

APPLYING PARTIAL LEAST SQUARES IN TOURISM AND HOSPITALITY RESEARCH

This page intentionally left blank

APPLYING PARTIAL LEAST SQUARES IN TOURISM AND HOSPITALITY RESEARCH

EDITORS:

FAIZAN ALI

University of South Florida Sarasota-Manatee, USA

S. MOSTAFA RASOOLIMANESH

Taylor's University, Malaysia

CIHAN COBANOGLU

University of South Florida Sarasota-Manatee, USA



United Kingdom – North America – Japan – India – Malaysia – China

Emerald Publishing Limited
Howard House, Wagon Lane, Bingley BD16 1WA, UK

First edition 2019

Copyright © 2019 Emerald Publishing Limited.

Reprints and permissions service

Contact: permissions@emeraldinsight.com

No part of this book may be reproduced, stored in a retrieval system, transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without either the prior written permission of the publisher or a licence permitting restricted copying issued in the UK by The Copyright Licensing Agency and in the USA by The Copyright Clearance Center. Any opinions expressed in the chapters are those of the authors. Whilst Emerald makes every effort to ensure the quality and accuracy of its content, Emerald makes no representation implied or otherwise, as to the chapters' suitability and application and disclaims any warranties, express or implied, to their use.

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN: 978-1-78756-700-9 (Print)

ISBN: 978-1-78756-699-6 (Online)

ISBN: 978-1-78756-701-6 (Epub)



ISOQAR certified
Management System,
awarded to Emerald
for adherence to
Environmental
standard
ISO 14001:2004.

Certificate Number 1985
ISO 14001



INVESTOR IN PEOPLE

Contents

Preface	vii
Foreword	xi
About the Authors	xvii
Chapter 1 Minimum Sample Size Estimation in PLS-SEM: An Application in Tourism and Hospitality Research <i>Ned Kock</i>	1
Chapter 2 New Guidelines for the Use of PLS Path Modeling in Hospitality, Travel, and Tourism Research <i>Jörg Henseler, Tobias Müller and Florian Schuberth</i>	17
Chapter 3 Predictions from Partial Least Squares Models <i>Nicholas P. Danks and Soumya Ray</i>	35
Chapter 4 PLS Path Modeling in Hospitality and Tourism Research: The Golden Age and Days of Future Past <i>Hengky Latan</i>	53
Chapter 5 Hotel Employees' Use of Smartphones and Performance: Reflective–Formative Estimation Approach <i>Minwoo Lee, Kawon Kim, Kyung Young Lee and Jung Hwa Hong</i>	85
Chapter 6 Loyalty to Rural Tourism in Brazil: The Moderating Effect of the Purchase Frequency <i>Mara Mataveli and Alfonso J. Gil</i>	109
Chapter 7 HRM Practices and Employee Retention: The Moderating Effect of Work Environment <i>Palwasha Bibi, Ashfaq Ahmad and Abdul H. A. Majid</i>	129

Chapter 8 Do Museums' Websites Boost Visitors' Intentions? A PLS Multigroup Comparison <i>Jesús García-Madariaga, Nuria Recuero Virto, María Francisca Blasco López and Joaquín Aldas Manzano</i>	153
Chapter 9 The Effect of High and Low Environmental Consciousness Regarding Brazilian Restaurants: A Multigroup Analysis Using PLS <i>Carlos Alberto Alves, Claudio José Stefanini and Leonardo Aureliano da Silva</i>	185
Chapter 10 Analyzing the Effects of Online and Offline Communication in the Hotel and Restaurant Industry: The PLS Approach <i>Maja Šerić and Đurđana Ozretić-Došen</i>	211
<i>About the Editors</i>	233
<i>Index</i>	235

Preface

Faizan Ali (University of South Florida Sarasota Manatee)

S. Mostafa Rasoolimanesh (Taylor's University)

Cihan Cobanoglu (University of South Florida Sarasota Manatee)

Partial least squares-structural equation modeling (PLS-SEM) is a multivariate statistical technique and its usage in various disciplines is increasing. Considering this increase in the application of PLS-SEM, numerous scholars have reviewed its usage in accounting, business research, strategic management, marketing, management information system, tourism and hospitality research, etc. Review studies on the usage of PLS-SEM in tourism (do Valle & Assaker, 2016) and hospitality research (Ali, Rasoolimanesh, Sarstedt, Ringle, & Ryu, 2018) indicate an increasing dissemination of PLS-SEM in tourism and hospitality research. Researchers in tourism and hospitality seem to be aware of sample size issues in PLS-SEM, which have attracted considerable attention in recent years. In addition, the reporting practices regarding the assessment of reflective measurement models are clearly above standard but still warrant improvement. This is particularly true regarding discriminant validity assessment, which draws on metrics that recent research has debunked as ineffective in a PLS-SEM context. Similarly, the structural model assessment practices compare well with those in other disciplines but should consider more recent metrics that allow for assessing a model's out-of-sample predictive power. However, other aspects, such as formative measurement model assessment, clearly require improvement. Hospitality researchers disregard fundamental validation steps such as convergent validity and multicollinearity assessment.

While these studies indicate an increase in the application of PLS-SEM in tourism and hospitality research over last few years, it is noteworthy that PLS-SEM is clearly under-utilized as compared to the extensively used covariance-based SEM in these disciplines. Apart from providing insights into reporting practices, these review papers (Ali et al., 2018; do Valle & Assaker, 2016) also indicated that tourism and hospitality researchers seem unaware of recent advances/complimentary analysis techniques in the field. These advances and techniques clearly extend the scope of the analyses and help researchers gain more insights from the model and the data. Extensions include, but are not limited to, the weighted PLS algorithm, consistent PLS, methods for uncovering unobserved heterogeneity and impact-performance map analyses. Hence, we are editing this

handbook to provide tourism and hospitality researchers with the foundations when adopting the PLS-SEM method in their research.

This handbook on the “*Applying Partial Least Squares in Tourism and Hospitality Research*” includes 10 chapters, representing a comprehensive application of the current, original and the most advanced research in the domain of PLS methods with specific reference to their use in tourism and hospitality research. While most of the chapters comprise a thorough discussion of applications to problems from tourism and hospitality research, others focus on some key aspects of PLS analysis with a didactic approach. This handbook serves as both an introduction for those without prior knowledge of PLS and as a comprehensive reference for researchers and practitioners interested in the most recent advances in PLS methodology.

The use of PLS-SEM in tourism and hospitality research is on the rise, a trend that is in line with what has been taking place in many other fields where advanced multivariate statistical methods are employed. One of the most fundamental issues in PLS-SEM is that of minimum sample size estimation, where the “10-times rule” has been a favorite due to its simplicity of application, even though it tends to yield grossly imprecise estimates. In Chapter 1, Ned Kock discuss two related methods, based on mathematical equations, as alternatives for minimum sample size estimation in PLS-SEM: the inverse square root method and the gamma-exponential method. The application of the methods is illustrated based on a model derived from a tourism and hospitality research study. Both the methods are implemented in one of the leading PLS-SEM software tools, WarpPLS, starting in version 6.0.

There are five types of research that can be distinguished in the context of PLS-PM: (1) confirmatory, (2) explanatory, (3) predictive, (4) descriptive, and (5) exploratory. Each research type needs to be considered to select the appropriate assessment criteria. Chapter 2, by Jörg Henseler, Tobias Müller, and Florian Schuberth, sheds some light to these five research types and explains the differences by presenting empirical examples from the literature in hospitality, travel, and tourism (HTT) research. This chapter introduces new guidelines and enhancements for the use of PLS-PM in causal HTT research to assess overall model fit by using consistent PLS (PLSc) in combination with the bootstrap-based test, to measure discriminant validity with the heterotrait-monotrait ratio of correlations and assess the reliability of reflectively measured constructs via Dijkstra and Henseler’s ρ_A .

Apart from the theoretic explanations offered by the empirical models in the research papers, practitioners are also interested in the practical implications that they can apply to future cases. Being able to provide predictive diagnoses is an increasingly important issue linking theory and practice, and empirical researchers in tourism and hospitality should heed the call for predictive evaluations of their theoretical models. Fortunately, PLS path models are uniquely suited to predictive analytics. Chapter 3, by Nicholas P. Danks and Soumya Ray, offers a review of the emerging predictive methodology for PLS path models and a practical guide to what researchers can do to diagnose the predictive qualities of their models. These discussions are followed by a demonstration on a well-regarded model and dataset from the tourism literature.

Chapter 4 is contributed by Hengky Latan. It aims to update the field of knowledge regarding recent advances in PLS path modeling. This chapter uses eight assessment criteria that have been adapted in accordance with recent advances in PLS-PM. Specifically, this chapter explores all recent advances in the application of each PLS-PM technique. This chapter highlights serious misconceptions surrounding the use of PLS-PM in many disciplines, including hospitality and tourism research. This chapter also contributes to the improved practices and application of PLS-PM by proposing a new framework for reporting the results of PLS-PM.

Chapter 5 is contributed by Minwoo Lee, Kawon Kim, Kyung Young Lee, and Jung Hwa Hong. It is an application of PLS-SEM to identify smart-computing functions of smartphone's use at the workplace in the hospitality industry and examine the impact of using smart-computing functions on Mintzberg's managerial role performance and overall performance improvement. This chapter presents how both reflectively measured constructs and formatively measured constructs can be tested by using PLS-SEM.

Chapter 6, contributed by Mara Mataveli and Alfonso J. Gil, is an application of PLS-SEM to examine the impact of motivations on rural tourism on loyalty. In addition, this chapter also uses and reports moderating as well as mediation analysis.

Chapter 7, contributed by Palwasha Bibi, Ashfaq Ahmad, and Abdul H. A. Majid is also an application of PLS-SEM to measure the relationships between compensation, training and development, performance appraisal and employee retention, and the moderating role of work environment on the relationships between compensation, training and development, performance appraisal, and employee retention.

Chapter 8, contributed by Jesús García-Madariaga, Nuria Recuero Virto, María Francisca Blasco López, and Joaquin Aldas-Manzano, aims to identify how features of museum websites explain visitors' intentions to visit the museum as well as revisit intentions to the website. This chapter applies multigroup analysis (MGA) to assess visitors' intentions across the websites of the two most visited museums of Spain: Prado Museum and Reina Sofia Museum.

Chapter 9, contributed by Carlos Alberto Alves, Claudio José Stefanini, and Leonardo Aureliano da Silva, applies PLS-SEM and MGA to investigate if the presence or absence of an environmental conscious can change the relationship between environmental practices, environmental image, and attachment, and their effects on customer loyalty in restaurants based on the theory of reasoned action.

Chapter 10, contributed by Maja Šerić and Đurdana Ozretić-Došen, examines whether consumers' perceptions of online and offline communication consistency can increase their perceived service quality and brand loyalty in hospitality by applying PLS-SEM and MGA.

Even though the discussion on PLS method is increasing, its application in tourism and hospitality is under-whelming. Consequently, editors for this handbook selected high-quality papers for publication where some of them advance and explain the recent advances of PLS-SEM and others report application of

the method. The handbook provides a forum for topical issues that demonstrate PLS path modeling's usefulness in tourism and hospitality applications. A description of the method, its empirical applications, and potential methodological advancements, which increase its usefulness for research and practice, are specifically emphasized. The editors believe that this handbook will be the starting point for a more intensive use of PLS-SEM in the tourism and hospitality discipline and for additional advances that will exploit PLS's capabilities in this area. The editors and authors gratefully acknowledge Christian M. Ringle and Marko Sarstedt's comments, encouraging support, and suggestions during the preparation of this handbook. The reviewers also deserve the heartfelt recognition of the editors for their remarkable contribution to the quality of this handbook. As usual, they were diligent, meticulous, constructive, and extremely competent. The editors specifically express their gratitude to the following reviewers: Babak Taheri (Heriot-Watt University), Christian M. Ringle (Hamburg University of Technology), Gabriel Cepeda-Carrión (Universidad de Sevilla), Hengky Latan (STIE Bank BPD Jateng), José L. Roldán (Universidad de Sevilla), Jun-Hwa Cheah (Universiti Teknologi Malaysia), Marko Sarstedt (Otto von Guericke Universität Magdeburg), Murad Ali (King Abdulaziz University), and Rob Hallak (University of South Australia).

References

- Ali, F., Rasoolimanesh, S. M., Sarstedt, M., Ringle, C. M., & Ryu, K. (2018). An assessment of the use of partial least squares structural equation modeling (PLS-SEM) in hospitality research. *International Journal of Contemporary Hospitality Management*, 30(1), 514–538.
- do Valle, P. O., & Assaker, G. (2016). Using partial least squares structural equation modeling in tourism research: A review of past research and recommendations for future applications. *Journal of Travel Research*, 55(6), 695–708.

Foreword

Since its introduction by Herman O. A. Wold (1982) and Jan-Bernd Lohmöller (1989), partial least squares structural equation modeling (PLS-SEM) has undergone a broad adoption and numerous advances. The increasing dissemination of PLS-SEM in applied business is rooted in Wynne W. Chin's (1995, 1998) introductory articles and the availability of several software packages for PLS-SEM such as PLS-Graph, matrixpls, SmartPLS*, and WarpPLS, whereby a recent software review considers SmartPLS (Ringle et al., 2015) being "the most comprehensive" one (Kumar and Purani, 2018). Today, several textbooks (Garson, 2016; Hair, Hult, Ringle, & Sarstedt, 2017; Hair, Sarstedt, Ringle, & Gudergan, 2018; Ramayah, Cheah, Chuah, Ting, & Memon, 2016) and handbook articles (Esposito Vinzi, Chin, Henseler, & Wang, 2010; Henseler, Ringle, & Sarstedt, 2012; Rigdon, 2013; Sarstedt, Ringle, & Hair, 2017) provide researchers with the foundations when adopting the PLS-SEM method in their research. Numerous review studies on the use of PLS-SEM in various business research disciplines such as – accounting (Lee, Petter, Fayard, & Robinson, 2011; Nitzl, 2016), family business (Sarstedt, Ringle, Smith, Reams, & Hair, 2014), group and organization management (Sosik, Kahai, & Piovoso, 2009), hospitality management (Ali, Rasoolimanesh, Sarstedt, Ringle, & Ryu, 2018), human resource management (Ringle, Sarstedt, Mitchell, & Gudergan, 2018), information systems (Hair, Hollingsworth, Randolph, & Chong, 2017; Ringle, Sarstedt, & Straub, 2012), international marketing research (Henseler, Ringle, & Sinkovics, 2009; Richter, Sinkovics, Ringle, & Schlägel, 2016), marketing (Hair, Sarstedt, Ringle, & Mena, 2012), operations management (Peng & Lai, 2012), psychology (Willaby, Costa, Burns, MacCann, & Roberts, 2015), strategic management (Hair, Sarstedt, Pieper, & Ringle, 2012), supply chain management (Kaufmann & Gaeckler, 2015), and tourism (do Valle & Assaker, 2016) – not only substantiate the wide adoption of the method, but also provide an overview how researchers used PLS-SEM in their studies.

Accompanying the rapid pace of development, PLS-SEM has also witnessed controversies, with researchers sometimes even questioning the method's *raison d'être* (Rönkkö, Antonakis, McIntosh, & Edwards, 2016; Rönkkö & Evermann, 2013; Rönkkö, McIntosh, & Antonakis, 2015). However, most of the criticism has been refuted as inaccurate (Henseler et al., 2014) or grounded in different measurement philosophies (Rigdon, Sarstedt, & Ringle, 2017; Sarstedt, Hair, Ringle, Thiele, & Gudergan, 2016). These criticisms, however, helped furthering

* Christian Ringle acknowledges a financial interest in SmartPLS.

the method's theory base in terms of measurement and model estimation, triggering a wide range of follow-up research. New developments in PLS-SEM range from new estimators (e.g., Dijkstra & Henseler, 2015; Dolce, Esposito Vinzi, & Lauro, 2018; Schuberth & Cantaluppi, 2017) and model evaluation metrics (e.g., Aguirre-Urreta & Rönkkö, 2018; Franke & Sarstedt, in press; Henseler, Ringle, & Sarstedt, 2015; Sharma, Sarstedt, Shmueli, Thiele, & Kim, 2017; Shmueli, Ray, Velasquez Estrada, & Chatla, 2016) to complementary methods such as methods for uncovering unobserved heterogeneity (e.g., Ringle, Sarstedt, & Schlittgen, 2014; Schlittgen, Ringle, Sarstedt, & Becker, 2016), different multigroup analysis approaches (Matthews, 2018), testing measurement invariance of composites (Henseler, Ringle, & Sarstedt, 2016), and endogeneity assessment (Hult, Hair, Proksch, Sarstedt, Pinkwart, & Ringle, 2018). These advances have greatly extended researchers' methodological toolbox (Khan et al., 2018) and fueled the adoption of PLS-SEM in the social sciences and other fields.

This handbook by Faizan Ali, S. Mostafa Rasoolimanesh, and Cihan Cobanoglu on PLS-SEM application in tourism and hospitality research represents another important contribution to progress on the method. We would like to thank Faizan Ali, S. Mostafa Rasoolimanesh, and Cihan Cobanoglu for the effort of developing this important handbook. Congratulations to a job done well!

Christian M. Ringle

Hamburg University of Technology (TUHH), Hamburg, Germany,
and the University of Waikato, Hamilton, New Zealand

Marko Sarstedt

Otto-von-Guericke University Magdeburg, Magdeburg,
Germany, and Monash University Malaysia

References

- Aguirre-Urreta, M. I., & Rönkkö, M. (2018). Statistical inference with PLS using bootstrap confidence intervals. *MIS Quarterly*, forthcoming.
- Ali, F., Rasoolimanesh, S. M., Sarstedt, M., Ringle, C. M., & Ryu, K. (2018). An assessment of the use of partial least squares structural equation modeling (PLS-SEM) in hospitality research. *International Journal of Contemporary Hospitality Management*, 30(1), 514–538.
- Chin, W. W. (1995). Partial least squares is to LISREL as principal components analysis is to common factor analysis. *Technology Studies*, 2(2), 315–319.
- Chin, W. W. (1998). The partial least squares approach to structural equation modeling. In G. A. Marcoulides (Ed.), *Modern methods for business research* (pp. 295–358). Mahwah, NJ: Erlbaum.
- Chin, W. W. (2003). *PLS-Graph 3.0*. Houston: Soft Modeling.
- Dijkstra, T. K., & Henseler, J. (2015). Consistent partial least squares path modeling. *MIS Quarterly*, 39(2), 297–316.
- do Valle, P. O., & Assaker, G. (2016). Using partial least squares structural equation modeling in tourism research: A review of past research and recommendations for future applications. *Journal of Travel Research*, 55(6), 695–708.
- Dolce, P., Esposito Vinzi, V., & Lauro, N. C. (2018). Non-symmetrical composite-based path modeling. *Advances in data analysis and classification*, forthcoming.

- Esposito Vinzi, V., Chin, W. W., Henseler, J., & Wang, H. (Eds.), (2010). *Handbook of partial least squares: Concepts, methods and applications (Springer handbooks of computational statistics series, vol. II)*. New York, NY: Springer.
- Franke, G. R., & Sarstedt, M. (in press). Heuristics versus statistics in discriminant validity testing: A comparison of four procedures. *Internet Research*, forthcoming.
- Garson, G. D. (2016). *Partial least squares regression and structural equation models*. Asheboro, NC: Statistical Associates.
- Hair, J. F., Hollingsworth, C. L., Randolph, A. B., & Chong, A. Y. L. (2017). An updated and expanded assessment of PLS-SEM in information systems research. *Industrial Management & Data Systems*, 117(3), 442–458.
- Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). *A primer on partial least squares structural equation modeling (PLS-SEM)* (2nd ed.). Thousand Oaks, CA: Sage.
- Hair, J. F., Sarstedt, M., Pieper, T. M., & Ringle, C. M. (2012). The use of partial least squares structural equation modeling in strategic management research: A review of past practices and recommendations for future applications. *Long Range Planning*, 45(5–6), 320–340.
- Hair, J. F., Sarstedt, M., Ringle, C. M., & Gudergan, S. P. (2018). *Advanced issues in partial least squares structural equation modeling (PLS-SEM)*. Thousand Oaks, CA: Sage.
- 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.
- Henseler, J., Dijkstra, T. K. (2015). *ADANCO 2.0*. Kleve: Composite Modeling, <http://www.compositemodeling.com>.
- Henseler, J., Dijkstra, T. K., Sarstedt, M., Ringle, C. M., Diamantopoulos, A., Straub, D. W., Ketchen, D. J., Jr., Hair, J. F., Hult, G. T. M., Calantone, R. J. (2014). Common beliefs and reality about partial least squares: Comments on Rönkkö & Evermann (2013). *Organizational Research Methods*, 17(2), 182–209.
- Henseler, J., Ringle, C. M., & Sarstedt, M. (2012). Using partial least squares path modeling in international advertising research: Basic concepts and recent issues. In S. Okazaki (Ed.), *Handbook of research in international advertising* (pp. 252–276). Cheltenham: Edward Elgar Publishing.
- Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115–135.
- Henseler, J., Ringle, C. M., & Sarstedt, M. (2016). Testing measurement invariance of composites using partial least squares. *International Marketing Review*, 33(3), 405–431.
- Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing. In R. R. Sinkovics & P. N. Ghauri (Eds.), *Advances in international marketing* (Vol. 20, pp. 277–320). Bingley: Emerald Publishing.
- Hult, G. T. M., Hair, J. F., Proksch, D., Sarstedt, M., Pinkwart, A., & Ringle, C. M. (2018). Addressing endogeneity in international marketing applications of partial least squares structural equation modeling. *Journal of International Marketing*, forthcoming.
- Kaufmann, L., & Gaeckler, J. (2015). A structured review of partial least squares in supply chain management research. *Journal of Purchasing and Supply Management*, 21(4), 259–272.
- Khan, G. F., Sarstedt, M., Shiau, W.-L., Hair, J. F., Ringle, C. M., & Fritze, M. (2018). Methodological research on partial least squares structural equation modeling (PLS-SEM): An analysis based on social network approaches. *Internet Research*, forthcoming.

- Kumar, D. S., & Purani, K. (2018). Model specification issues in PLS-SEM: Illustrating linear and non-linear models in hospitality services context. *Journal of Hospitality and Tourism Technology*.
- Lee, L., Petter, S., Fayard, D., & Robinson, S. (2011). On the use of partial least squares path modeling in accounting research. *International Journal of Accounting Information Systems*, 12(4), 305–328.
- Lohmöller, J.-B. (1989). *Latent variable path modeling with partial least squares*. Heidelberg, Germany: Physica.
- Matthews, L. (2018). Applying multi-group analysis in PLS-SEM: A step-by-step process. In H. Latan & R. Noonan (Eds.), *Partial least squares structural equation modeling: Basic concepts, methodological issues and applications* (pp. 219–243). Heidelberg, Germany: Springer.
- Nitzl, C. (2016). The use of partial least squares structural equation modelling (PLS-SEM) in management accounting research: Directions for future theory development. *Journal of Accounting Literature*, 37(December), 19–35.
- Peng, D. X., & Lai, F. (2012). Using partial least squares in operations management research: A practical guideline and summary of past research. *Journal of Operations Management*, 30(6), 467–480.
- Ramayah, T., Cheah, J.-H., Chuah, F., Ting, H., & Memon, M. A. (2016). *Partial least squares structural equation modeling (PLS-SEM) using SmartPLS 3.0: An updated and practical guide to statistical analysis*. Singapore: Pearson.
- Richter, N. F., Sinkovics, R. R., Ringle, C. M., & Schlägel, C. (2016). A critical look at the use of SEM in international business research. *International Marketing Review*, 33(3), 376–404.
- Rigdon, E. E. (2013). Partial least squares path modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling. A second course* (2nd ed., pp. 81–116). Charlotte, NC: Information Age Publishing.
- Rigdon, E. E., Sarstedt, M., & Ringle, C. M. (2017). On comparing results from CB-SEM and PLS-SEM. Five perspectives and five recommendations. *Marketing ZFP*, 39(3), 4–16.
- Ringle, C. M., Sarstedt, M., Mitchell, R., & Gudergan, S. P. (2018). Partial least squares structural equation modeling in HRM research. *The International Journal of Human Resource Management*, forthcoming.
- Ringle, C. M., Sarstedt, M., & Schlittgen, R. (2014). Genetic algorithm segmentation in partial least squares structural equation modeling. *OR Spectrum*, 36(1), 251–276.
- Ringle, C. M., Sarstedt, M., & Straub, D. W. (2012). A critical look at the use of PLS-SEM in MIS quarterly. *MIS Quarterly*, 36(1), iii–xiv.
- Ringle, C. M., Wende, S., & Becker, J.-M. (2015). *SmartPLS 3*. Bönningstedt: SmartPLS. Retrieved from <http://www.smartpls.com>.
- Ringle, C. M., Wende, S., & Will, A. (2005). *SmartPLS 2*. Hamburg: SmartPLS. Retrieved from www.smartpls.de.
- Rönkkö, M., Antonakis, J., McIntosh, C. N., & Edwards, J. R. (2016). Partial least squares path modeling: Time for some serious second thoughts. *Journal of Operations Management*, 47–48(November), 9–27.
- Rönkkö, M., & Evermann, J. (2013). A critical examination of common beliefs about partial least squares path modeling. *Organizational Research Methods*, 16(3), 425–448.
- Rönkkö, M., McIntosh, C. N., & Antonakis, J. (2015). On the adoption of partial least squares in psychological research: Caveat emptor. *Personality and Individual Differences*, 87, 76–84.

- Sarstedt, M., Hair, J. F., Ringle, C. M., Thiele, K. O., & Gudergan, S. P. (2016). Estimation issues with PLS and CBSEM: Where the bias lies! *Journal of Business Research*, 69(10), 3998–4010.
- Sarstedt, M., Ringle, C. M., & Hair, J. F. (2017). Partial least squares structural equation modeling. In C. Homburg, M. Klarmann, & A. Vomberg (Eds.), *Handbook of market research*. Heidelberg, Germany: Springer.
- Sarstedt, M., Ringle, C. M., Smith, D., Reams, R., & Hair, J. F. (2014). Partial least squares structural equation modeling (PLS-SEM): A useful tool for family business researchers. *Journal of Family Business Strategy*, 5(1), 105–115.
- Schlittgen, R., Ringle, C. M., Sarstedt, M., & Becker, J.-M. (2016). Segmentation of PLS path models by iterative reweighted regressions. *Journal of Business Research*, 69(10), 4583–4592.
- Schuberth, F., & Cantaluppi, G. (2017). Ordinal consistent partial least squares. In H. Latan & R. Noonan (Eds.), *Partial least squares path modeling: Basic concepts, methodological issues and applications* (pp. 109–150). Cham, Switzerland: Springer International Publishing.
- Sharma, P. N., Sarstedt, M., Shmueli, G., Thiele, K. O., & Kim, K. H. (2017). *Model selection in MIS research using PLS-SEM*. Working Paper.
- Shmueli, G., Ray, S., Velasquez Estrada, J. M., & Chatla, S. B. (2016). The elephant in the room: Evaluating the predictive performance of PLS models. *Journal of Business Research*, 69(10), 4552–4564.
- Sosik, J. J., Kahai, S. S., & Piovoso, M. J. (2009). Silver bullet or voodoo statistics? A primer for using the partial least squares data analytic technique in group and organization research. *Group Organization Management*, 34(1), 5–36.
- Willaby, H. W., Costa, D. S. J., Burns, B. D., MacCann, C., & Roberts, R. D. (2015). Testing complex models with small sample sizes: A historical overview and empirical demonstration of what partial least squares (PLS) can offer differential psychology. *Personality and Individual Differences*, 84, 73–78.
- Wold, H. O. A. (1982). Soft modeling: The basic design and some extensions. In K. G. Jöreskog & H. O. A. Wold (Eds.), *Systems under indirect observations: Part II* (pp. 1–54). Amsterdam, The Netherlands: North-Holland.

This page intentionally left blank

About the Authors

Ashfaq Ahmad is an Assistant Professor at College of Business, Universiti Utara Malaysia, Kedah, Malaysia.

Carlos Alberto Alves is a Professor of the Graduate Program in Hospitality at Universidade Anhembi Morumbi, São Paulo, Brazil.

Joaquin Aldas-Manzano is a Professor of Marketing in the Department of Marketing, Faculty of Economics, Universitat de València, Researcher of Valencia Economics Research Institute (Ivie), and Visiting Research Fellow of the University of Glasgow, Scotland, UK.

Leonardo Aureliano da Silva is a Professor of the Graduate Program in Food and Beverage Management at Universidade Anhembi Morumbi and a Professor at Escola Superior de Propaganda e Marketing – ESPM.

Palwasha Bibi is an Assistant Professor at College of Business, Universiti Utara Malaysia, Kedah, Malaysia.

Nicholas P. Danks is a PhD Candidate at the Institute of Service Science in National Tsing Hua University, Taiwan.

Jesús García-Madariaga is a Coordinator of Marketing Discipline at Complutense University. He is currently an Associate Professor in Economics and Business Administration, Director of the University Research team “Markco2,” and Visiting Research Fellow of the California Polytechnic University, San Luis Obispo, CA, USA.

Alfonso J. Gil is an Assistant Professor in Business Administration at the University of La Rioja, La Rioja. He is a Tutor Professor at the National Distance Education University UNED, Madrid, Spain, and a Visiting Professor at the State University of the West of Paraná, Brazil.

Jörg Henseler is the Chair of Product-Market Relations and heads the Department of Design, Production & Management at the Faculty of Engineering Technology of the University of Twente, Enschede, The Netherlands. Moreover, he is visiting professor at Nova Information Management School, Universidade Nova de Lisboa, Portugal.

Jung Hwa Hong is an Associate Professor in Hotel Management and Dean of Division of Tourism in the University of Dongseo in South Korea.

Claudio José Stefanini is a Professor of the Graduate Program in Hospitality at Universidade Anhembi Morumbi, São Paulo, Brazil.

Kawon Kim is an Assistant Professor at the School of Hotel, Restaurant, and Tourism Management at the University of South Carolina.

Ned Kock is a Killam Distinguished Professor and Chair of the Division of International Business and Technology Studies, A.R. Sanchez, Jr. School of Business, at Texas A&M International University.

Hengky Latan is a Senior Researcher in the area of accounting. He has worked at the STIE Bank BPD Jateng in Department of Accounting, and at the HLC Consulting, Indonesia.

Kyung Young Lee is an Associate Professor in the Rowe School of Business at Dalhousie University, Canada.

Minwoo Lee is an Assistant Professor at Conrad N. Hilton College of Hotel and Restaurant Management, University of Houston, USA.

Maria Francisca Blasco López is a Senior Lecturer at the Universidad Complutense de Madrid. She is currently the Vice Dean and Director of the MBA in Marketing Management and the master's degree program in Neuromarketing and Consumer Behaviour at the Universidad Complutense de Madrid.

Abdul H. A. Majid is an Associate Professor at College of Business, Universiti Utara Malaysia, Kedah, Malaysia.

Mara Mataveli is a Researcher in Economics and Business Administration at the University of La Rioja, Spain.

Tobias Müller is an external PhD candidate at the Chair of Product-Market Relations of the University of Twente, Enschede, The Netherlands.

Durdana Ozretić-Došen is a Professor at the Marketing Department of the Faculty of Economics & Business, University of Zagreb, Croatia.

Soumya Ray is an Associate Professor at the Institute of Service Science in National Tsing Hua University, Taiwan.

Florian Schuberth is an Assistant Professor at the Chair of Product-Market Relations of the University of Twente, Enschede, The Netherlands.

Maja Šerić is an Assistant Professor at the Department of Marketing of the Faculty of Economics, University of Valencia, Spain.

Nuria Recuero Virto is a Post-doctoral Researcher at Universidad Complutense de Madrid, with a special grant of Santander Bank.

Chapter 1

Minimum Sample Size Estimation in PLS-SEM: An Application in Tourism and Hospitality Research

Ned Kock

Introduction

The field of tourism and hospitality research is closely linked with the innovative implementation and use (Rasoolimanesh, Dahalan, & Jaafar, 2016; Rasoolimanesh, Jaafar, Kock, & Ahmad, 2017) of the partial least squares (PLS) technique (Kock, 2010; Kock & Hadaya, 2018). This technique has been extensively used in a variety of fields within the broader scholarly realm of business and social sciences research, a practice that has extended to many other fields over the years, to analyze path models with variables that are measured indirectly through other variables. These indirectly measured variables are generally known as latent variables (Kline, 1998; Kock & Lynn, 2012). The approach to analyzing path models with latent variables is broadly known as structural equation modeling (SEM). Thus, the acronym “PLS-SEM” is used here to refer to SEM employing PLS.

One of the most fundamental issues in PLS-SEM is that of minimum sample size estimation. A widely used minimum sample size estimation method in PLS-SEM is the “10-times rule” method (Hair, Ringle, & Sarstedt, 2011), which builds on the assumption that the sample size employed in an empirical study should be greater than 10 times the maximum number of inner or outer model links pointing at any latent variable in the model. While this method’s simplicity of application makes it a favorite among PLS-SEM users, it has been shown in the past to lead to grossly inaccurate estimates (Goodhue, Lewis, & Thompson, 2012; Kock & Hadaya, 2018).

We propose two related methods, based on mathematical equations, for minimum sample size estimation in PLS-SEM. The first method is called the inverse square root method because it uses the inverse square root of a sample’s size for

standard error estimation – an important step in minimum sample size estimation. The second method is called the gamma-exponential method because it relies on gamma and exponential smoothing function corrections applied to the first method.

The application of the methods is illustrated based on a model derived from a tourism and hospitality research study. Both methods are implemented in one of the leading PLS-SEM software tools, WarpPLS, starting in version 6.0 (Kock, 2017). Informed by various Monte Carlo experiments (Kock & Hadaya, 2018), we show that both methods are fairly accurate, with the inverse square root method being also particularly attractive in terms of its simplicity of application.

Illustrative Study

We use a study in the field of tourism and hospitality research as a basis for the development of a model to illustrate our discussion of minimum sample size estimation in PLS-SEM. Fig. 1 summarizes the results in terms of path coefficients (e.g., $\beta = 0.121$) and R^2 coefficients (e.g., $R^2 = 0.296$). The study used as a basis for the model was authored by Rasoolimanesh, Jaafar, Kock, and Ahmad, and published in the *Journal of Sustainable Tourism* in 2017 (Rasoolimanesh et al., 2017). A solid theoretical development underlies the study. It was published in an outlet, namely the *Journal of Sustainable Tourism*, which has long been considered a very selective and elite academic journal in the field of tourism and hospitality research. To simplify our discussion, our model is a modified version of the model developed and tested by Rasoolimanesh et al. (2017).

The latent variables shown as ovals are assumed to be measured reflectively through multiple indicators, primarily on Likert-type scales with multiple points (e.g., 5 points). Community involvement refers to the degree of local resident involvement in the management of a World Heritage Site (WHS). A WHS is typically a geographical area or landmark that has been formally deemed by the United Nations Educational, Scientific and Cultural Organization (UNESCO) as having major cultural, historical, or scientific significance.

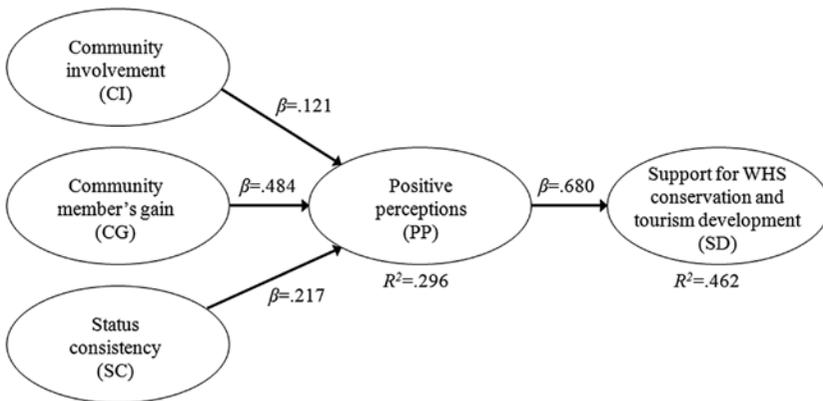


Fig. 1: Illustrative Model.

Community member's gain refers to perceived gains (e.g., in household income and quality of life) to the resident's friends and relatives. Status consistency refers to the resident's satisfaction and sense of belonging to the community where the WHS is located. Positive perceptions is a variable that refers to positive perceptions about the community to which the WHS belongs as arising from the WHS status granted by the UNESCO. Support for WHS conservation and tourism development refers to degree to which a resident perceives involvement in the management of a WHS as being important.

Power, Effect Size, and Minimum Sample Size

Statistical *power* (Cohen, 1988, 1992; Goodhue et al., 2012; Kock, 2016; Muthén & Muthén, 2002), often referred to simply as “power,” is a statistical test's probability of avoiding type II errors, or false negatives. Power is often estimated for a particular coefficient of association and sample size, for samples drawn from a population, at a given significance level (usually, $p < 0.05$). For example, let us assume that a PLS-SEM test is applied to a path in a model where the corresponding coefficient is associated with a “real” effect at the population level of magnitude 0.2; which would be referred to as the “true” path coefficient. Let us also assume that the test correctly recognizes the path coefficient as significant 83% of the time when samples of size 150 are randomly taken from the population. Under these circumstances, we would conclude that the power of the test is 83%, or 0.83.

The *effect size* (Cohen, 1988, 1992; Kock, 2014b) is a measure of the magnitude of an effect that is independent of the size of the sample analyzed. Two main measures of effect size are commonly used in PLS-SEM. One is Cohen's f^2 coefficient (Cohen, 1988, 1992), which is calculated as $\Delta R^2 / (1 - R^2)$, where ΔR^2 is the incremental contribution of a predictor latent variable to the R^2 of the criterion latent variable to which it points. The other measure of effect size commonly used in PLS-SEM is the absolute contribution of the predictor latent variable (Kock, 2014b; Mandal, Mukhopadhyay, Bagchi, & Gunasekaran, 2012), namely the numerator ΔR^2 of Cohen's f^2 equation, without the denominator correction. This second measure tends to yield lower results, thus being a more conservative effect size estimate. By convention, effect sizes of 0.02, 0.15, and 0.35 are, respectively, termed small, medium, and large (Cohen, 1992; Kock, 2014b).

The *minimum sample size* at which a PLS-SEM test achieves an acceptable level of power (usually 0.8) depends on the effect size associated with the path coefficient under consideration (Cohen, 1988, 1992; Goodhue et al., 2012; Kock, 2014b). The higher is the magnitude of a path coefficient at the population level, the higher is usually its effect size, and the greater is the probability that a true effect will be properly detected with a small sample. Therefore, strong path coefficients at the population level, whether they are negative or positive, tend to require very small sample sizes for their proper identification. So, if a researcher knows that all of the path coefficients of a model will be strong prior to collecting empirical data, leading to large effect sizes, the researcher may consider using a small sample size in a PLS-SEM analysis. As we will see later, we can use the

notion of effect size for a general minimum sample size recommendation that does not rely on predictions about path strength.

More often than not, PLS-SEM is presented as being a desirable multivariate data analysis method due to its remarkable ability to achieve acceptable power at very small sample sizes (Hair et al., 2011; Hair, Hult, Ringle, & Sartedt, 2014). While this may be true for models containing only strong path coefficients and large effect sizes, it is not true for models with path coefficients of more modest magnitudes, and certainly not true for models with fairly weak path coefficients. (At points in our discussion, we deviate somewhat from strict technical statistical jargon, for simplicity. For example, in the previous sentence we refer to “weak” path coefficients, meaning positive or negative path coefficients whose absolute values are low.) As demonstrated by Kock and Hadaya (2018), PLS-SEM’s power is consistent with what one would expect from ordinary least squares regression, as well as other methods with similar mathematical underpinnings.

The Monte Carlo Simulation Method

Employing the Monte Carlo simulation method (Kock, 2016; Paxton, Curran, Bollen, Kirby, & Chen, 2001; Robert & Casella, 2013) for minimum sample size estimation in PLS-SEM requires the researcher to choose a number of sample size points (e.g., 15, 20, 30, and 40), generate a number of samples (e.g., 1,000) for each sample size point, calculate the percentages of samples in which significant effects (e.g., for which $p < 0.05$) were found for each sample size point (the power associated with each sample size), and estimate via interpolation the minimum sample size at which power reaches the desired threshold (i.e., 0.8).

Therefore, the Monte Carlo simulation method often requires two or more related simulations. The initial simulations are experimental to define an appropriate set of sample size points. These would be followed by a final simulation, whereby one would estimate via interpolation the minimum sample size at which power reaches the desired threshold of 0.8.

The samples (e.g., 1,000) generated for each sample size point via the Monte Carlo simulation method are based on a population model defined by the researcher. The Monte Carlo simulation method is a complex way by which minimum sample sizes can be determined, and for which technical methodological expertise is required.

Minimum sample size estimation via the Monte Carlo simulation method may be a very time-consuming alternative, even for experienced methodological researchers with good computer programming skills. Nevertheless, it is a fairly precise method for minimum sample size estimation. As such, we use it here to obtain a baseline estimate of the true minimum required sample size in our illustrative model, a baseline estimate against which other methods are compared.

Classic Methods: 10-Times Rule and Minimum R^2

In this section, we discuss two methods for minimum sample size estimation in PLS-SEM that have been used in the past, which we refer to here as “classic”

methods. The first method presented here is the 10-times rule method (Goodhue et al., 2012; Hair et al., 2011), which is the most commonly used in PLS-SEM. The second method, the minimum R^2 method, has been proposed by Hair et al. (2014, p. 21) as an alternative to the 10-times rule method.

The 10-Times Rule Method

The most widely used minimum sample size estimation method in PLS-SEM is the “10-times rule” method (Hair et al., 2011; Peng & Lai, 2012). Among the variations of this method, the one usually seen is based on the rule that the sample size should be greater than 10 times the maximum number of inner or outer model links pointing at any latent variable in the model (Goodhue et al., 2012).

Minimum sample size estimation via the 10-times rule method does not depend on the magnitude of the path coefficients in the model. For example, in the model used in our illustrative study, the 10-times rule method leads to the minimum sample size estimation of 30, regardless of the strengths of the path coefficients. This is because the maximum number of model links pointing at any variable in the model is 3, which when multiplied by 10 yields 30.

As demonstrated by Kock and Hadaya (2018), this method can lead to grossly inaccurate estimations of minimum required sample size. This is confirmed in the context of our illustrative model, for which the true minimum required sample size was estimated by us to be 407 via a Monte Carlo simulation. A sample size of 30 would be less than one-tenth of the true minimum required sample size, and would in fact lead to an unacceptably low power level if used in a PLS-SEM analysis.

The Minimum R^2 Method

In their pioneering book on PLS-SEM, Hair et al. (2014, p. 21) discuss an alternative to the 10-times rule for minimum sample size estimation. We refer to this method as the “minimum R^2 method,” because the minimum R^2 in the model is prominently used for minimum sample size estimation. This method, which builds on Cohen’s (1988, 1992) power tables for least squares regression, relies on a table listing minimum required sample sizes based on three elements.

The first element of the minimum R^2 method is the maximum number of arrows pointing at a latent variable (a.k.a. construct) in a model. The second is the significance level used. The third is the minimum R^2 in the model. Table 1 is a reduced version of the table presented by Hair et al. (2014, p. 21). This reduced version focuses on the significance level of 0.05, which is the most commonly used significance level in PLS-SEM, and assumes that power is set at 0.8.

For example, in the model used in our illustrative study, the maximum number of arrows pointing at a latent variable is 3, and the minimum R^2 in the model is 0.296. There is no cell in the table for the minimum R^2 method for which these two values intersect, but the closest cell shows a minimum sample size of 59, which we use as the minimum R^2 method’s estimate. This is in fact an overestimation based on this method because it assumes a minimum R^2 in the model of 0.25, which is lower than the actual value of 0.296.

Table 1: Tabulation of the Minimum R^2 Method.

Maximum Number of Arrows Pointing at A Construct	Minimum R^2 in the Model			
	0.10	0.25	0.50	0.75
2	110	52	33	26
3	124	59	38	30
4	137	65	42	33
5	147	70	45	36
6	157	75	48	39
7	166	80	51	41
8	174	84	54	44
9	181	88	57	46
10	189	91	59	48

While this method appears to be an improvement over the 10-times rule method, as it takes as an input at least one additional element beyond the network of links in the model, [Kock and Hadaya \(2018\)](#) demonstrated that the minimum R^2 method can also lead to grossly inaccurate estimations of minimum required sample size.

This is confirmed in the context of our illustrative model. The true minimum required sample size was estimated by us to be 407 via a Monte Carlo simulation, as noted earlier. A sample size of 59 would in fact lead to a power level of well below 50% in the context of our illustrative model, which is unacceptably low and comparable to the estimate obtained via the 10-times rule method in terms of its deleterious impact on power.

New Methods: Inverse Square Root and Gamma-Exponential

In this section, we discuss two related methods, based on mathematical equations, for minimum sample size estimation in PLS-SEM. Neither method relies on Monte Carlo simulations or on elements that make up the 10 times rule or the minimum R^2 methods. The first method, called the inverse square root method, uses the inverse square root of a sample's size for standard error estimation – hence, its name. The second method, called the gamma-exponential method, relies on gamma and exponential smoothing function corrections applied to the standard error estimation employed in the first method.

The Inverse Square Root Method

Whenever one or more researchers analyze samples taken from a population using PLS-SEM, each analysis generates various path coefficients. Each path coefficient (β) will have a standard error (S) associated with it. If we plot the distribution of the ratio β/S , also indicating the location of a critical T ratio ([Kock, 2015](#);

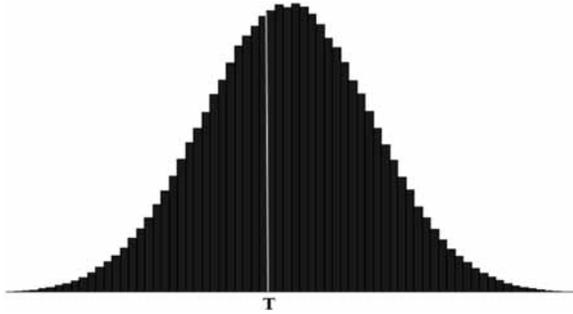


Fig. 2: Distribution of the Ratio β/S .

Note: T = Critical T ratio for a specific significance level chosen.

Weakliem, 2016) for a specific significance level chosen, we obtain a graph that has the general shape shown in Fig. 2. For each instance where the ratio β/S surpasses the critical T ratio, the effect associated with the path coefficient β will be correctly deemed as statistically significant. This assumes that the path coefficient refers to an effect that exists at the population level; that is, a “true” effect.

The magnitude of the ratio β/S increases with increases in the magnitude of the path coefficient β and decreases in the standard error S . This standard error decreases with increases in sample size, as will be seen shortly in the following. Therefore, with increases in the magnitude of the path coefficient and of the sample size analyzed, the probability that the ratio β/S will surpass the critical T ratio will increase. As a result, the likelihood that an effect that does exist at the population level will be mistakenly rejected will decrease. In other words, the power of the test will increase because the probability that false negatives will occur will decrease.

As we can see from the figure, the power of a test associated with a given path coefficient for which a sign has been hypothesized can be defined as the probability that the ratio $|\beta|/S$ will be greater than the critical T ratio for a specific significance level chosen (Cohen, 1988; Goodhue et al., 2012; Kock, 2015). Here, $|\beta|$ is the absolute value of β , as a path coefficient strength’s influence on power is exerted whether the coefficient is positive or negative. The significance level normally chosen is 0.05 (i.e., $p < 0.05$), for which the critical T ratio can be denoted as $T_{0.05}$. This can be expressed mathematically as follows:

$$W = P\left(\frac{|\beta|}{S} > T_{0.05}\right). \quad (1)$$

Statistical power is denoted as W in (1), and $P(\cdot)$ is the probability function. If we set power to be above a given level, most commonly 0.8, Eq. (1) can be expressed employing a cumulative probability function $\Phi(\cdot)$ for the standard normal distribution. Assuming that path coefficients are normally distributed, we can say that power will be greater than 0.8 when the cumulative distribution function for the standard normal distribution indicated in the following equation is greater than 0.8:

$$\Phi\left(\frac{|\beta|}{S} - T_{0.05}\right) > 0.8. \quad (2)$$

The assumption that path coefficients are normally distributed generally holds for PLS-SEM, because coefficients calculated based on sample sets taken randomly from a population tend to be distributed in conformity with the central limit theorem (Kipnis & Varadhan, 1986; Miller & Wichern, 1977).

Taking Eq. (2) as a basis, we obtain Eq. (3) in terms of the standardized score associated with the value 0.8 of the cumulative distribution function for the normal distribution ($z_{0.8}$). To obtain Eq. (3), we also take into consideration the property that $T_{0.05} = z_{0.95}$.

$$\begin{aligned} \frac{|\beta|}{S} - T_{0.05} &> z_{0.8} \rightarrow \\ \frac{|\beta|}{S} &> T_{0.05} + z_{0.8} \rightarrow \\ \frac{|\beta|}{S} &> z_{0.95} + z_{0.8}. \end{aligned} \quad (3)$$

Any given z -score z_x can be calculated based on a standard normal distribution, which is a normal distribution with a mean of 0 and a standard deviation of 1. The score is a value associated with the probability x that a random variable takes on a value that is equal to or less than z_x . In MATLAB, it is obtained using the function `norminv(x,0,1)`. In Excel, it is obtained using the function `NORMINV(x,0,1)` or the function `NORMSINV(x)`.

An estimate \hat{S} of the true standard error (S) can be produced using Eq. (4). This estimate lends the name to the method presented here, the inverse square root method, and is known to be biased (Gurland & Tripathi, 1971; Kock, 2014a, 2018), consistently underestimating the corresponding true value at very small samples (i.e., $1 < N \leq 10$), and consistently overestimating it at greater sample sizes (i.e., $N > 10$). Shortly, we will discuss two approaches to correct this bias, which are combined in our second proposed minimum sample size estimation method, the gamma-exponential method.

$$\hat{S} = \frac{1}{\sqrt{N}}. \quad (4)$$

Employing the Excel function `NORMSINV(x)`, we obtain the values for $z_{0.95}$ and $z_{0.8}$, or `NORMSINV(0.95)` and `NORMSINV(0.8)`, which are, respectively, 1.645 and 0.842. The sum $z_{0.95} + z_{0.8}$ is thus 2.486. Combining Eqs. (3) and (4), with $|\beta|_{\min}$ replacing $|\beta|$ and representing the absolute value of the statistically significant path coefficient with the minimum magnitude in the model, we then have:

$$\begin{aligned}
 |\beta|_{\min} \sqrt{\hat{N}} &> z_{0.95} + z_{0.8} \rightarrow \\
 \hat{N} &> \left(\frac{z_{0.95} + z_{0.8}}{|\beta|_{\min}} \right)^2 \rightarrow \\
 \hat{N} &> \left(\frac{2.486}{|\beta|_{\min}} \right)^2.
 \end{aligned}
 \tag{5}$$

Based on our proposed inverse square root method, the minimum sample size is estimated as the smallest positive integer that satisfies Eq. (5). As such, it can be calculated by rounding the result of the calculation of the right side of the equation to the next integer. In MATLAB, it can be obtained using the function `ceil((2.486/bmin)2)`, where *b_{min}* is a variable that stores the value of $|\beta|_{\min}$. In Excel, it can be obtained using the function `ROUNDUP((2.486/bmin)2,0)`, where *b_{min}* is the name of a cell that stores the value of $|\beta|_{\min}$.

The Gamma-Exponential Method

As we noted earlier, our estimate \hat{S} of the true standard error (*S*), obtained through the formula $1/\sqrt{N}$, is known to be biased. A classic gamma function correction of the bias for very small sample sizes (i.e., $1 < N \leq 10$) was proposed by [Gurland and Tripathi \(1971\)](#):

$$\hat{S} = \frac{1}{c\sqrt{N}},$$

where

$$c = \sqrt{\frac{N-1}{2}} \frac{\Gamma\left(\frac{N-1}{2}\right)}{\Gamma\left(\frac{N}{2}\right)} \text{ and } \Gamma(\cdot) \text{ is the gamma function.}$$

With the gamma function correction proposed by [Gurland and Tripathi \(1971\)](#), the resultant Eq. (6) to obtain the minimum required sample size \hat{N} becomes more complex. This equation can be solved by means of a computer program that starts with $\hat{N} = 1$ and progressive increments the value of \hat{N} to 2, 3, etc., until the smallest positive integer that satisfies the equation is obtained. In MATLAB, the value of $\Gamma(x)$ is obtained using the function `gamma(x)`. In Excel, it is obtained using the two-function formula `EXP(GAMMALN(x+1))`.

$$|\beta|_{\min \sqrt{\hat{N}}} \frac{\Gamma\left(\frac{\hat{N}-1}{2}\right)}{\Gamma\left(\frac{\hat{N}}{2}\right)} > 2.486.
 \tag{6}$$

The gamma function correction equation has no effect, in terms of minimum required sample size estimation, for $N > 10$. The reason for this is that the correction coefficient c quickly converges to 1 for $N > 10$. An exponential smoothing function correction of the standard error bias was proposed and validated by Kock (2014a, 2018) in the context of PLS-SEM for sample sizes greater than those covered by the gamma function correction (i.e., $N > 10$):

$$\hat{S} = \frac{1}{\sqrt{N}} e^{-\left(\frac{c|\beta|}{\sqrt{N}}\right)}.$$

With this exponential smoothing function correction, Eq. (7) to obtain the minimum required sample size \hat{N} also ends up being more complex:

$$|\beta|_{min} \sqrt{\hat{N}} e^{-\left(\frac{c|\beta|_{min}}{\sqrt{\hat{N}}}\right)} > 2.486. \tag{7}$$

As with the gamma function correction equation, this equation can be solved with a computer program that starts with $\hat{N} = 1$ and progressive increments its value to 2, 3, etc., until the smallest positive integer that satisfies the equation is obtained. In MATLAB, the value of e^x is obtained using the function `exp(x)`. In Excel, it is obtained using the function `EXP(x)`. The gamma-exponential method can be seen as a refinement of the inverse square root method.

The New Methods in WarpPLS

The software WarpPLS 6.0 (Kock, 2017) implements the two new methods for minimum sample size estimation in PLS-SEM discussed earlier. The menu option “Explore statistical power and minimum sample size requirements” (see Fig. 3)

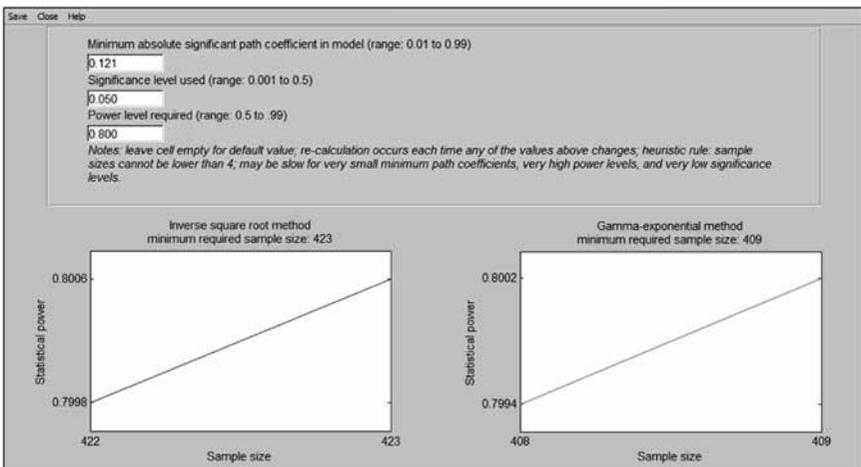


Fig. 3: Minimum Sample Size Estimation in WarpPLS.

allows users to obtain estimates of the minimum required sample sizes for empirical studies based on the following model elements: the minimum absolute significant path coefficient in the model (e.g., 0.121), the significance level used for hypothesis testing (e.g., 0.05), and the power level required (e.g., 0.8).

We can see in the figure that the two new methods are fairly precise, yielding minimum required sample size estimates of 423 and 409, which are very close to the true value estimated by us to be 407 via a Monte Carlo simulation. The degree of precision of these new methods becomes more impressive when we compare the latter with the estimates produced by the 10-times rule and minimum R^2 methods for the same model, which were 30 and 59, respectively. The latter are clearly gross underestimations.

The inverse square root method tends to slightly overestimate the minimum required sample size, while the gamma-exponential method provides a more precise estimate. Given this, empirical researchers are advised to report both estimates, and try to meet the estimate generated by the more conservative of the two methods (i.e., the inverse square root method), which will ensure that the power level achieved by their study will be above the one sought.

Minimum Sample Size Estimation After Data Collection and Analysis

When minimum sample size estimation is conducted after data collection and analysis, its results can be used as a basis for additional data collection, as well as adjustments in the analysis and in the hypothesis-testing assumptions. Minimum sample size estimation after data collection and analysis is known as retrospective estimation; as opposed to prospective estimation, conducted before data collection and analysis. Although there is debate on this topic, the latter (prospective) approach is generally recommended (Gerard, Smith, & Weerakody, 1998; Nakagawa & Foster, 2004).

Additional data collection involves not only collecting additional data points, but also re-testing the model with the new dataset to ensure that the path coefficient with the minimum absolute magnitude has not decreased. Let us assume that a researcher collects 100 data points to test our illustrative model using a data set collected in a different country, and finds that the path coefficient with the minimum absolute magnitude in the model is now 0.237. Using the inverse square root method, the minimum required sample size is estimated to be 111. The researcher then proceeds to collect 11 additional data points and re-tests the model. If the path coefficient with the minimum absolute magnitude in the model is still 0.237 or higher, then the minimum sample size requirement is met.

Instead of collecting additional data points, the researcher may rely on adjustments in the analysis and in the hypothesis-testing assumptions. Taking the previous example as a basis, instead of collecting 11 additional data points, the researcher may simplify the model somewhat by removing one or more competing links (i.e., links from multiple predictors to one criterion latent variable); particularly, links competing with the link (or path) whose coefficient is 0.237. Clearly,

this should be informed by theory and past research; otherwise, the empirical study becomes a data-fitting exercise.

Let us say that the researcher removed one link competing with the link whose path coefficient is 0.237. This removal would have a good chance of increasing the path coefficient with the minimum absolute magnitude, because each additional competing link tends to decrease the path coefficients for other competing links. Let us say that the path coefficient with the minimum absolute magnitude is 0.286 after the removal of one competing link. Using the inverse square root method, the minimum required sample size is estimated to be 76. This minimum required sample size is already met by the 100 data points originally collected. In this case, a simplification of the research model obviates the need for additional data collection.

An alternative to simplifying the model, which does not involve collecting more data either, is to regard a path coefficient that is too low in the context of a given sample size to be nonsignificant regardless of the corresponding p value. For example, let us assume that, with 100 data points, the two path coefficients with the smallest absolute magnitudes are 0.237 and 0.253, both found to be significant at $p < 0.05$ in an empirical study. Using the inverse square root method, the minimum required sample sizes associated with these two coefficients would, respectively, be 111 (as noted before) and 97. Here, the researcher would regard the analysis to have failed to support the hypothesis associated with the 0.237 path, and succeeded in its support of the hypothesis associated with the 0.253 path.

Minimum Sample Size Estimation Before Data Collection and Analysis

Minimum sample size estimation before data collection and analysis, or prospective estimation, is generally recommended over the retrospective approach of estimation after data collection and analysis (Gerard et al., 1998; Nakagawa & Foster, 2004). In prospective estimation, the researcher must decide at the outset the acceptable value of the path coefficient with the minimum absolute magnitude. This is likely to drive hypothesis testing beyond considerations regarding p values. In this context, an important question is: What is a reasonable acceptable value of the path coefficient with the minimum absolute magnitude in a model?

Based on Cohen's (1988, 1992) power-assessment guidelines, a reasonable answer to this question would be a value that would satisfy $\beta^2/(1 - \beta^2) > 0.02$ in a very simple model with only one predictor and one criterion latent variable. In other words, the effect size measured via Cohen's f^2 coefficient in a model with only two variables X and Y , linked as $X \rightarrow Y$, would have to be greater than Cohen's (1988, 1992) minimum acceptable effect size of 0.02.

Models that are more complex would tend to lead to lower effect sizes, because such models would likely include more competing links. Given this, we could set as our target an effect size that is twice Cohen's (1988, 1992) minimum acceptable, namely an effect size of 0.04. Our continuing research on this topic, including a variety of targeted Monte Carlo simulations, suggests that this rule of thumb covers the vast majority of models; including fairly complex models, as long as they

are free of vertical and lateral collinearity (Kock & Lynn, 2012). The corresponding inequality for this proposed rule of thumb would be $\beta^2/(1 - \beta^2) > 0.04$, whose solution is $\beta \geq 0.197$.

Using the inverse square root method, the previously mentioned would lead to a minimum required sample size of 160. Given this, another general rule of thumb could be proposed; this one as an answer to the following question: What is a reasonable value for minimum sample size, if we do not know in advance the value of the path coefficient with the minimum absolute magnitude? The answer would be 160, based on the inverse square root method. Based on the gamma-exponential method, the answer would be 146.

A different approach for prospective minimum sample size estimation is to set the acceptable value of the path coefficient with the minimum absolute magnitude based on past empirical research or the results of a pilot study. Either of these could suggest a large path coefficient of minimum absolute magnitude, which would lead to a relatively small sample size requirement. The danger here is in underestimating the minimum required sample size, which would call for conservative prospective estimations of the path coefficient of minimum absolute magnitude.

For example, if past empirical research or a pilot study suggests a path coefficient of minimum absolute magnitude of 0.35, the inverse square root method would yield a minimum required sample size of 51. Still, after having collected and analyzed 51 data points in an empirical study, a researcher would have to make sure that the path coefficient of minimum absolute magnitude was not lower than the expected 0.35. (A path coefficient of minimum absolute magnitude equal to or higher than 0.35 would have been acceptable.) If the path coefficient of minimum absolute magnitude turned out to be lower than the expected 0.35, the researcher would have to rely on approaches similar to those discussed earlier in connection with retrospective estimation (e.g., additional data collection).

Discussion and Conclusion

One of the most fundamental issues in PLS-SEM is that of minimum sample size estimation, where the “10-times rule” method has been a favorite (Hair et al., 2011) due to its simplicity of application – it builds on the rule that the sample size should be greater than 10 times the maximum number of inner or outer model links pointing at any latent variable in the model.

In spite of the 10-times rule method’s simplicity of application, it has been shown in the past to lead to grossly inaccurate estimates (Goodhue et al., 2012; Kock & Hadaya, 2018). We proposed two related methods, based on mathematical equations, as alternatives for minimum sample size estimation in PLS-SEM: the inverse square root method and the gamma-exponential method. Informed by the results of various Monte Carlo experiments, we show that both methods are fairly accurate. We also showed that the first method is particularly attractive in terms of its simplicity of application. We employed a study in the field of tourism and hospitality research as a basis for the development of an illustrative model, which we used to ground our discussion of minimum sample size estimation on a practical example.

Our results suggest that the methods we propose here for minimum sample size estimation are not affected by capitalization on error, a common problem with PLS-SEM that is characterized by path coefficient overestimations when both coefficients and samples are small. The main reason for this robustness is that our proposed methods tend to generate minimum sample size estimates for small path coefficients that are far above the sample sizes at which capitalization on error occurs. The fact that our methods focus on high power values (i.e., greater than 0.8) for minimum sample size estimation is a key element in avoiding bias due to capitalization on error. If we had tried to develop methods to estimate minimum sample sizes for fairly low power values (e.g., 0.2), capitalization on error might become an issue.

Collecting and analyzing data at multiple levels of analysis can have an impact on minimum sample size estimation; for example, collecting data at the individual and team levels. For example, let us say that 17 teams comprising five individuals each were used for data collection in an empirical study. If we collected and analyzed data at the team level of analysis, this would lead to a rather small sample size of 17, and team aggregation of the individual data. If we had considered the individual team member to be the unit of analysis, the sample size would have been 85, but the analysis would have to become more complex in order to control for the effect of team membership on various hypothesized relationships. [Kock and Hadaya \(2018\)](#) provide a more detailed discussion on this; see Appendix F of that article.

The software WarpPLS 6.0 ([Kock, 2017](#)) implements the two new methods for minimum sample size estimation in PLS-SEM. We showed in our analysis using WarpPLS that the inverse square root method tends to slightly overestimate the minimum required sample size, while the gamma-exponential method provides a more precise estimate. Therefore, we advise empirical researchers to provide both estimates in research reports, and try to meet the estimate generated by the more conservative of the two methods (i.e., the inverse square root method). This will ensure that the power level achieved by their study will be safely above the one sought.

Acknowledgments

This book chapter is based on a previous version published as an article in the *Information Systems Journal*, and contains revised text, originally written by the author, from articles published in the *Journal of Sustainable Tourism*, the *International Journal of e-Collaboration*, and the *Journal of Modern Applied Statistical Methods*. This is done with full permission by the copyright holders, and with the goal of timely and wide dissemination of scholarly work. The author is the developer of the software WarpPLS, which has over 7,000 users in more than 33 different countries at the time of this writing, and moderator of the PLS-SEM email distribution list. He is grateful to those users, and to the members of the

PLS-SEM email distribution list, for questions, comments, and discussions on topics related to SEM and to the use of WarpPLS.

References

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159.
- Gerard, P. D., Smith, D. R., & Weerakkody, G. (1998). Limits of retrospective power analysis. *Journal of Wildlife Management*, 62(2), 801–807.
- Goodhue, D. L., Lewis, W., & Thompson, R. (2012). Does PLS have advantages for small sample size or non-normal data? *MIS Quarterly*, 36(3), 981–1001.
- Gurland, J., & Tripathi R. C. (1971). A simple approximation for unbiased estimation of the standard deviation. *American Statistician*, 25(4), 30–32.
- Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarteed, M. (2014). *A primer on partial least squares structural equation modeling*. Thousand Oaks, CA: Sage.
- Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *The Journal of Marketing Theory and Practice*, 19(2), 139–152.
- Kipnis, C., & Varadhan, S. R. S. (1986). Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Communications in Mathematical Physics*, 104(1), 1–19.
- Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York, NY: Guilford Publications.
- Kock, N. (2010). Using WarpPLS in e-collaboration studies: An overview of five main analysis steps. *International Journal of e-Collaboration*, 6(4), 1–11.
- Kock, N. (2014a). *Stable P value calculation methods in PLS-SEM*. Laredo, TX: ScriptWarp Systems.
- Kock, N. (2014b). Advanced mediating effects tests, multi-group analyses, and measurement model assessments in PLS-based SEM. *International Journal of e-Collaboration*, 10(3), 1–13.
- Kock, N. (2015). One-tailed or two-tailed P values in PLS-SEM? *International Journal of e-Collaboration*, 11(2), 1–7.
- Kock, N. (2016). Non-normality propagation among latent variables and indicators in PLS-SEM simulations. *Journal of Modern Applied Statistical Methods*, 15(1), 299–315.
- Kock, N. (2017). *WarpPLS User Manual: Version 6.0*. Laredo, TX: ScriptWarp Systems.
- Kock, N. (2018). Should bootstrapping be used in PLS-SEM: Toward stable *p*-value calculation methods. *Journal of Applied Structural Equation Modeling*, 2(1), 1–12.
- Kock, N., & Hadaya, P. (2018). Minimum sample size estimation in PLS-SEM: The inverse square root and gamma-exponential methods. *Information Systems Journal*, 28(1), 227–261.
- Kock, N., & Lynn, G. S. (2012). Lateral collinearity and misleading results in variance-based SEM: An illustration and recommendations. *Journal of the Association for Information Systems*, 13(7), 546–580.
- Mandal, P., Mukhopadhyay, S., Bagchi, K., & Gunasekaran, A. (2012). The impact of organisational strategy, culture, people and technology management on organisational practice and performance: An empirical analysis. *International Journal of Information Systems and Change Management*, 6(2), 160–176.

- Miller, R. B., & Wichern, D. W. (1977). *Intermediate business statistics: Analysis of variance, regression and time series*. New York, NY: Holt, Rinehart and Winston.
- Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, 9(4), 599–620.
- Nakagawa, S., & Foster, T. M. (2004). The case against retrospective statistical power analyses with an introduction to power analysis. *Acta Ethologica*, 7(2), 103–108.
- Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., & Chen, F. (2001). Monte Carlo experiments: Design and implementation. *Structural Equation Modeling*, 8(2), 287–312.
- Peng, D. X., & Lai, F. (2012). Using partial least squares in operations management research: A practical guideline and summary of past research. *Journal of Operations Management*, 30(6), 467–480.
- Rasoolimanesh, S. M., Dahalan, N., & Jaafar, M. (2016). Tourists' perceived value and satisfaction in a community-based homestay in the Lenggong Valley World Heritage Site. *Journal of Hospitality and Tourism Management*, 26(1), 72–81.
- Rasoolimanesh, S. M., Jaafar, M., Kock, N., & Ahmad, A. G. (2017). The effects of community factors on residents' perceptions toward World Heritage Site inscription and sustainable tourism development. *Journal of Sustainable Tourism*, 25(2), 198–216.
- Robert, C., & Casella, G. (2013). *Monte Carlo statistical methods*. New York, NY: Springer.
- Weakliem, D. L. (2016). *Hypothesis testing and model selection in the social sciences*. New York, NY: Guilford Publications.