0.959
SPP	0.977	0.955	0.953
CA	0.989	0.959	0.986
OP	0.959	0.922	0.915
Output	0.964	0.931	0.926

All the indices for the PLS-SEM were higher due to the presence of the internal consistency, based on the average correlation amongst the items (multicollinearity).

Secondly, the inner model quality was assessed by investigating the indices of the coefficient of determination, bootstrapping, redundancy index, and the Goodness of Fit (GoF) index. The structural model assessment includes the testing of the relationships between all

model constructs shown in Tables 3 and 4. The RPLS-SEM model showed no significant fluctuations, which showed that the RPLS-SEM was better than the PLS-SEM model. Esposito Vinzi et al. (2010) stated that the assessment of the non-significant path coefficients should be carried out carefully, due to the presence of multicollinearity. Finally, the PLS-SEM model showed a higher coefficient of determination, redundancy index, and GoF values since these indices were based on the correlation (multicollinearity issue).

Table 3 and Table 4 present the results of the bootstrapping technique conducted on the different resampled datasets. The significant fluctuations noted in the results were based on the differing number of resampling data groups, except in 500 re- sampled data sets, where the RPLS-SEM model showed a good performance.

Table 3. Structural PLS-SEM Model Analysed Using the Bootstrap Process

Relationship	***T* – Statistic**	***P – value***
DW →Output	3.317	0.198
SPP →Output	2.358	0.000**
CA →Output	0.515	0.284
OP →Output	2.501	0.019*
SPP →DW	5.874	0.548
SPP →CA	1.073	0.607
CA →DW	0.601	0.044*
OP →DW	2.017	0.019*
OP →SPP	5.340	0.000**
OP →CA	1.973	0.013*

* indicates the significance at 0.05 level of significance.

** indicates the significance at 0.01 level.

Table 4. RPLS-SEM Structural Model Assessment Using the Bootstrap Process

Relationship	**T – Statistic**	**P – value**
DW →Output	2.287	0.023*
SPP →Output	9.073	0.000**
CA →Output	3.883	0.000**
OP →Output	2.072	0.039*
SPP →DW	2.865	0.004**
SPP →CA	1,071	0.285
CA →DW	3.803	0.000**
OP →DW	2.352	0.019*
OP →SPP	3.316	0.001**
OP →CA	11.346	0.000**

* Significance at 0.05 level ** significance at 0.01 level

The data showed that multicollinearity existed in the PLS-SEM model (Table 5); whereas the variance inflation factors (VIF) values in the RPLS-SEM were seen to be less than 5 (Table 6). Hence, the researcher proposed the RPLS-SEM for overcoming the multicollinearity in the study.

Table 5. VIF Values for the Outer PLS-SEM Model

Predictor	VIF
D1	**6.545**
D1	**6.545**
S1	**5.861**
S2	**5.861**
C1	**20.491**
C2	**18.104**
C3	**12.612**
C4	**11.195**
O1	3.463
O2	3.463
P1	3.880
P2	3.880

Table 6. VIF Values for the Outer RPLS-SEM Model

Predictor	VIF
D1	1.920
D1	1.920
S1	1.464
S2	1.464
C1	3.116
C2	3.239
C3	3.883
C4	2.049
O1	2.193
O2	2.193
P1	2.177
P2	2.177

The results compared the performances of the PLS-SEM and the RPLS-SEM and showed that the RPLS-SEM was more effective than the PLS-SEM model in overcoming the multicollinearity problem.

Conclusion

The results and the analysis of the data set derived from the Libyan Oil Refining sector showed that the novel RPLS-SEM model was very effective and robust. This model showed a higher efficiency and displayed a better predictive capacity compared to the conventional

PLS-SEM model. Finally, it was stated that this robust model was able to efficiently cope with the data set and provide robust predictions.

Conflict of interest

The authors declare no potential conflict of interest regarding the publication of this work. In addition, the ethical issues including plagiarism, informed consent, misconduct, data fabrication and, or falsification, double publication and, or submission, and redundancy have been completely witnessed by the authors.

Funding

The author (s) received no financial support for the research, authorship, and /or publication of this article

References

Chin, W. W. (2010). How to write up and report PLS analyses. In *Handbook of partial least squares* (pp. 655–690). Springer.

Clark, R. G. (1995). Winsorization methods in sample surveys.

Critical Reviews in Analytical Chemistry, *36*(3–4), 221–242.

Croux, C., & Rousseeuw, P. J. (1992). Time-efficient algorithms for two highly robust estimators of scale. In *ComputationalStatistics* (pp. 411–428). Springer.

Enaami, M. E., Mohamed, Z., & Ghani, S. A. (2013). Model development for wheat production: Outliers and multicollinearityproblem in Cobb-Douglas production function. *Emirates Journal of Food and Agriculture*, 81–88.

Esposito Vinzi, V., Chin, W. W., Henseler, J., & Wang, H. (2010). *Handbook of partial least squares: Concepts, methods and applications*. Heidelberg, Dordrecht, London, New York: Springer.

Favre-Martinoz, C., Haziza, D., & Beaumont, J.-F. (2015). A method of determining the winsorization threshold, with an application to domain estimation. *Survey Methodology*, *41*(1), 57–77.

Filzmoser, P., Maronna, R., & Werner, M. (2008). Outlier identification in high dimensions. *Computational Statistics &Data Analysis*, *52*(3), 1694–1711.

Fornell, C. (1994). Partial least squares. *Advanced Methods of Marketing Research*.

Garson, G. D. (2016). Partial Least Squares: Regression and structural equation models. Statistical Associates Blue BookSeries. Statistical Associates Publishing: Asheboro, USA.

Hair Jr, J. F., Sarstedt, M., Ringle, C. M., & Gudergan, S. P. (2017). *Advanced issues in partial least squares structuralequation modeling*. SAGE Publications.

Hair, J. F., Sarstedt, M., Ringle, C. M., & Mena, J. A. (2012). An assessment of the use of partial least squares structural equation modeling in marketing research. *Journal of the Academy of Marketing Science*, *40*(3), 414–433.

Jannoo, Z., Yap, B. W., Auchoybur, N., & Lazim, M. A. (2014). The effect of no normality on CB-SEM and PLS-SEM path estimates. *International Journal of Mathematical, Computational,*

Physical and Quantum Engineering, *8*(2), 285–291. Maronna, R. A., & Zamar, R. H. (2002). Robust estimates of location and dispersion for high-dimensional datasets. *Technometrics*, *44*(4), 307–317.

Kalogirou, S. A. (2013). *Solar energy engineering: processes and systems*. Academic Press.

Maronna, R. A., Martin, D., & Yohai, R. S. (2006). Wiley Series in Probability and Statistics. *Robust Statistics: Theory andMethods*, 404–414.

Ringle, C. M., Wende, S., & Becker, J.-M. (2015). SmartPLS 3. *Boenningstedt: SmartPLS GmbH, Http://Www. Smartpls.Com*.

Rousseeuw, P. J., & Hubert, M. (2018). Anomaly detection by robust statistics. *Wiley Interdisciplinary Reviews: DataMining and Knowledge Discovery*, *8*(2), e1236.

Rousseeuw, P. J., Debruyne, M., Engelen, S., & Hubert, M. (2006). Robustness and outlier detection in chemometrics.

Sarstedt, M., Ringle, C. M., & Hair, J. F. (2017). Partial least squares structural equation modeling. In *Handbook of marketresearch* (pp. 1–40). Springer.

Bibliographic information of this paper for citing:

Zandi. GholamReza; Shakhim, Fadya Ramadan; Mohamed, Zulkifley & Alshargawi, Amel Saad (2022). Comparison between the Conventional Partial Least Squares (Pls) and the Robust Partial Least Squares (Rpls-Sem) Through Winsorization Approach. *Journal of Information Technology Management*, 14 (4), 87-94. https://doi.org/ 10.22059/jitm.2022.88291