Advances in Mathematics: Scientific Journal **9** (2020), no.3, 1053–1060
ISSN: 1857-8365 (printed); 1857-8438 (electronic)
https://doi.org/10.37418/amsj.9.3.30

GENERAL DISCUSSION ON SMALL SAMPLE SIZE IN CONSISTENT PLS

ASYRAF AFTHANORHAN

ABSTRACT. An optimum sample size is an essential component of any research. If a study does not have an optimum sample size, the significance of the results may not be detected. This implies that the study would lack power to detect the significance of differences because inadequate of sample size. There is some veracity claim about the usefulness of Consistent PLS when research apply small sample size. Therefore, this study is aimed to discuss particularly on how effective of Consistent PLS when using the small sample size. Based on evidences, one can concluded that the Consistent PLS is not appropriate for the small size and the researchers need to collect sufficient sample size to meet the minimum threshold of statistical power. The recommendation for improvement also discussed.

1. INTRODUCTION

One of the most important asked questions of a research method is how large a sample is needed for a specific research project (Maxwell, Kelley, and Rausch, 2008, cite20). As consequences, they refer Cohen (1992) [3], Israel (1992) [15], Hair et al. [16], and Westland (2010), [26] approaches for determining the appropriateness of sample size for their works. Sample size planning is a vital aspect of research for those aspires to confirm their works using Null Hypothesis Statistical Significance Testing (NHSST). Hence, one must use the

2010 *Mathematics Subject Classification.* 62D05.
Key words and phrases. Consistent PLS, Small Sample, Statistical Power.

correct approach for computing the sample size appropriate to the study design and its subtype.

There are countless interpretation to describe the behavior of Type I and Type II error rates. Type I error rate is defined when the result showed the effect is statistically occurred in the sample but it is actually does not exist in the true population which also could be considered as false positive. In contrast, Type II error rate is inferred when the result showed the effect does not occurred in the sample but it is actually exist in the true population which could be noted as false negative (Goodhue et al., 2006 [12]; Dijkstra and Henseler, 2015 [7]; Aimran et al., 2017 [1]; O'Brien and Castelloe, 2007 [23]; Ronkko and Evermann, 2013 [24]; Maxwell, Kelley, and Rausch, 2008 [20]; Nickerson, 2000 [22]).

The effects of both types of crucial error should be reduced if the researchers interest to obtain accuracy in parameter estimates (Maxwell, Kelley, and Rausch, 2008 [20]; Cohen, 1994 [4]; Goodman and Berlin, 1994 [13]; Kelley and Rausch, 2006 [17]) especially when using the confirmation technique.

Large sample size is generally leads to increase precision and accuracy when estimating parameter values which usually need to have sufficient statistical power (Cohen, 1994 [4]). Means that, high sample could reduce the risk of crucial errors for obtaining accurate parameter estimates. This, however, in some situations, this can result from the presence of systematic error or strong dependence in the data when the sample is too large (Bartlett, Kotrlik, and Higgins, 2001 [2]; Fugard and Potts, 2015 [10]; Francis et al., 2010 [9]).

After recognizing that the traditional PLS is not suitable for examining the common factor model, there were many efforts to develop the alternative approaches as Consistent PLS for estimating measurement and structural model. In particular, the Consistent PLS has an ability to estimate consistently on those common factor models. In addition, Hair et al. (2016) [14] conjectured that Consistent PLS is even useful for small sample size because it inherits the strength of traditional PLS and thus is more benefited than of CBSEM. Apart from that, Marcoulides and Saunders, 2006 [19], Goodhue, Lewis and Thompson (2012) [11] believe that there is some "magic" in the PLS method that allows to obtain the result with less than 100 of sample size.

This, however, the usefulness of Consistent PLS with small sample has never being simply stated by the proponent of Consistent PLS (Dijkstra and Henseler, 2015 [7]; Dijkstra, 2010 [5]), instead, they claimed that more observations are needed to intensify the power analysis of Consistent PLS. The statement carries the meaning that small sample cannot be performed with Consistent PLS, thus indicates that future research should be devoted for this work.

2. Findings

In this section, the researcher attempt to explore whether the behavior of Consistent PLS can be useful for small sample size or not. By doing so, the given formula addressed by Dijkstra (2010) [6] for factor loading is discussed. By addressing on this statements, the reason of this study opting factor loading is because this is the important part when dealing the measurement model. Based on the previous literatures, the PLS has being improved by attenuating the measurement error through Minimum Residual approach (MINRES) in order to vanish the inconsistency estimates problem. As such, the veracity claim Consistent PLS is may appropriate for small sample size is myth where this statements are never claimed by any proponents themselves as followed.

$$y_i = \lambda_i \eta_i + \varepsilon_i \,,$$

where y_i is the vector of indicator is composed of at least two sub vectors (measuring a unique latent variable and sub vector contains at least two components). λ is the vector of factor loadings and the unobservable latent variable η is real-valued, and ϵ is the vector of idiosyncratic error that have the same dimensions as y_i. At this stage, the vector of idiosyncratic error is included in the vector of indicator where the real weight for vector of indicator cannot be estimated accurately. In this essence, the factor loading with Consistent PLS is may overestimated due to the presence of idiosyncratic error where the pure error should be assessed

$$\Sigma_{ii} = E y_i y_i^T + \Theta_i \,. \tag{2.1}$$

The latter has zero mean and unit variance. Based on this, Consistent PLS does not able to run the model with missing value where the mean of this vector has being set value as zero. The covariance matrix Σ_{ij} of y_i can be written as Equation (2.1). Θ_i is the covariance matrix of measurement error that is diagonal with non-negative element (positive definite). With this restricted rule in covariance matrix, one can be inferred that Consistent PLS may has Heywood Cases in certain situations where the traditional PLS did not (Dijkstra and Henseler, 2015 [7]). In this case, Consistent PLS may able to detect the mis-specified model as this effect also occurred for CBSEM.

$$\Sigma_{ii} = Ey_i y_i^T = \rho_{ij} \lambda_i \lambda_j^T ,$$

where ρ_{ij} is the correlation between η_i and η_j. One of the reason why this approach being called as composite variable. The composite is not only inheriting the measurement error but also dependent on the correlation between variable to produce the covariance matrix. As a consequence, the way this approach has being accounted is similar to item-total correlation in the case of reliability. Other than that, Dijkstra attempt to reduce the effects of measurement error by maximizing the reliability of indicators. Therefore, Dijkstra and Henseler (2015) [7] proposed a new reliability to make sure this reliability is comparable as composite reliability when assessing the measurement model. In the first place, the sample data should be assumed to be standardized before being analyze and eventually the observed data has zero mean and variance.

$$\bar{w}_i = \sum_{j \in C(i)} sign_{ij}.\ S_{ij}\ \bar{w}_j\ . \tag{2.2}$$

In the assessment of measurement model, Consistent PLS use the mode A algorithm generally as it was verified as the most stable algorithm (Lohmoller, 2013 [18]). Because the algorithm converges very fast for estimate the weight vector (for example, for the model analyzed in this thesis, the iteration needed are less than the CBSEM). With these weights (see Equation (2.2)), sample proxies are defined for the latent variable or construct with the same dimension as y_i. Since of the probability limit is approaching to one when the sample size is large, the model will converge with a probability tended to one (Dijkstra, 1983; Dijkstra, 2010 [5,6]). Means that, if the small sample size is utilized, the model will be fail to converge arbitrary to one. Using these weights, the factor loading of the corresponding latent variables are used.

$$\bar{w}_i = \propto \sum_{j \in C(i)} sign_{ij} \sum_{ij} \bar{w}_j = \sum_{j \in C(i)} sign_{ij}.\rho_{ij} \lambda_i \lambda_j^T\ \bar{w}_j\ . \tag{2.3}$$

Equation (2.3) shows the probability limit for vector of weight that can be derived from Equation (2.2). Based on this equation, $\bar{w}$ is directly proportional to λ_i. Because of the unit variance, the proportionality constant has to be such that

$\bar{w}_i^T \sum_{ii} \bar{w}_i = 1$ This entails that:

$$\bar{w}_i = \lambda_i \div (\lambda_i^T {\sum}_{ii} \lambda_i)^{1/2} . \tag{2.4}$$

Finally, the factor loading can be estimated directly using of this scalar. As can be seen at the Equation (2.4), scalar vector is not well-defined for the small sample size (Dijkstra and Henseler, 2015) where the loading will always have overestimated. Other than that, if the pairs of error of the same block are suspected to be correlated (check their VIF), one can delete corresponding terms both numerical and denominator. Unfortunately, there is no close form for minimizing the Euclidean distance after re-specify the model. Therefore, the estimated values may deviate from the true values. As a consequence, this study summarized the output from three models for small sample size for verification.

$$C_i = \left[\frac{\bar{w}_i^T (S_{ii} - diag(S_{ii}))\bar{w}_i}{\bar{w}_i^T (\bar{w}_i \bar{w}_i^T - diag(\bar{w}_i \bar{w}_i^T))\bar{w}_i} \right] .$$

3. Discussion on Small Sample Size

After recognizing that traditional PLS is unsuitable for examining the common factor model, there have been some efforts to develop alternative approaches as Consistent PLS for estimating measurement and structural model. In particular, the Consistent PLS has an ability to estimate consistently on those types of models. In addition, Hair et al. (2016) [14] conjectured that Consistent PLS is useful even for small sample size as traditional PLS which one could address that this novel approach is more benefited than of CBSEM. In the case of small sample, we have recently been witnessing the performance of Consistent PLS which explicitly not competent as they are. This marketing effort has been extremely demonstrated with the usefulness of consistent PLS by Hair et al. (2016) in A Primer on Partial Least Square Path Modeling and Advance Issue in PLS-PM (Hair et al. 2016). Concurrently, however, many researchers, including some paper about method appear to have forgotten that Consistent PLS is not equivalent to traditional PLS.

The question of whether Consistent PLS is suitable for small sample size as recommended by that literature is not useful, because the proponent of Consistent PLS clearly explain that it is not able for that sample (Dijkstra and Henseler, 2015; Dijkstra, 2010; Dijkstra and Schmerlleh-Engel, 2014 [8]). In earlier discussion, the formula for Consistent PLS is addressed, appear that positive bias of indicator loadings are indeed existing with Consistent PLS. At the same time, the result showed that the scalar vector is not well-defined for small sample but it is sufficient for large samples (Dijkstra and Schmerlleh- Engel, 2014) regardless of the quality indicators.

The most common capabilities that researchers usually want in SEM techniques are the ability to estimate the model with small sample size, as well as no stringent assumptions (Shah and Goldstein, 2006 [25]; Mohamad et al., 2019 [21]). Based on this finding thus far, it is clear that Consistent PLS does not provide these capabilities.

For this examples, at small sample, the Consistent PLS yield large standard error than CBSEM when detecting less complex model, lack of discriminant when applied under Fornell and Larcker criterion, SRMR index always failed, parameter estimates of path coefficients are biased, and low statistical power. Moreover, when detecting for the parameter accuracy (i.e., indicator loadings and construct correlations), the Consistent PLS is not permissible under all conditions but remain acceptable for theory testing because the presence of consistency estimates. Therefore, simply stated Consistent PLS is useful for small sample to place the technique becoming more and more pronounced in management research can undeservedly enhance its reputation merely by association.

Consistent PLS as an alternative to CBSEM technique for small sample not only obscures what the method actually does and implies capabilities it does not have, but also leads to omission of erroneous analyses that could be affect wholly research specific project. At best, the researchers should collect sufficient data to avoid this problem (Ronkko and Evermann, 2013 [24]) because more observation are required to obtain more information from the samples (Westland, 2015). Thus, the power analysis could be enhanced if using more observation in the case of Consistent PLS context (Dijkstra and Henseler, 2015). Another potential useful option is to extend the use of two stage least square estimator (2SLS) which making this estimator is permitted for recursive model as well to estimate consistent parameter estimates

REFERENCES

[1] A. N. AIMRAN, S. AHMAD, A. AFTHANORHAN, Z. AWANG: *The development of comparative bias index*, In AIP Conference Proceedings, **1870**(1) (2017), 060008.

[2] BARTLETT, J. E. KOTRLIK, C. C. HIGGINS: *Organizational research: Determining appropriate sample size in survey research*, Information technology, learning, and performance journal, **19**(1) (2001), 43–47.

[3] J. COHEN: *A power primer*, Psychological bulletin, **112**(1) (1992), 145–155.

[4] J. COHEN: *The earth is round (p <.05)*, Am. Psychol., **49** (1994), 997–1003.

[5] T. K. DIJKSTRA: *Some comments on maximum likelihood and partial least squares methods*, Journal of Econometrics, **22**(1) (1983), 67–90.

[6] T. K. DIJKSTRA: *Latent variables and indices: Herman Wold's basic design and partial least squares*, In Handbook of partial least squares, 23–46, 2010.

[7] T. K. DIJKSTRA, J. HENSELER: *Consistent and asymptotically normal PLS estimators for linear structural equations*, Computational Statistics and Data Analysis, **81** (2015), 10–23.

[8] T. K. DIJKSTRA, K. SCHERMELLEH-ENGEL: *Consistent partial least squares for nonlinear structural equation models*, Psychometrika, **79**(4) (2014), 585–604.

[9] J. J. FRANCIS, M. JOHNSTON,C. ROBERTSON, L. GLIDEWELL, V. ENTWISTLE, M. P. ECCLES, J. M. GRIMSHAW: *What is an adequate sample size? Operationalising data saturation for theory-based interview studies*, Psychology and Health, **25**(10) (2010), 1229–1245.

[10] A. J. FUGARD, H. W. POTTS: *Supporting thinking on sample sizes for thematic analyses: a quantitative tool*, International Journal of Social Research Methodology, **18**(6) (2016), 669–684.

[11] D. L. GOODHUE, W. LEWIS, R. THOMPSON: *Does PLS have advantages for small sample size or nonnormal data*, Mis Quarterly, **36**(3) (2012), 891–1001.

[12] D. GOODHUE, W. LEWIS, R. THOMPSON: *PLS, small sample size, and statistical power in MIS research*, In Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06), **8** (2006), 202b-202b.

[13] S. N. GOODMAN, J. A. BERLIN: it The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results, Annals of Internal Medicine, **121**(3) (1994), 200–206.

[14] J. F. HAIR, G. T. M. HULT, C. RINGLE, M. SARSTEDT: *A primer on partial least squares structural equation modeling,* Sage Publications, 2016.

[15] J. F. HAIR, W. C. BLACK, B. J. BABIN, R. E. ANDERSON, R. L. TATHAM: *Analise multivariada de dados,* Bookman Editora, 2009.

[16] G. D. ISRAEL: *Determining sample size: University of Florida Cooperative Extension Service*, Institute of Food and Agriculture Sciences, EDIS, 1992.

[17] K. KELLEY, J. R. RAUSCH: *Sample size planning for the standardized mean difference: accuracy in parameter estimation via narrow confidence intervals*, Psychological methods, **11**(4) (2016), 363–373.

[18] J. B. LOHMOLLER: *Latent variable path modeling with partial least squares*, Springer Science and Business Media, 2013.

[19] G. A. MARCOULIDES, C. SAUNDERS: *Editor's comments: PLS: a silver bullet?* MIS quarterly, 2006.

[20] S. E. MAXWELL, K. KELLEY, J. R. RAUSCH: *Sample size planning for statistical power and accuracy in parameter estimation*, Annu. Rev. Psychol., **59** (2008), 537–563.

[21] A. MOHAMAD, A. AFTHANORHAN, Z. AWANG, M. MOHAMMAD: *Comparison Between CB-SEM and PLSSEM: Testing and Confirming the Maqasid Syariah Quality of Life Measurement Model*, The Journal of Social Sciences Research, **5**(3) (2019), 608–614.

[22] R. S. NICKERSON: *Null hypothesis significance testing: a review of an old and continuing controversy*, Psychological methods, **5**(2) (2000), 241–251.

[23] R. G. O'BRIEN, J. CASTELLOE: *Sample size analysis for traditional hypothesis testing: concepts and issues*, Pharmaceutical statistics using SAS: A practical guide, (2007), 237–271.

[24] M. RONKKO, J. EVERMANN: *A critical examination of common beliefs about partial least squares path modeling*, Organizational Research Methods, (2013), 1094428112474693.

[25] R. SHAH, S. M. GOLDSTEIN: it Use of structural equation modeling in operations management research: Looking back and forward, Journal of Operations Management, **24**(2) (2006), 148–169.

[26] J. C. WESTLAND: *Lower bounds on sample size in structural equation modeling*, Electronic Commerce Research and Applications, **9**(6) (2010), 476–487.

FACULTY OF BUSINESS AND MANAGEMENT
UNIVERSITI SULTAN ZAINAL ABIDIN (UNISZA)
E-mail address: asyrafafthanorhan@unisza.edu.my