MEASUREMENT AND RESEARCH METHODS IN INTERNATIONAL MARKETING

ADVANCES IN INTERNATIONAL MARKETING

Series Editors: Volumes 1–20 – S. Tamer Cavusgil
Volume 21 – Shaoming Zou

Recent Volumes:

Volume 5:	Industrial Networks
Volume 6:	Export Marketing: International Perspectives
Volume 7:	Marketing in Asia Pacific and Beyond
Volume 8:	Parts I & II
Volume 9:	International Marketing and Purchasing
Volume 10:	Globalization, the Multinational Firm, and Emerging Economies
Volume 11:	Reassessing the Internationalization of the Firm
Volume 12:	New Directions in International Advertising Research
Volume 13:	Study Abroad: Perspectives and Experiences from Business Schools
Volume 14:	Reviving Traditions in Research on International Market Entry
Volume 15:	Research on International Service Marketing: A State of the Art
Volume 16:	Relationship between Exporters and their Foreign Sales and Marketing Intermediaries
Volume 17:	International Market Research: Opportunities and Challenges in the 21st Century
Volume 18:	Cross-Cultural Buyer Behaviour
Volume 19:	Michigan State University Contributions to International Business and Innovation
Volume 20:	New Challenges to International Marketing
Volume 21:	International Marketing: Emerging Markets

ADVANCES IN INTERNATIONAL MARKETING
VOLUME 22

MEASUREMENT AND RESEARCH METHODS IN INTERNATIONAL MARKETING

EDITED BY

MARKO SARSTEDT
Ludwig-Maximilians-Universität München, Germany

MANFRED SCHWAIGER
Ludwig-Maximilians-Universität München, Germany

CHARLES R. TAYLOR
Villanova School of Business, USA

Emerald

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

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

First edition 2011

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-1-78052-094-0
ISSN: 1474-7979 (Series)

LIST OF CONTRIBUTORS

Rolph E. Anderson	Drexel University, Philadelphia, PA, USA
Hae-Kyong Bang	Villanova School of Business, Villanova, PA, USA
C. Luke Bowen	Villanova School of Business, Villanova, PA, USA
Peggy E. Chaudhry	Villanova School of Business, Villanova, PA, USA
Adamantios Diamantopoulos	University of Vienna, Vienna, Austria
Sandra Diehl	Alpen-Adria University of Klagenfurt, Klagenfurt, Austria
George R. Franke	University of Alabama, Tuscaloosa, AL, USA
Siegfried P. Gudergan	University of Newcastle, Newcastle, NSW, Australia
Jörg Henseler	Radboud University Nijmegen, Nijmegen, The Netherlands
John S. Hill	University of Alabama, Tuscaloosa, AL, USA
Ronald Paul Hill	Villanova School of Business, Villanova, PA, USA
Hyokjin Kwak	Drexel University, Philadelphia, PA, USA
Ryan Luchs	Duquesne University, Pittsburgh, PA, USA
Barbara Mueller	San Diego State University, San Diego, CA, USA

Shintaro Okazaki	Universidad Autónoma de Madrid, Madrid, Spain
Marina Puzakova	Drexel University, Philadelphia, PA, USA
Sascha Raithel	Ludwig-Maximilians-Universität München, Munich, Germany
Jase Ramsey	University of Alabama, Tuscaloosa, AL, USA
R. Glenn Richey	University of Alabama, Tuscaloosa, AL, USA
Petra Riefler	University of Vienna, Vienna, Austria
Edward E. Rigdon	Georgia State University, Atlanta, GA, USA
Christian M. Ringle	Hamburg University of Technology (TUHH), Hamburg, Germany and University of Newcastle, Newcastle, NSW, Australia
Joseph F. Rocereto	Monmouth University, West Long Branch, NJ, USA
Matthew E. Sarkees	West Virginia University, Morgantown, WV USA
Marko Sarstedt	Ludwig-Maximilians-Universität München, Munich, Germany and University of Newcastle, Newcastle, NSW, Australia
Sebastian Scharf	Biesalski & Company, Munich, Germany
Manfred Schwaiger	Ludwig-Maximilians-Universität München, Munich, Germany
Stephen A. Stumpf	Villanova School of Business, Villanova, PA, USA

Charles R. Taylor	Villanova School of Business, Villanova, PA, USA
Goksel Yalcinkaya	University of New Hampshire, Durham, NH, USA
Lorenz Zimmermann	Ludwig-Maximilians-Universität München, Munich, Germany

INTRODUCTION: MEASUREMENT AND RESEARCH METHODS IN INTERNATIONAL MARKETING

"Garbage in, garbage out" is a common expression that academics and practitioners use to emphasize that empirical analysis is only as good as the basis on which it relies. Although the importance of sound data and valid measures has long been acknowledged, it is nevertheless often problematic to follow required quality standards in concrete research situations. Potential sources of error are usually unknown, methods to ensure data quality are unavailable, and existing methods for scale development, index construction, data collection, and data analysis are insufficient or erroneously applied. This is especially true of international marketing research, which often makes great demands on the data and measures used, as well as on the research methodology applied. Against this background, this volume addresses issues pertaining to measurement and research methodology in an international marketing context. Thanks to the efforts of authors and reviewers, we are pleased to present nine articles that deal with cutting-edge topics such as formative measurement, response-bias in cross-cultural research, marketing efficiency measurement, and segmentation methods.

We feel confident this selection of research papers will help researchers and practitioners alike address quality issues related to measurement and data analysis in international marketing correctly.

The first part of the volume deals with measurement in international marketing research.

This part's first chapter, titled "Using Formative Measures in International Marketing Models: A Cautionary Tale Using Consumer Animosity as an Example," by Adamantios Diamantopoulos and Petra Riefler deals with formative measurement which, after years of neglect, is finally receiving more attention in various social and behavioral science disciplines (e.g., Bagozzi 2011; Bollen 2011; Diamantopoulos, 2011; Diamantopoulos, Riefler, & Roth, 2008). However, despite the increasing number of formative measurement models found in the literature, little is known about the potential consequences of their use for substantive theory testing.

Against this background, the authors highlight some problems that may arise when formative instead of reflective measures are used to test even simple theoretical models. They also illustrate approaches that help overcome these problems, and pinpoint the results' potential interpretation difficulties with regard to re-specified measurement models. Thus, the chapter stimulates discussion on the implications for theory development when models with formative measures are established and tested.

The second chapter in this part by George R. Franke, John S. Hill, Jase Ramsey, and R. Glenn Richey titled "Difference Scores, Analysis Levels, and the (Mis)interpretation of Cultural Distance" demonstrates previously unrecognized problems with the conceptualization, analysis, and interpretation of cultural distance measures. The authors' analytical and empirical analyses show that the difference scores that are implicit in measures of cultural distance usually imply unrealistic constraints on relationships between variables. Furthermore, analyzing cultural distance at the level of organizations rather than countries exaggerates the available sample size and may result in inaccurate statistical tests. On the basis of their findings, the authors suggest methods for improvements in cultural distance research.

The third chapter, titled "The Role of Response Formats on Extreme Response Style: A Case of Likert-Type vs. Semantic Differential Scales," by Joseph F. Rocereto, Marina Puzakova, Rolph E. Anderson, and Hyokjin Kwak deals with response bias in cross-cultural research, which may systematically differ from one culture to another, violating the assumption of measurement equivalence. Specifically, the authors investigate the role of response format type on extreme response style in different cultures, showing that differences occur with regard to Likert-type scales, whereas no significant differences arise when utilizing the semantic differential format.

Lastly, in their study "A Multicountry Advertising Research Framework: Lessons Learned From Testing Global Consumer Culture Positioning," Shintaro Okazaki, Barbara Mueller, and Sandra Diehl propose a framework that is useful for conducting multicountry marketing and advertising research. To illustrate the series of steps involved in conducting such investigations, the authors present a six-country study examining global consumer culture positioning. The suggested steps are relevant for the exploration of a wide variety of marketing and advertising-related topics.

Chapters in the second part deal with methods to measure marketing efficiency in an international marketing context. The first chapter by Matthew E. Sarkees and Ryan Luchs, titled "Stochastic Frontier Estimation in International Marketing Research: Exploring Untapped Opportunities," describes the basic characteristics of stochastic frontier estimation, which allows researchers to model efficiency issues using combinations of inputs and outputs. The method has been commonly applied in economics, but despite its relevance for international marketing research, its application is comparatively scarce in this field. By discussing the method's advantages, providing an application on data from the pharmaceutical industry, and discussing further potential applications in international marketing, the authors make a strong case for the method's suitability for tackling a broad range of research questions.

One of these research questions is the subject of the second chapter in this part titled "Marketing Accountability: Applying Data Envelopment Analysis to Assess the Impact of Advertising Efficiency on Shareholder Value," by Sascha Raithel, Sebastian Scharf, Charles R. Taylor, Manfred Schwaiger, and Lorenz Zimmermann. Specifically, the authors pick up the ongoing discussion on the accountability of marketing (e.g., Jacobson & Mizik, 2009; Luo & Homburg, 2008; Raithel, Sarstedt, Scharf, & Schwaiger, 2011) and examine advertising efficiency's effect on a firm's stock market performance. The authors illustrate how data envelopment analysis can be used to measure marketing expenditures' efficiency and combine the method with a stock return response modeling technique to evaluate marketing performance effects over time. Their results imply that managers should not limit their tactics to increasing market-based assets at any cost and raising budgets if they wish to send a positive signal to investors.

The third part deals with methodological advances in international marketing.

The first chapter, titled "The State of Methodological Practice in International Marketing Research," is by Charles R. Taylor, C. Luke Bowen, and Hae-Kyong Bang. The authors conduct a content analysis of papers published in leading marketing and advertising journals from 2005–2010 to examine whether accepted principles for conducting cross-national research are being followed. To this end, the chapter begins by outlining several guidelines that should be followed, including issues such as providing a compelling rationale for the countries being studied, conducting back-translations, measuring cultural factors, and conducting post hoc equivalence tests. The chapter also examines the theoretical bases, primary methodologies, and analytical techniques applied that have evolved over the years. The analysis also compares the degree to which general marketing, advertising, and international marketing/business journals follow the guidelines.

The second chapter in this part by Edward E. Rigdon, Christian M. Ringle, Marko Sarstedt, and Siegfried P. Gudergan, titled "Assessing Heterogeneity in Customer Satisfaction Studies: Across Industry Similarities and within Industry Differences,'' broaches the issue of heterogeneity in the context of partial least squares (PLS) path modeling (e.g., Hair, Ringle, & Sarstedt, 2011), which has recently attracted considerable research interest in different fields (e.g., Hair, Sarstedt, Ringle, & Mena, 2012; Navarro, Acedo, Losada, & Ruzo, 2011; Rigdon, Ringle, & Sarstedt, 2010; Sarstedt, Becker, Ringle, & Schwaiger, 2011). Specifically, the authors look at evidence for observed and unobserved heterogeneity within data underlying the American Customer Satisfaction Index (ACSI) model. Using the finite mixture PLS path modeling (FIMIX-PLS; Hahn, Johnson, Herrmann, & Huber, 2002; Sarstedt et al., 2011) method, the authors uncover three latent segments of customers within each industry which differ significantly in terms of the impact of customer loyalty's anteceding constructs. As such, the chapter underlines the need to be open to differences across different populations, rather than attempting to impose one rigid model across distinct groups or cultures.

Regardless of whether researchers partition their data based on observable characteristics or derive latent segments by means of response-based segmentation techniques, both procedures share the final step of analysis: A comparison of parameter estimates across the identified segments, using PLS path modeling-based approaches to multi-group analysis. The second chapter, titled ''Multigroup Analysis in Partial Least Squares (PLS) Path Modeling: Alternative Methods and Empirical Results,'' by Marko Sarstedt, Jörg Henseler, and Christian M. Ringle deals with this topic by critically reviewing available approaches to multigroup analysis within a PLS path modeling framework. The authors also illustrate their characteristics by means of empirical data and propose two novel approaches to compare two and more groups of data.

There are a few people whom we would like to thank for their contribution to this project. First, we thank Shaoming Zou for encouraging us to edit an issue on this topic and for his advice along the way. It is his vision for *Advances in International Marketing* that allows these volumes to consistently contain high-quality work by leading scholars. We also thank Emerald, and particularly Mary Miskin and Stephanie Hull for their help in editing the manuscript. Finally, we appreciate the patience and support of our families during the time we devoted to this project.

Marko Sarstedt
Manfred Schwaiger
Charles R. Taylor
Guest Editors

REFERENCES

Bagozzi, R. P. (2011). Measurement and meaning in information systems and organizational research: Methodological and philosophical foundations. *MIS Quarterly*, *35*(2), 261–292.

Bollen, K. A. (2011). Evaluating effect, composite, and causal indicators in structural equation models. *MIS Quarterly*, *35*(2), 359–372.

Diamantopoulos, A. (2011). Incorporating formative measures into covariance-based structural equation models. *MIS Quarterly*, *35*(2), 335–358.

Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research*, *61*(12), 1203–1218.

Hahn, C., Johnson, M. D., Herrmann, A., & Huber, F. (2002). Capturing customer heterogeneity using a finite mixture PLS approach. *Schmalenbach Business Review*, *54*(3), 243–269.

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*, forthcoming.

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory and Practice*, *19*(2), 139–151.

Jacobson, R., & Mizik, N. (2009). The financial markets and customer satisfaction: Reexamining possible financial market mispricing of customer satisfaction. *Marketing Science*, *28*(5), 836–845.

Luo, X., & Homburg, C. (2008). Satisfaction, complaint, and the stock value gap. *Journal of Marketing*, *72*(4), 29–43.

Navarro, A., Acedo, F. J., Losada, F., & Ruzo, E. (2011). Integrated model of export activity: Analysis of heterogeneity in managers' orientations and perceptions on strategic management in foreign markets. *Journal of Marketing Theory & Practice*, *19*(2), 187–204.

Raithel, S., Sarstedt, M., Scharf, S., & Schwaiger, M. (2011). On the value relevance of customer satisfaction. Multiple drivers and multiple markets. *Journal of the Academy of Marketing Science*, forthcoming.

Rigdon, E. E., Ringle, C. M., & Sarstedt, M. (2010). Structural modeling of heterogeneous data with partial least squares. In: N. K. Malhotra (Ed.), *Review of marketing research* (Vol. 7, pp. 255–296). Armonk: Sharpe.

Sarstedt, M., Becker, J.-M., Ringle, C. M., & Schwaiger, M. (2011). Uncovering and treating unobserved heterogeneity with FIMIX-PLS: Which model selection criterion provides an appropriate number of segments? *Schmalenbach Business Review*, *63*(1), 34–62.

USING FORMATIVE MEASURES IN INTERNATIONAL MARKETING MODELS: A CAUTIONARY TALE USING CONSUMER ANIMOSITY AS AN EXAMPLE

Adamantios Diamantopoulos and Petra Riefler

ABSTRACT

Purpose – *Despite the increasing use of formative measurement models in literature, little is known about potential consequences for substantive theory testing. Against this background, the aims of this chapter are (1) to highlight some problems that may arise when formative instead of reflective measures are used to test even simple theoretical models with covarianced-based methodologies, (2) to illustrate some approaches that might help overcome these problems, (3) to pinpoint potential interpretation difficulties of the results involving re-specified measurement models, and (4) to stimulate discussion on the implications for theory development when models are tested with formative measures.*

Methodology/approach – *Potential consequences of formative measurement models for theory testing are highlighted using an empirical study on consumer animosity as an illustrative example and applying covarianced-based structural equations modeling procedures for estimation purposes.*

Measurement and Research Methods in International Marketing
Advances in International Marketing, Volume 22, 11–30

ISSN: 1474-7979/doi:10.1108/S1474-7979(2011)0000022004

Findings – *The empirical study shows (a) that some scaling options for the (composite) latent variable result in non-convergence problems, (b) that, assuming convergence, parameter estimates, standard errors, and significance levels vary depending on the scaling method used, and (c) that goodness-of-fit statistics cannot be used as diagnostic measures for the appropriateness of divergent results.*

Originality/value of paper – *The contribution of this chapter is two-fold: First, it shows that to enable estimation, it is often necessary to modify (i.e., expand) the original theoretical model in a conceptually reasonable manner and to do so before data collection. Second, it demonstrates that alternative scaling options for composite latent variables may result in inconsistent substantive conclusions. Consequently, the impact of formative measurement on theory testing is a critical topic and needs to receive further attention in future literature.*

Keywords: Consumer animosity; formative measurement; scaling; theory testing

INTRODUCTION

In recent years, formative measurement has received increasing attention in various disciplines including organization research (e.g., Podsakoff, MacKenzie, Lee, & Podsakoff, 2003), strategy (e.g., Podsakoff, Shen, & Podsakoff, 2006), information systems (e.g., Petter, Straub, & Rai, 2007), and marketing (e.g., Jarvis, MacKenzie, & Podsakoff, 2003). With specific reference to international marketing, formative measurement models have been applied to operationalize such constructs as export performance (Diamantopoulos, 1999), industry drivers (Johansson & Yip, 1994), export coordination (Diamantopoulos & Siguaw, 2006), psychic distance (Brewer, 2007), and product-country images (Diamantopoulos & Papadopoulos, 2010).[1]

Extant literature on formative measurement is primararily of a technical nature focusing on issues such as model specification, identification, and estimation (e.g., Bollen & Davis, 2009; Cenfetelli & Bessellier, 2009; Diamantopoulos & Winklhofer, 2001; Petter et al., 2007). This technical focus of the relevant literature is both understandable and commendable in light of mounting evidence that misspecification of measurement models can have wide-reaching consequences in terms of parameter estimation, fit assessment and, ultimately, the substantive conclusions drawn (e.g., Jarvis et al., 2003). What seems to be missing, however, are contributions focusing on the implications of using formatively measured constructs for theory testing purposes (for a notable exception, see Franke, Preacher, & Rigdon, 2008). Although it is by now acknowledged that many constructs are best operationalized under a formative perspective, little is known how the implementation of such a perspective in (existing) theoretical models affects substantive results.

In this chapter, we use consumer animosity (Klein, Ettenson, & Morris, 1998) as an illustrative construct to discuss some problems that might arise when seeking to test an established theoretical model but with one of its key constructs re-specified (from reflective to formative) for measurement purposes. Consumer animosity, is defined as ''remnants of antipathy related to previous or ongoing military, political or economic events'' (Klein et al., 1998, p. 90) and has become a construct of key interest in international marketing literature as reflected in a large number of published studies (e.g., Ettenson & Klein, 2005; Funk, Arthurs, Treviño, & Joireman, 2010; Huang, Phau, & Lin, 2010; Jiménez & San Martin, 2010; Klein, 2002; Leong et al., 2008; Nijssen & Douglas, 2004; Shin, 2001). Originally specified as a (second-order) reflective measurement model, Riefler and Diamantopoulos (2007, p. 105) argued that ''a formative measurement model … may be more consistent with the conceptual definition of animosity''. The latter authors, however, did not try to test Klein et al.'s (1998) original model with animosity operationalized formatively. Moreover, most subsequent research studies on consumer animosity have kept the (conceptually problematic) reflective specification (e.g., Funk et al., 2010; Jiménez & San Martin, 2010; Leong et al., 2008; a notable exception is Maher and Mady's (2010) study).

Against this background, we draw from an empirical investigation of consumer animosity in Austria in order to (1) highlight some problems that may arise when formative instead of reflective measures are used to test (even simple) theoretical models, (2) illustrate some approaches that might help to overcome these problems, (3) pinpoint potential interpretation difficulties of the results involving re-specified measurement models, and (4) stimulate a discussion on the implications for theory development when models are tested with formative measures.

Note that our intended contribution is methodological and our use of the animosity construct is purely for illustrative purposes. Note also that we focus exclusively on the testing of models via covariance-based methodologies (as implemented, for example, in LISREL, EQS, or AMOS) and do not consider variance-based methodologies such as partial least squares (as implemented, for example, in SmartPLS or PLSGraph). The reason for this is that ''PLS's lack of a global optimization function and consequently measures of global goodness of model fit definitely limits the use of PLS for the theory testing'' (Henseler, Ringle, & Sinkovics, 2009, p. 297).

REVISITING KLEIN ET AL.'S (1998) ANIMOSITY MODEL OF FOREIGN PRODUCT PURCHASE

The original animosity model of foreign product purchase by Klein et al. (1998) incorporates consumer animosity and product judgment as two core predictor constructs, and willingness to buy products of the target country as the dependent construct (Fig. 1).[2] The lack of a causal link between animosity and product judgment is based on the rationale that ''consumers might avoid products from the offending nation not because of concern about the quality of goods, but because the exporting nation has engaged in military, political or economic acts that a consumer finds both grievous and difficult to forgive'' (Klein et al., 1998, p. 90). Using a sample of Chinese consumers, the authors found a negative impact of consumer animosity on willingness to buy Japanese products, whereas, as expected, product judgment impacted positively willingness to buy. These results have been confirmed in a number of replication studies in different country settings (e.g., Klein, 2002; Nijssen & Douglas, 2004; Shin, 2001).

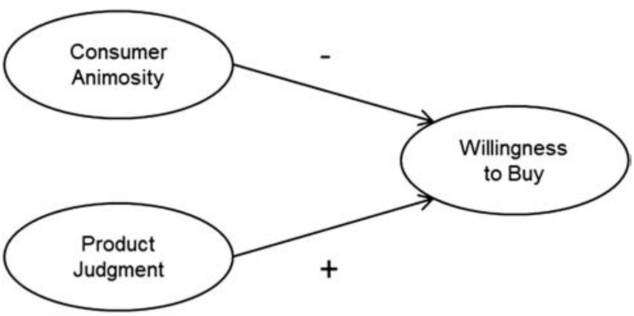

Fig. 1. The Core Animosity Model of Foreign Product Purchase (modified from Klein et al., 1998).

Klein et al.'s (1998) original measurement model is a higher-order model which consists of two reflective first-order constructs (war and economic animosity, respectively) plus one directly measured item capturing general animosity. War animosity is measured by three reflective items (e.g., "I will never forgive Japan for the Nanjing Massacre"), economic animosity by five reflective items (e.g., "Japan is not a reliable trading partner") and general animosity by a single item ("I dislike the Japanese"). According to this model specification, the second-order latent variable (i.e., animosity) causes variation in the first-order latent variables (i.e., war and economic animosity) and, in turn, the latter cause variation in the items (Jarvis et al., 2003; Law, Wong, & Mobley, 1998). From a substantive perspective, what this means is that any change (i.e., increase or decrease) in animosity will *necessarily* result in a change in *both* war *and* economic (and general) animosity.

Based on a comprehensive review of animosity literature, Riefler and Diamantopoulos (2007) take issue with the measurement model of Klein et al. (1998). Specifically, drawing on the conceptual description of animosity as being "the result of hostility *stemming from* consumers' perceptions of a particular nation's action" (Klein & Ettenson, 1999, p. 7, added emphasis), Riefler and Diamantopoulos (2007) argue that the implied causality from the construct to the measures represents a fundamental misspecification of the measurement model. In particular, they point out that economic and war animosity (as well as any other sources of animosity such as politically or culturally based animosity) would determine the overall animosity felt by consumers; they state that consumers "are likely to dislike a country because of the inexcusableness of its acting, and not the other way around" (Riefler & Diamantopoulos, 2007, p. 105). Based on this argument, they propose a formative specification of the animosity construct whereby economic, war, political, and cultural animosities are modeled as distinct components/facets which cause overall animosity towards the offending country.

RESPECIFYING KLEIN ET AL.'S (1998) ORIGINAL ANIMOSITY MODEL

Although estimation of the original animosity model in Fig. 1 is straightforward when consumer animosity is operationalized under a reflective measurement perspective, this is no longer the case if a formative operationalization is opted for. This is because the re-specified model does not satisfy the "2+ emitted paths rule" and is thus under-identified

(Bollen & Davis, 2009; MacCallum & Browne, 1993; Temme, 2006). This identification rule requires that the formative latent variable emits *at least two paths* to other (reflective) constructs or indicators. This requirement is clearly not satisfied in the model in Fig. 1 because there is only a single outcome variable (willingness to buy). To enable estimation, it is thus necessary to add a second emitting path from animosity to another construct. Unfortunately, adding the (obvious) path from animosity to product judgment (as shown in Fig. 2) does *not* solve the identification problem. This is because product judgment and willingness to buy are themselves related (with the former impacting the latter), whereas according to the "2+ emitted paths rule," the outcome variables *cannot* be interrelated; neither a structural path between them (as is the case here) nor correlated disturbance terms are allowed (Bollen & Davis, 2009; Temme, 2006). Of course, one can solve the identification by simply removing the link from product judgment to willingness to buy (as shown in Fig. 3). Doing so, however, would result in a theoretically flawed model for two reasons. First, arguing that consumers are willing to buy products irrespective of whether they evaluate such products as being good or bad would be tantamount to stating that buying behavior is completely random and consumers wholly irrational. Second, eliminating this link no longer makes it possible to test whether animosity impacts willingness to buy independently of product judgment. Given that this is a major theoretical premise of the consumer animosity model by Klein et al. (1998), being unable to test it is clearly unsatisfactory.

From the above, it becomes evident than even a very simple model – such as the one in Fig. 1 – can become problematic for testing purposes simply because the measurement of one predictor construct has been re-specified to be formative

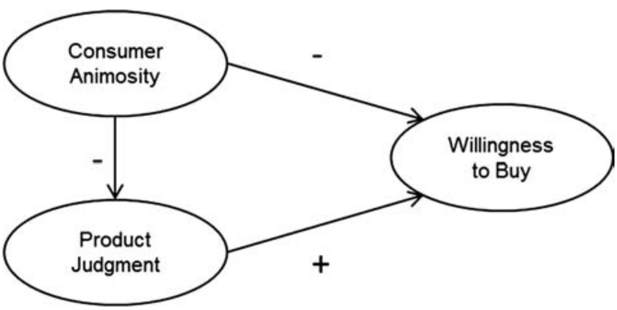

Fig. 2. Unidentified Revised Model.

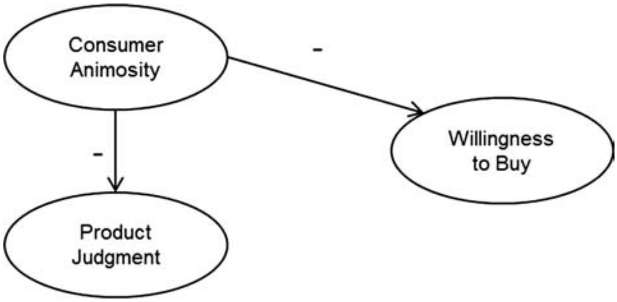

Fig. 3. Theoretically Untenable Revised Model.

rather than reflective. To enable estimation, one must therefore modify the original model so as to solve identification problems but, at the same time, do so in a theoretically defensible manner. For our illustrative model in Fig. 1, these requirements imply that (1) at least one additional (reflectively measured) construct needs to be introduced in an endogenous position, (2) this construct must be theoretically influenced by consumer animosity, and (3) this construct must *not* be related to at least one other endogenous construct which is directly influenced by animosity (so that the "2+ emitted paths rule" is satisfied).

Fig. 4 shows a revised model following these guidelines and incorporating country affect as the additional construct. Country affect refers to "positive or negative emotions, other subjective states or also to a state of arousal, which consumers can experience toward any (foreign) country and which further lead to particular action tendencies and explicit actions" (Burger, 2009, p. 20) and is expected to be negatively influenced by consumer animosity (i.e., the greater the animosity toward a given country, the more negative the feelings of consumers toward that country).

The model in Fig. 4 is, at first sight, appealing because it solves identification problems while incorporating the original links between animosity, product judgment, and willingness to buy (as shown in Fig. 2) within the larger model. However, theoretically, this expanded model is also assailable since it postulates a zero-relationship between country affect and product judgment. Given that, conceptually, country affect is an integral component of the overall country image construct (Roth & Diamantopoulos, 2009) and given that country image has (repeatedly) been found to impact consumers' product evaluations (e.g., Jaffe & Nebenzahl, 2006; Papadopoulos & Heslop, 1993; Wilcox, 2005), positing a zero link between country affect and product judgment is theoretically untenable (and empirically likely to result in very poor model fit).

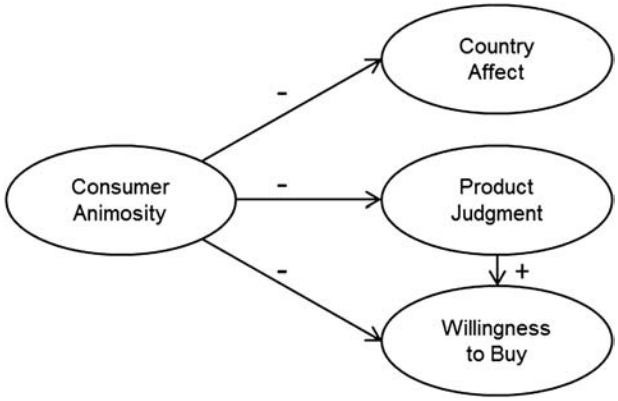

Fig. 4. Expanded Animosity Model I.

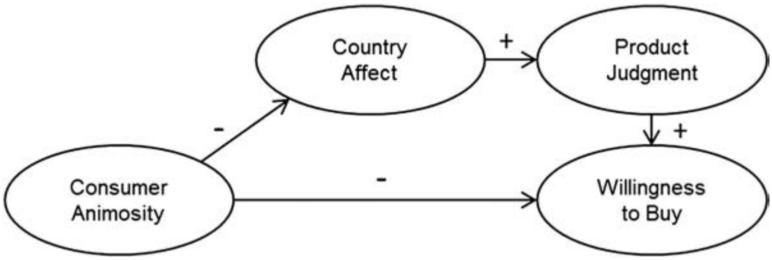

Fig. 5. Expanded Animosity Model II.

Fig. 5 shows a much more theoretically defensible model employing the same sets of constructs as in Fig. 4 while ensuring that identification conditions are still satisfied.

In the model in Fig. 5, identification is achieved as two paths to reflectively measured and unrelated constructs (country affect and willingness to buy) emanate from the formatively measured animosity construct. Here, country affect is allowed to impact product judgment (which is theoretically tenable) and product judgment is directly related to willingness to buy (which is also theoretically tenable). Moreover, the model allows for testing the focal relationship of the original animosity model, that is, the direct link between animosity and willingness to buy (see Fig. 1). Thus the model in Fig. 5 is both theoretically defensible *and* operationally testable. In the following, we will use this model to investigate the potential consequences of the formative re-specification from an empirical perspective.

EMPIRICAL INVESTIGATION OF THE REVISED ANIMOSITY MODEL

Pre-Study: Indicator Development

Literature recommends that "paying tribute to the context-specific nature of the animosity construct, *tailored* measures should be developed based on the reasons revealed by consumers instead of unquestionably adopting the Klein et al. (1998) scale as a generally applicable measure" (Riefler & Diamantopoulos, 2007, p. 144, added emphasis). Consequently, we conducted a pre-study using 107 Austrian consumers (51% female, 17–87 years) to identify (a) the most frequent target country for animosity feelings among Austrian consumers, and (b) the major underlying reasons for these animosity feelings. This pre-study revealed that the USA ranked first as animosity target country (22% of mentions) at the time of the study, which is in line with the finding of Riefler and Diamantopoulos' (2007) earlier exploratory study.

In a first step, the pool of reasons for consumer animosity was reduced by eliminating reasons that were overall of little importance to respondents. This was done using two selection criteria, namely (1) the absolute number of respondents having mentioned a particular reason, and (2) the average number of points allocated to the reason by respondents based on a constant sum task included in this pre-study. This purification led to a reduced indicator pool of 19 empirically relevant animosity reasons. The top-three reasons for animosity were the foreign policy of the USA, the US president in office at that time (George W. Bush), and the legal existence of the death penalty.

Based on this, a total of 19 items capturing various reasons for animosity were formulated. In writing up the items, conventional guidelines regarding clarity, length, lack of ambiguity, and avoidance of jargon were followed (e.g., DeVellis, 2003). Subsequently, five scholars experienced in international marketing were recruited for an item-sort task. We provided them with descriptions of six potential sources of consumer animosity described in literature (i.e., war, political, economic, cultural, religious, and personal animosity) as well as a list of the 19 indicators. We then instructed the participants to carefully read each indicator and assign it to the source which, according to their individual judgment, it best reflects. This sorting task resulted in eight political, six cultural, three economic, and two war-related animosity items; religious and personal animosity was found not relevant for the Austria–USA study setting.

Data Collection and Measurement: Main Study

For the main study, we incorporated the pool of 19 items in a questionnaire. A pre-test with 12 consumers using the protocol approach (Reynolds & Diamantopoulos, 1998) resulted in minor changes in wording to ensure that all items were comprehensible and no difficulties in answering occurred. In a street intercept in various locations of the capital city Vienna, a total of 300 consumers completed the questionnaire (approximately 20% participation rate). After removing respondents with nationalities other than Austrian, a sample of 261 Austrian consumers (50% female, 13–69 years, mean age of 31.5 years) was used for subsequent analysis.

The consumer animosity items were measured using balanced, six-point semantic differential scaling including a "no opinion" answering option. Country affect was assessed on a four-item reflective scale (based on Holbrook, 1981; Shamdasani, Stanaland, & Tan, 2001). Product judgment and willingness to buy were both assessed using reflective six-item scales (based on Darling & Arnold, 1988; Wood & Darling, 1993) as used in the original study by Klein et al. (1998). For all reflective scales, seven-point Likert-type answering formats were used. Applying a forward translation (Craig & Douglas, 2005), the items were independently translated from English to German by three German native speakers fluent in English. The three versions were compared and revised accordingly.[3] The internal consistency was satisfactory for all three reflective scales (country affect: Cronbach's $\alpha = 0.92$, product judgment: $\alpha = 0.76$, and willingness to buy: $\alpha = 0.82$).

Development and Purification of Animosity Index

In the first stage of analysis, the 19 formative items were subjected to a multicollinearity check. Multicollinearity is an undesirable property in formative measurement models as since (a) it can result in unstable estimates for the indicator coefficients, (b) it becomes difficult to separate the distinct influence of individual indicators on the latent variable, and (c) it leads to difficulties in assessing indicator validity on the basis of the magnitude of the coefficients (Cenfetelli & Bessellier, 2009; Diamantopoulos & Winklhofer, 2001; Petter et al., 2007). Following recommendations in literature (e.g., see Kleinbaum, Kupper, Muller, & Azhar, 1998), we used a variance inflation factor (VIF) $>$ 10 as cut-off; all items showed VIFs $<$ 10 and were therefore retained. Finally, following Albers and Hildebrandt (2006), we combined the indicators for each animosity source into an index by calculating the arithmetic mean and using the latter as a single-item measure in an attempt to reduce model complexity. We thus specified four (composite) formative indicators capturing economic, political, war, and cultural animosity resources, respectively.

MODEL ESTIMATION

Scaling Options

In both reflective and formative models, latent variables have no scale on their own and consequently require a unit of measurement for parameter estimates to be statistically identified (Bollen, 1989). Literature suggests three alternative options for scaling formatively measured constructs, namely (1) fixing a path from one (of the four summated) animosity indicators to 1 (e.g., see MacCallum & Browne, 1993), which corresponds to the scaling approach commonly used for reflectively measured constructs; (2) fixing a path from the formatively measured construct to one of the outcome constructs to 1 (e.g., Bollen & Davis, 2009); or (3) standardizing the variance of the formatively measured construct (e.g., Edwards, 2001). Although none of these alternative options is considered as superior in literature, Gonzalez and Griffin (2001) show that standard errors and, consequently, significance tests of model parameters can be affected by the choice of scaling method (see also Diamantopoulos, 2011; Franke et al., 2008). Building on these initial insights, we empirically compare the results obtained for the model in Fig. 5 using the three alternative scaling methods to scale the animosity construct.[4] Country affect, product judgment, and willingness to buy are conventionally scaled by setting one of their (reflective) indicators to 1. The results are summarized in Table 1.

Table 1. SEM Results.

	Scaling Method Used			
	Path from ECON, POL, or CULT Fixed to 1	Path from WAR Fixed to 1	Path from ANIM to AFFECT Fixed to 1	Standardization
Measurement paths				
ECON→ANIM	n/a	2.293 (1.630)	0.233* (3.131)	−0.206* (−2.975)
POL→ANIM		0.134 (0.145)	0.014 (0.150)	−0.012 (−0.150)
CULT→ANIM		4.651 (1.943)	0.474* (7.541)	−0.418* (−5.916)
WAR→ANIM		1.000 (fixed)	0.102* (2.015)	−0.089* (−1.971)
Structural paths				
ANIM→AFFECT	n/a	0.102* (2.015)	1.000 (fixed)	−1.133* (−7.910)
AFFECT→PJ		−0.747* (−9.460)	−0.747* (−9.460)	−0.747* (−9.460)
PJ→WILL		−0.139* (−1.875)	−0.139 (−1.875)	−0.139 (−1.875)
ANIM→WILL		0.071 (1.930)	0.697* (5.874)	−0.790* (−6.185)
Variance explained (R^2)				
ANIM	n/a	0.518	0.518	0.518
AFFECT		0.879	0.879	0.879
PJ		0.364	0.364	0.364
WILL		0.431	0.431	0.431
Fit statistics				
χ^2 (df=67)	n/a	133.072	133.072	133.072
RMSEA		0.062	0.062	0.062
NNFI		0.973	0.973	0.973
CFI		0.980	0.980	0.980
SRMR		0.043	0.043	0.043

Notes: Unstandardized parameter estimates, values in () are t-values; $^*p<0.05$; n/a, solution did not converge.
ECON, economic animosity; POL, political animosity; CULT, cultural animosity; WAR, war animosity; ANIM, consumer animosity; PJ, product judgment; and WILL, willingness to buy. Note that AFFECT and WILL are reverse-scored; hence the negative coefficients between AFFECT and PJ and positive coefficients between ANIM and WILL. Note also that the standardization option results in negative coefficients for all formative indicators and, hence, the signs of the paths between ANIM and AFFECT as well as WILL are reversed.

Problems of Nonconvergence

For the scaling method of fixing one formative indicator path to animosity to 1, we had four alternative options of implementation, namely choosing one of the four different sources of (economic, political, war, and cultural) animosity as marker variable. As Table 1 (column 1) reports, three of these four implementation options resulted in nonconvergence of the model; only the model using war animosity as the marker variable converged (see column 2). The other two scaling methods (see columns 3 and 4 in Table 1) did not cause any convergence problems.

Model Fit

Focusing on those cases where convergence was attained, as Table 1 shows, the choice of scaling method does not seem to influence goodness of fit. All fit statistics (χ^2, RMSEA, NNFI, CFI, and SRMR) are identical across the three scaling options and show good fit for the expanded animosity model earlier presented in Fig. 5.

Variance Explained

As was the case with model fit, the explained variances of the endogenous variables in the model are invariant across scaling methods for all constructs (Table 1). Hence, the choice of scaling method does not affect the explanatory power of the theoretical model. The exogenous construct, consumer animosity, was well explained by the four formative indicators ($R^2 = 0.52$).[5] Satisfactory R^2 values were also attained for country affect ($R^2 = 0.88$), product judgment ($R^2 = 0.36$), and willingness to buy ($R^2 = 0.43$).

Parameter Estimates

The parameter estimates obtained for the measurement paths from the formative indicators sources to the animosity construct showed to be inconsistent across scaling methods (Table 1). First, the parameter values differ substantially across methods (for an example, see the estimates for the path economic animosity to animosity). However, closer investigation reveals that the raw parameter estimates across different scalings differ by a constant factor; for example, one can derive the measurement path coefficients in the third column of Table 1 (Path from ANIM to AFFECT fixed to 1) by multiplying the values of the corresponding coefficients in the first column (Path from WAR fixed to 1) by 0.102.[6] Having said that, no such constant relationship can be discerned between the values of the standard errors of the parameters concerned and this has important implications for the relevant significance tests as discussed later.

Second, and more importantly, using a critical *t*-value of 1.96 (corresponding to a 5% significance level), the number of significant indicator paths varies across scaling methods. Specifically, the paths from economic as well as cultural animosity to animosity are significant when using two scaling methods while nonsignificant when using the third (i.e., fixing the path from war animosity to animosity). This result is most probably the outcome of instability in standard errors noted earlier and confirms previous findings by Franke et al. (2008) and Diamantopoulos (2011) indicating that the choice of scaling may influence a researcher's conclusions regarding the validity of different formative indicators (e.g., economic animosity would be considered a valid indicator when standardization is opted for but not when war animosity is used as a marker variable).

As far as structural parameters are concerned, the paths AFFECT to PJ as well as PJ to WILL show to be stable across all three scaling methods (Table 1); hence the choice of scaling option for the formative latent variable does not affect parameter estimates and associated standard errors for relationships between endogenous variables. In contrast, focusing on the paths between the exogenous and the endogenous variables (i.e., ANIM to AFFECT and ANIM to WILL) reveals unstable results across scaling methods. For example, the path from consumer animosity to willingness to pay is not significant (at $p<0.05$) when war animosity is used as a marker variable (see second column in Table 1) but highly significant ($p<0.001$) when the variance of animosity is standardized. Given that the path involved is of central theoretical importance in the model in Fig. 5, different substantive conclusions would be drawn simply because of the choice of scaling method.

DISCUSSION AND CONCLUSIONS

Although formative measurement has increasingly been adopted in international marketing literature, little is known about possible consequences on theory testing due to re-specifications of reflective measurement models into formative models. In this chapter we sought to induce investigation and discussion of such consequences by highlighting a number of critical observations when empirically using formative measurement models. For this purpose, a structural model testing the effects of consumer animosity was used as an illustrative example.

Our analysis demonstrates several critical points. To begin with, due to problems of under-identification inherent to formative measurement, (even simple) theoretical models might need to be modified. This modification should not only focus on solving the identification problem, but also on the theoretical soundness of model extensions or changes. In this context, it is important to recognize the need for modifying the original model *before* data collection in order to collect data on any *additional* constructs that might be required. In our illustrative example, had we simply engaged in data collection for the constructs incorporated in the original animosity model (see Fig. 1), we would have faced major problems at the estimation stage. This is because the only model we could have estimated with a formative specification of the animosity construct is the model in Fig. 3 (which, as previously noted is theoretically untenable). To estimate our final model (see Fig. 5), data on the additional construct (country affect) needed to be collected.

Our findings also confirm previous research (Diamantopoulos, 2011; Franke et al., 2008) showing that the results derived from models incorporating formatively measured constructs may vary substantially depending on the scaling option used. First, some scaling options result in nonconvergence problems; this happens when formative indicators are used as "marker" (i.e., reference) variables. Second, assuming that convergence is attained, the standard errors of parameter estimates and associated significance levels of formative indicators depend on the choice of scaling. Given that the meaning of a formatively measured construct is inextricably linked to its indicators (Bollen & Lennox, 1991; Diamantopoulos & Winklhofer, 2001), this means that the substantive interpretation of a formatively measured construct can vary even within the *same* model.[7] Third, the interpretation of structural relationships may also be affected by the choice of scaling method. In our illustrative example, results on the hypothesized relationship between animosity and willingness to buy make the consequences of alternative scaling options most visible. Using three alternative options, two lead to a negative significant relationship (which corresponds to previous findings on animosity) and one to a nonsignificant relationship (hence contradicting previous findings based on reflective operationalizations of animosity). Consequently, different researchers estimating exactly the same model but using differing scaling options would draw very different substantial conclusions regarding the extent to which animosity impacts purchase decisions or not. Fourth, this situation is aggravated by the fact that all models yield identical goodness-of-fit statistics, regardless of the scaling option used. In other words, model fit does *not* provide any diagnostic value for the appropriateness of the obtained results. Thus if different researchers were to use different scaling approaches and obtained divergent results, all of them would be able to justify their model (and conclusions) based on acceptable model fit!

In light of the above, an obvious question becomes "where do we go from here"? One option for bypassing scaling problems is to adopt variance-based modeling methodologies (e.g., PLS) rather than covariance-based methodologies (e.g., LISREL). Of course, doing so comes at a price since the latter allow to *test* for the appropriateness of the specified model, whereas PLS assumes that the model has been correctly specified in the first place. Another alternative is to avoid using scaling options that appear to be particularly problematic. In this context, it has been observed that "using a formative indicator as a reference variable for scaling purposes can result in questionable estimates and/or standard errors" (Diamantopoulos, 2011, p. 350) and that "standardizing the formative construct and the observed variables, then interpreting the traditional Wald tests reported by most programs for structural equation modeling, appears to be a promising alternative" (Franke et al., 2008, p. 1236). Needless to say that future research should engage in simulation studies on the effects of alternative scaling options on convergence, parameter estimates, and standard errors in order to identify any systematic patterns of biases and provide an unambiguous basis for deciding in favor of one or another scaling option.

In conclusion, incorporating formatively measured constructs in theoretical models presents several challenges. It is hoped that the issues highlighted in this chapter will assist international marketing researchers in better dealing with them.

NOTES

1. The differences of reflective and formative measurement models have been extensively discussed in previous literature and need not be reiterated here (for a review see Diamantopoulos, Riefler and Roth (2008) and references given therein). Reflective measurement models have a long tradition in social sciences and are directly based on classical test theory (Lord & Novick, 1968). In these models, causality is specified from the construct to the measures (Bollen & Lennox, 1991; Edwards & Bagozzi, 2000) and measurement items are (necessarily) positively intercorrelated (Bollen, 1984). Formative measurement models, in contrast, specify causality from the measures to the construct and, being exogenous variables,

measurement items are not "forced" to display any specific intercorrelation pattern (Diamantopoulos & Winklhofer, 2001).

2. Klein et al. (1998) also include "consumer ethnocentrism" as a third predictor constructs impacting both product judgment and willingness to buy; however, as this is not a core part of the model (it does not moderate or mediate the impact of animosity), it has been excluded from Fig. 1 for simplicity purposes. The same applies to the "product ownership" construct as an outcome of willingness to buy.

3. As the animosity items were originally developed in German (see pre-study), translation was not an issue.

4. When scaling animosity, we did not consider the option of setting the path to willingness to pay equal to 1 as this represents a key theoretical relationship for testing. In principle, however, a solution could have also been obtained with this scaling option.

5. Note, however, that the amount of explained variance in animosity is not solely a function of the included formative indicators but also of the specific outcome constructs included in the model (Diamantopoulos, 2006; Franke et al., 2008).

6. The authors would like to thank Armin Monecke for pointing out the relationship among values of the unstandardized coefficients under different scalings.

7. Note that this is a different issue from the "context specificity" issue associated with formative models, whereby the estimates for the formative indicators depend on the specific outcomes variables included in the model (see Kim, Shin, & Grover, 2010; Wilcox, Howell, & Breivik, 2008).

REFERENCES

Albers, S., & Hildebrandt, L. (2006). Methodische Probleme bei der Erfolgsfaktorenforschung – Messfehler, formative versus reflektive Indikatoren und die Wahl des Strukturgleichungs-Modells. *Zeitschrift für betriebswirtschaftliche Forschung*, *58*(February), 2–33.

Bollen, K. A. (1984). Multiple indicators: Internal consistency or no necessary relationship? *Quality and Quantity*, *18*, 377–385.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Bollen, K. A., & Davis, W. R. (2009). Causal indicator models: Identification, estimation, and testing. *Structural Equation Modeling*, *16*, 498–522.

Bollen, K. A., & Lennox, R. (1991). Conventional widsom on measurement: A structural equation perspective. *Psychological Bulletin*, *110*(2), 305–314.

Brewer, P. A. (2007). Operationalizing psychic distance: A revised approach. *Journal of International Marketing*, *15*(1), 44–66.

Burger, K. (2009). Country affect: Scale development and validation. Diploma thesis, University of Vienna, Vienna.

Cenfetelli, R. T., & Bessellier, G. (2009). Interpretation of formative measurement in information systems research. *MIS Quarterly*, *33*(4), 689–708.

Craig, C. S., & Douglas, S. P. (2005). *International marketing research* (3rd ed.). Chichester, UK: Wiley.

Darling, J. R., & Arnold, D. R. (1988). The competitive position abroad of products and marketing practices of the United States, Japan and Selected European Countries. *Journal of Consumer Marketing*, *5*(Fall), 61–68.

DeVellis, R. F. (2003). *Scale development: Theory and applications. (2nd ed.) Applied social research methods series Vol. 26*. Thousand Oaks, CA: Sage.

Diamantopoulos, A. (1999). Export performance measurement: Reflective versus formative indicators. *International Marketing Review*, *16*(6), 444–457.

Diamantopoulos, A. (2006). The error term in formative measurement models: Interpretation and modeling implications. *Journal of Modelling in Management*, *1*(1), 7–17.

Diamantopoulos, A. (2011). Incorporating formative measures into covariance-based structural equation models. *MIS Quarterly*, *35*(2), 335–358.

Diamantopoulos, A., & Papadopoulos, N. (2010). Assessing the cross-national invariance of formative measures: Guidelines for international business researchers. *Journal of International Business Studies*, *41*(2), 360–370.

Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research*, *61*(12), 1203–1218.

Diamantopoulos, A., & Siguaw, J. A. (2006). Formative versus reflective indicators in organizational measure development: A comparison and empirical illustration. *British Journal of Management*, *17*, 263–282.

Diamantopoulos, A., & Winklhofer, H. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research*, *38*(2), 269–277.

Edwards, J. R. (2001). Multidimensional constructs in organizational behavior research: An integrative analytical framework. *Organizational Research Methods*, *4*(2), 144–192.

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, *5*(2), 155–174.

Ettenson, R. E., & Klein, J. G. (2005). The fallout from French Nuclear Testing in the South Pacific. *International Marketing Review*, *22*(2), 199–224.

Franke, G. R., Preacher, K. J., & Rigdon, E. E. (2008). Proportional structural effects of formative indicators. *Journal of Business Research*, *61*(12), 1229–1237.

Funk, C. A., Arthurs, J. D., Treviño, L. J., & Joireman, J. J. (2010). Consumer animosity in the global value chain: The effect of international production shifts on willingness to purchase hybrid products. *Journal of International Business Studies*, *41*, 639–651.

Gonzalez, R., & Griffin, D. (2001). Testing parameters in structural equation modeling: Every "one" matters. *Psychological Methods*, *6*, 258–269.

Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing. *Advances in International Marketing*, *20*, 277–319.

Holbrook, M. B. (1981). Integrating compositional and decompositional analyses to represent the intervening role of perceptions in evaluative judgments. *Journal of Marketing Research*, *18*(February), 13–28.

Huang, Y.-A., Phau, I., & Lin, C. (2010). Consumer animosity, economic hardship, and normative influence: How do they affect consumers' purchase intentions? *European Journal of Marketing*, *44*(7/8), 909–937.

Jaffe, E. D., & Nebenzahl, I. D. (2006). *National image and competitive advantage: The theory and practice of place branding* (2nd ed.). Copenhagen: Copenhagen Business School Press.

Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research*, *30*(2), 199–218.

Jiménez, N. H., & San Martin, S. (2010). The role of country-of-origin, ethnocentrism and animosity in promoting consumer trust: The moderating role of familiarity. *International Business Review*, *19*(1), 34–45.

Johansson, J. K., & Yip, G. S. (1994). Exploiting globalization potential: U.S. and Japanese strategies. *Strategic Management Journal*, *15*(8), 579–601.

Kim, G., Shin, B., & Grover, V. (2010). Investigating two contradictory views of formative measurement in information systems research. *MIS Quarterly*, *34*(2), 345–365.

Klein, J. G. (2002). Us versus them, or us versus everyone? Delineating consumer aversion to foreign goods. *Journal of International Business Studies*, *33*(2), 345–363.

Klein, J., & Ettenson, R. (1999). Consumer animosity and consumer ethnocentrism: An analysis of unique antecedents. *Journal of International Consumer Marketing*, *11*(4), 5–24.

Klein, J. G., Ettenson, R., & Morris, M. D. (1998). The animosity model of foreign product purchase: An empirical test in the People's Republic of China. *Journal of Marketing*, *62*, 89–100.

Kleinbaum, D. G., Kupper, L. L., Muller, K. E., & Azhar, N. (1998). *Applied regression analysis and other multivariate methods*. Pacific Grove, CA: Duxbury Press.

Law, K. S., Wong, C.-S., & Mobley, W. H. (1998). Toward a taxonomy of multidimensional constructs. *Academy of Management Review*, *23*, 741–755.

Leong, S. M., Cote, J. A., Ang, S. H., Tan, S. J., Jung, K., Kau, A. K., & Pornpitakpan, C. (2008). Understanding consumer animosity in an international crisis: Nature, antecedents, and consequences. *Journal of International Business Studies*, *39*, 996–1009.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

MacCallum, R. C., & Browne, M. W. (1993). The use of causal indicators in covariance structure models: Some practical issues. *Psychological Bulletin*, *114*(3), 533–541.

Maher, A. A., & Mady, S. (2010). Animosity, subjective norms, and anticipated emotions during an international crisis. *International Marketing Review*, *27*(6), 630–651.

Nijssen, E., & Douglas, S. (2004). Examining the animosity model in a country with a high level of foreign trade. *International Journal of Research in Marketing*, *21*, 23–38.

Papadopoulos, N., & Heslop, L. A. (1993). *Product-country images: Impact and role in international marketing*. New York: International Business Press.

Petter, S., Straub, D., & Rai, A. (2007). Specifying formative constructs in information systems research. [Research Essay]. *MIS Quarterly*, *31*(4), 623–656.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879–903.

Podsakoff, N. P., Shen, W., & Podsakoff, P. M. (2006). The role of formative measurement models in strategic management research: review, critique, and implications for future research. *Research Methodology in Strategy and Management*, *3*, 197–252.

Reynolds, N., & Diamantopoulos, A. (1998). The effect of pretest method on error detection rates: Experimental evidence. *European Journal of Marketing*, *32*(5/6), 480–498.

Riefler, P., & Diamantopoulos, A. (2007). Consumer animosity: A literature review and a reconsideration of its measurement. *International Marketing Review*, *24*(1), 87–119.

Roth, K. P., & Diamantopoulos, A. (2009). Advancing the country image construct. *Journal of Business Research*, *62*, 726–740.

Shamdasani, P. N., Stanaland, A. J. S., & Tan, J. (2001). Location, location, location: Insights for advertising placement on the web. *Journal of Advertising Research*, *41*(4), 7–21.

Shin, M. (2001). The animosity model of foreign product purchase revisited: Does it work in Korea? *Journal of Empirical Generalisations in Marketing Science*, *6*, 6–14.

Temme, D. (2006). Die Spezifikation und Identifikation formativer Messmodelle der Marketingforschung in Kovarianzstrukturanalysen. *Marketing ZFP*, *28*(3), 183–196.

Wilcox, J. B., Howell, R. D., & Breivik, E. (2008). Questions about formative measurement. *Journal of Business Research*, *61*(12), 1219–1228.

Wilcox, R. R. (2005). *Introduction to robust estimation and hypothesis testing* (2nd ed.). San Diego, CA: Academic Press.

Wood, V. R., & Darling, J. R. (1993). The marketing challenges of the Newly Independent Republics: Product competitiveness in global markets. *Journal of International Marketing*, *1*(1), 77–102.

DIFFERENCE SCORES, ANALYSIS LEVELS, AND THE (MIS)INTERPRETATION OF CULTURAL DISTANCE

George R. Franke, John S. Hill, Jase Ramsey and R. Glenn Richey

ABSTRACT

Purpose – *Cultural distance (CD) reflects differences in cultural values across countries. Many studies have used CD to explain strategies and outcomes in international business practices, although often with limited success. This chapter demonstrates previously unrecognized problems with the conceptualization, analysis, and interpretation of CD measures and suggests methods for improvements in CD research.*

Design/methodology/approach – *Problems with traditional methods in CD research are demonstrated analytically and illustrated with correlation and regression analyses of secondary data. One analysis shows that individual cultural dimensions may provide alternative explanations for hypothesized effects of distance. Two other examples illustrate the incorrect conclusions that traditional analysis approaches may suggest.*

Measurement and Research Methods in International Marketing
Advances in International Marketing, Volume 22, 31–51

ISSN: 1474-7979/doi:10.1108/S1474-7979(2011)0000022005

Findings – *The difference scores that are implicit in measures of CD usually imply unrealistic constraints on relationships between variables. Analyzing CD at the level of organizations rather than countries exaggerates the available sample size and may result in inaccurate statistical tests.*

Research limitations/implications – *The empirical examples illustrate problems with methodology for CD research. They are not proposed as substantive, generalizable tests of hypotheses.*

Originality/value of the chapter – *This chapter provides original arguments to augment existing criticisms of CD research. It shows that findings from extant CD studies may not support the conclusions that have been reported in the literature. Future research should use methods that lead to correct interpretations of CD effects.*

Keywords: Culture; cultural distance; psychic distance; difference scores; multilevel analysis

Culture has long fascinated international marketing researchers. Cultural differences can create challenges when marketing abroad, such as choosing markets to enter, modifying or developing products for international markets, developing global promotional programs, and negotiating with foreign suppliers and customers. Greater differences may lead to greater problems so that understanding the cultural gaps between markets may help explain business decisions and outcomes (e.g., Johanson & Vahlne, 1977). However, early cultural frameworks, grounded in anthropology (e.g., Kluckhohn & Strodtbeck, 1961), were unwieldy and difficult to apply in international business research. Hofstede's (1980, 2001) contributions to quantifying cultural values suggested a way to measure differences between pairs of countries. Using Hofstede's cultural indexes, Kogut and Singh (1988) introduced a formula for combining data on multiple cultural differences into a composite measure of cultural distance (CD). Thus, "by offering a seemingly simple and standardized measure of culture differences, the CD construct offered a tangible and convenient tool with which to bypass the complexities and intricacies of culture" (Shenkar, 2001, p. 519).

The Kogut–Singh formula has been used in more than 100 empirical studies (Håkanson & Ambos, 2010), and many alternative conceptualizations and formulas for CD have also been developed (e.g., Berry, Guillén, & Zhou, 2010; Dow & Karunaratna, 2006; Evans & Mavondo, 2002; Katsikeas, Skarmeas, & Bello, 2009). But in spite of the growing literature on CD, results have not lived up to expectations. Several meta-analyses of CD studies find small effects in general, with larger effects appearing only within particular categories of various moderator variables (Magnusson, Wilson, Zdravkovic, Zhou, & Westjohn, 2008; Tihanyi, Griffith, & Russell, 2005; Zhao, Luo, & Suh, 2004).

Previous authors have identified important limitations of the construct. Shenkar (2001) critiques conceptual and methodological aspects of CD in general as well as specific elements of Kogut and Singh's (1988) formula. Tung and Verbeke (2010) build on Shenkar's (2001) analysis to identify generic limitations, remediable weaknesses in research design, and weaknesses requiring reconceptualization of CD. Kirkman, Lowe, & Gibson (2006, p. 303) "strongly encourage researchers to *avoid further use of the overall [CD] index*" (italics in original). Harzing (2004) perceives the CD literature as inducing cultural myopia. Sarala and Vaara (2010) note criticisms of CD measures based on Hofstede's dimensions and measure CD using a broader and more recent set of nine cultural dimensions (House, Hanges, Javidan, Dorfman, & Gupta, 2004). Berry et al. (2010) recommend calculation of Mahalanobis distance that reflects covariances between multiple distance dimensions (e.g., economic distance, financial distance, and political distance).

With very few exceptions, extant criticisms of CD have not considered methodological problems that are inherent in the analysis of difference scores (e.g., Edwards, 1993, 1994, 2001; Griffin, Murray, & Gonzalez, 1999; Johns, 1981; Peter, Churchill, & Brown, 1993). They have also not fully considered the statistical implications of using individual- or company-level data to test country-level hypotheses (Sousa & Bradley, 2006). The number of countries, not the individual business practices or outcomes, determines the relevant sample size. Thus, CD research tends to vastly overstate the power of statistical tests, reporting, for example, a meta-analytic correlation of $-.036$ as significant at $p < .01$ (Magnusson et al., 2008).

The purpose of this study is to develop implications of difference scores and analysis levels for studying CD. The findings should make several contributions to the existing literature. In general, they should help in interpreting existing CD research and implementing future studies. More specifically, the results show that in some cases, tests of CD are indistinguishable from tests of cultural levels (i.e., not distances per se). Recognizing that CD measures involve difference scores also shows that tests of distance hypotheses may have a superficial appeal that disguises unacceptable underlying assumptions. Using the appropriate country-level sample sizes to interpret past research or conduct future research will give more statistically justifiable tests of statistical significance. Thus, the study shows that continuing with current practices in the CD literature would be inappropriate.

As a foundation of background material, the next section goes into more detail on the CD construct and its recognized shortcomings. Then, the chapter presents a more detailed analysis of difference scores, with an empirical illustration based on Chinese international travel statistics. Reanalyzing data from a study on equity joint ventures (EJVs) in China (Pan, 1996) illustrates the differences between country-level and individual-level statistical tests. An assessment of individual-level measures of distances, often called psychic distance (Håkanson & Ambos, 2010; Sousa & Bradley, 2006), shows that this approach is also subject to problems with the analysis of difference scores. The concluding sections include a discussion of the results, with suggestions for future research.

CULTURAL DISTANCE

Theoretical Foundations

Greater geographic distances between suppliers and customers can increase costs and inefficiencies in international business. The distance metaphor suggests remoteness and unfamiliarity, implying uncertainties and difficulties of doing business in foreign markets. Beckerman (1956) introduced the concept of "psychic distance" (subjective evaluations based on "fewer language difficulties, and so on," p. 38) as a complement to "economic distance" (transportation costs) in explaining patterns of trade between countries. Researchers at Uppsala University in Sweden treated psychic distance as "the sum of factors preventing the flow of information from and to the market, [including] differences in language, education, business practices, culture, and industrial development" (Johanson & Vahlne, 1977, p. 24). According to Kogut and Singh (1988, p. 430), "Cultural distance is, in most respects, similar to the 'psychic distance' used by the Uppsala school." However, subsequent practice commonly treats psychic distance as individuals' perceptions of differences between countries, with CD referring to composites of differences among national-level data (Håkanson & Ambos, 2010; Sousa & Bradley, 2006).[1]

Although the drawbacks of higher distances seem apparent, they do not necessarily lead to unambiguous predictions. Harzing (2004) discusses contradictory theoretical arguments for effects of CD on preferences for equity versus non-equity modes of market entry, for full versus shared control, and for acquisitions versus greenfields. Evans, Treadgold, and Mavondo (2000) suggest that greater distance may be a positive influence, as when an organization from a highly developed country has resources that give a greater competitive edge in developing countries. Wang and Schaan (2008) posit curvilinear (inverted-U) relationships between CD and both preference for joint ventures and performance in foreign markets. Brock, Shenkar, Shoham, and Siscovick (2008) argue that distance effects can be asymmetric, leading to different outcomes for the same pairs of countries depending on which is the home and which is the host. Magnusson et al. (2008) meta-analytically identify several moderators of CD effects across studies, including the effects of measurement approach (CD versus psychic distance) on estimated firm performance.

CD Measurement

The CD literature does not provide clear theoretical grounds for forming composite distance measures. Berry et al. (2010), for example, provide a thorough analysis of which national dimensions to consider and how to combine them to best satisfy various measurement criteria. However, they do not address the question of whether component variables should be combined into aggregate distance measures in the first place. This approach follows the precedent of Kogut and Singh (1988), which takes as given that composite measures are usable indexes of cultural differences.

Kogut and Singh (1988) introduced an index of CD based on four cultural dimensions from Hofstede (1980, 2001): power distance, uncertainty avoidance, individualism-collectivism, and masculinity-femininity:

$$\mathrm{CD}_j = \sum \frac{(I_{ij} - I_{iu})^2 / V_i}{4} \tag{1}$$

where I_{ij} and I_{iu} refer to values on the ith cultural dimension ($i = 1$–4) in country j and the United States (or any other baseline country), respectively, V_i is the variance of the ith dimension across countries, and CD_j is the cultural distance between country j and the United States. This index increases the weight given to larger differences on particular dimensions but reduces the weight given to dimensions with greater variability. As a composite of four underlying dimensions, similar CD scores between pairs of countries can arise from varied patterns of cultural differences in values and ways of thinking.

Kogut and Singh's formula is extensively used in CD research; Håkanson and Ambos (2010) report finding 148 empirical applications. Minor modifications to Eq. (1) include taking absolute values rather than squared differences (e.g., Grosse & Trevino, 1996) or using the square root of the calculated distance (e.g., Chang & Rosenzweig, 2001). A more elaborate modification that takes into account the covariances among dimensional variables shows a very high correlation of $r = .87$ between the new and original formulas (Berry et al., 2010). Thus, Eq. (1) may be considered both the prototypical and the predominant measure of CD in empirical research.

Criticisms of CD Research

Shenkar (2001) presents a detailed evaluation of "illusions" and assumptions in CD research. One illusion is that the symmetry of distance applies to cultural differences. Instead, as mentioned earlier, the effects of culture may differ for the home and host countries in a business relationship. The illusion of stability applies at the aggregate level – that is, cultures change over time and available measures may become outdated – and at the institutional level, such that apparent CD declines with increasing experience in a country. The illusion of discordance wrongly implies that all cultural differences are problematic, when in fact they "may be complementary and hence have a positive synergetic effect on investment and performance" (p. 524). This implication is similar to the "assumption of equivalence," which treats the four original Hofstede (1980) dimensions as equally important, while generally ignoring the subsequent dimension of long-term orientation (LTO) (Hofstede, 2001).

Shenkar's "illusion of linearity" is not an intrinsic problem in that researchers are free to treat CD in nonlinear ways (e.g., as implied by the squared difference terms in Kogut and Singh's, 1988, formula, or explicitly allowing nonlinear effects of CD as in Griffith & Myers, 2005). Shenkar strangely treats the hypothesis that CD has a causal influence on business decisions and performance as an illusion "that culture is the only determinant of distance with relevance ..." (p. 524). Certainly, other differences than the four cultural dimensions considered by Kogut and Singh (1988) may be relevant, and empirical research on CD incorporates error terms that recognize the existence of additional unmeasured causes. Shenkar's (2001) assumptions of corporate and spatial homogeneity also involve factors that are reflected in empirical error terms. That is, some organizations may respond differently to CD than others, just as an organization located near a border with another country may have a lower CD relative to that country than indicated by national-level data.

Tung and Verbeke (2010, p. 1262) discuss the relevance of Shenkar's (2001) comments "to broader applied work that takes on board cultural distance dimensions and measures." They emphasize that CD is not the same as psychic distance, which is actually more relevant to managerial decision-making. They describe problematic assumptions in CD research as "masks" that researchers hide behind to disguise the weaknesses of their methods. For example, combining individual distance measures into overall CD scores may conceal more than it reveals in terms of producing useful findings for international business.

One potential mask in distance research is that CD scores for different home or host countries may reflect very different cultural dimensions. This pattern is shown in Table 1, which illustrates the contribution of individual cultural dimensions to overall CD using Kogut and Singh's formula. For 66 baseline countries, distances relative to the other countries are correlated with those countries' scores on the Hofstede dimensions. Expanding Kogut and Singh's formula to include LTO reduces the available countries to 36. The table notes two LTO correlations, one for CD scores based on four dimensions and another based on all five. For simplicity, Table 1 notes detailed results for just the 20 largest countries by GDP (The World Factbook, 2009). Some correlations are so strong that CD can essentially be interpreted in terms of a single underlying cultural dimension. For example, CD largely reflects relative levels of individualism or collectivism for Australia, Indonesia, and the United States, with absolute correlations above .80.

Further drawbacks in CD studies may be hidden by any inappropriate analysis of difference scores. These problems are discussed next.

ANALYSIS OF DIFFERENCE SCORES

A fundamental problem with standard approaches to researching CD is that "*similarity is not a general quality. It is possible to discuss similarity only with respect to specified dimensions*" (Cronbach & Gleser, 1953, p. 457, italics in original). Thus, differences between countries in latitude and longitude can be combined into a single measure of geographic distance.[2] However, if the components of a CD index do not share a common meaning other than "distance," they should be unbundled rather than aggregated (Tung &

Table 1. Correlations between Culture Scores and Cultural Distance Scores.

Country	Power Distance[a]	Uncertainty Avoidance[a]	Individualism Collectivism[a]	Masculinity Femininity[a]	Long-Term Orientation[a]	Long-Term Orientation[b]
Australia	0.73	0.26	−0.83	−0.20	0.37	0.63
Belgium	0.03	−0.68	−0.23	−0.09	0.28	0.53
Brazil	−0.47	−0.60	0.53	0.09	−0.08	−0.23
Canada	0.74	0.30	−0.77	−0.04	0.32	0.69
China	−0.61	0.22	0.61	−0.32	−0.30	−0.80
France	−0.01	−0.65	−0.14	0.19	0.25	0.52
Germany	0.64	0.01	−0.63	−0.31	0.30	0.64
India	−0.44	0.26	0.28	−0.25	−0.22	−0.32
Indonesia	−0.67	−0.12	0.82	0.12	−0.37	–
Italy	0.39	−0.21	−0.57	−0.45	0.30	0.59
Japan	−0.15	−0.46	0.04	−0.72	0.12	−0.14
Korea	−0.40	−0.63	0.66	0.33	−0.11	−0.36
Mexico	−0.67	−0.55	0.59	−0.32	−0.17	–
Netherlands	0.66	0.13	−0.52	0.60	0.26	0.39
Russia	−0.65	−0.69	0.60	0.23	−0.11	–
Spain	−0.02	−0.69	0.13	0.30	0.09	0.64
Switzerland	0.65	0.14	−0.65	−0.39	0.29	0.54
Turkey	−0.39	−0.70	0.50	0.20	−0.06	–
UK	0.70	0.44	−0.79	−0.26	0.29	0.62
USA	0.69	0.32	−0.83	−0.23	0.35	0.64
Maximum absolute *r*, 66 countries	0.83	0.87	0.83	0.81	0.42	0.80
Mean absolute *r*, 66 countries	0.48	0.42	0.52	0.29	0.21	0.46

[a]Cultural distance scores based on four cultural dimensions (66 countries).
[b]Cultural distance scores based on five cultural dimensions including long-term orientation (36 countries).

Verbeke, 2010). Composite indexes of CD have ''face validity [that] is very low in some country comparisons'' (Harzing, 2004, p. 102). According to Eq. (1), for example, Brazil is culturally closer to Turkey, Bulgaria, and Iran than to any South American country.

Considering specific dimensions rather than composites of multiple dimensions does not avoid all problems with difference scores. A generic model with a single difference score as predictor illustrates some of the issues. Suppose a variable x_i (a cultural dimension in country i) is compared to a constant c (the value of the cultural dimension for a specific baseline country), as in

$$y = b_0 + b_1\{f(x_i - c)\} + e \tag{2}$$

where y is the dependent variable, b_0 is the intercept, b_1 is the regression coefficient, e is the error term, and f is a function representing alternative forms of differences (e.g., raw, squared, or absolute). A raw difference would be relevant when a higher (or lower) value is always better, as when institutions in one country prefer to invest in countries where economic development is higher or corruption is lower than at home. If only one home or host country is considered, which is the usual case in CD studies, in this model, *the value of the baseline country is irrelevant*. The intercept b_0 is affected by the value of c, but the correlation between x and y is the same regardless of c, or whether c is even included in the analysis (Griffin et al., 1999).

The function f in Eq. (2) could be absolute differences such that differences between x_i and c have equal effects on y regardless of whether the difference is positive or negative. If the baseline country is near an extreme on the cultural dimension, such as the masculinity of Japan or the uncertainty avoidance of Singapore, x_i in other countries will generally be above or below c, and including c in the model has little effect on the results (Edwards, 1994; Griffin et al., 1999). Generalizations to countries with less extreme values on the dimension would be ambiguous, because the analysis does little to distinguish between the effects of culture versus CD.

Because squared distance scores are common in CD research, Eq. (2) is worth expanding to indicate the effects of squaring differences:

$$y = b_0 + b_1(x_i - c)^2 + e = b_0 + b_1(x_i^2 - 2x_ic + c^2) + e \tag{3}$$

Because the squared term for the baseline country, c^2, is a constant, it is redundant with the intercept and has no effect on the variance explained by the model. More important conceptually, the model implies that the squared value of x_i has an effect on y that is reduced by exactly twice the value of x_i times the baseline culture's score. This hypothesis may be difficult to justify on theoretical grounds, although it is implicit in the squared-difference formulation. To test the constraints inherent in Eq. (3), its results should be compared with those from a less constrained model (cf. Edwards, 1994; Griffith & Myers, 2005):

$$y = b_0 + b_1x_i + b_2x_i^2 + e \tag{4}$$

The model in Eq. (3) is supported if it explains as much variance within sampling error as Eq. (4) and if in Eq. (4) b_1 is not significantly different from $-2cb_2$. The intercept in Eq. (4) should also equal $b_0 + b_1c^2$ from Eq. (3). If these constraints do not hold, the squared difference formulation in Eq. (3) should be rejected in favor of the more general formulation in Eq. (4). That is, a significant coefficient for a difference score analyzed as in Eq. (3) does not by itself justify the difference formulation.

To illustrate the alternative models, Table 2 summarizes regression results based on Chinese international passenger travel to 14 other countries (ETC Market Intelligence Group, 2007). Tourists may prefer to fly to culturally similar countries, such that CD is negatively related to destination choice (Ng, Lee, & Soutar, 2007). For simplicity, CD is calculated from a single cultural dimension, LTO, that is, $CD_{LTO} = (LTO_i - LTO_c)^2$, where i refers to country i and c refers to China. (Given the use of a single cultural dimension, dividing by the variance of LTO is irrelevant.) Table 2 summarizes the regression of passenger counts on CD_{LTO}, which despite the small sample size produces a significant ($p<.04$), negative standardized coefficient of $-.57$. Regressing passenger counts on other countries' LTO and LTO squared, as implied by Eq. (4), is significant overall ($p<.002$) and for each predictor (LTO, $p<.03$; LTO squared, $p<.01$). Restricting the regression coefficient for LTO to equal -2 times China's LTO of 118 times the regression coefficient for LTO squared (i.e., $b_1 = -2cb_2$ in Eq. (4)) produces exactly the same R^2 and significance levels as the CD_{LTO} model, showing their equivalence. However, the constraint significantly worsens the

Table 2. Constrained and Unconstrained Models of Chinese International Travel.

Term	CD Model		Constrained Model[a]		Unconstrained Model	
	b	p	b	p	b	p
Intercept	2,894.95	.01	−1,985.94	.15	4,204.79	.05
CD_{LTO}	−.35	.03				
LTO			82.73	.03	−212.23	.03
LTO squared			−.35	.03	2.37	.01
R^2	.32		.32		.69	
F	5.70	.03	5.70	.03	12.35	.002
Model d.f.	1		1		2	
Error d.f.	12		12		11	

[a]Constrained so that $b(\text{LTO}) = -2 \times 118 \times b(\text{LTO squared})$, where b is the unstandardized regression coefficient and 118 is China's LTO score. The fit is significantly worse than the unconstrained model, $t = 3.63$, $p<.004$.

fit of the model ($p < .004$) relative to the unconstrained model. Thus, although the CD_{LTO} model appears to explain Chinese travel patterns, the constraints implied by the distance formulation are not consistent with the data.[3]

Expanding Eq. (3) to include a second distance measure, and assuming for simplicity that only a single baseline country is considered in a given analysis, reveals further theoretical difficulties with CD measurement:

$$\begin{aligned} y &= b_0 + b_1\{(x_{1i} - c_1)^2 + (x_{2i} - c_2)^2\} + e \\ &= b_0 + b_1(x_{1i}^2 - 2x_{1i}c_1 + c_1^2 + x_{2i}^2 - 2x_{2i}c_2 + c_2^2) + e \end{aligned} \tag{5}$$

As with Eq. (3), the squared terms for the baseline country influence the intercept, and the effect of each cultural dimension is reduced by twice its product with the baseline culture's score. An important additional constraint is that the squared values of the cultural dimensions are weighted equally (i.e., b_1 is the coefficient for both dimensions). This constraint is unrealistic if some cultural dimensions, such as uncertainty avoidance, are more relevant than others (e.g., Kogut & Singh, 1988). Rather than hiding the constraint within the CD calculation, researchers should test it against the more general model:

$$y = b_0 + b_1x_{1i} + b_2x_{1i}^2 + b_3x_{2i} + b_4x_{2i}^2 + e \tag{6}$$

More constraints are implied as more dimensions are used in CD calculations such as the four used by Kogut and Singh (1988) or the nine used by Sarala and Vaara (2010). Considering multiple baseline countries simultaneously, or weighting the regression coefficients by the inverses of the variances of the cultural dimensions (Kogut & Singh, 1988), complicates the formula of the more general model but does not change the earlier analysis in any fundamental way.[4]

Testing CD effects with models such as Eq. (6) addresses several of the limitations discussed by Shenkar (2001) and Tung and Verbeke (2010). The model reveals the effects of individual cultural dimensions and allows for both nonlinear relationships and positive effects of CD. The major drawback of this approach is that high correlations between cultural dimensions and between cultural scores and their squared values will often lead to inferential errors. Collinearity effects hamper estimation of standard errors except when sample sizes are large and the model explains high amounts of variance in the dependent variable (Mason & Perreault, 1991). Because the relevant sample size for CD research involves countries, typical sample sizes in international business research may not allow reliable estimates of the individual regression coefficients (cf. Franke & Richey, 2010; Hult et al., 2008).

ANALYSIS LEVELS FOR DISTANCE EFFECTS

Focusing on country-level analyses is inconsistent with common research practices on CD, but it follows from Hofstede's (2001, pp. 15–17) discussion of ecological or between-society correlations as opposed to within-society correlations. For example, examining ownership preferences of businesses in one country relative to another would provide a single test of CD effects, regardless of how many individual acquisitions, joint ventures, or greenfield investments are considered. Larger samples provide a stronger basis for comparing those countries, but a very limited basis for generalizing to other countries based on CD scores (Franke & Richey, 2010).

Multilevel models provide a means for analyzing country-level and business-level effects simultaneously. Van de Vijver, Van Hemert, and Poortinga (2008) describe three defining characteristics of multilevel models. One is that they deal with phenomena at two or more levels such as cultures and organizations. Another is hierarchical structure with varying degrees of proximal and distal effects such as cultural (distal) and organizational (proximal) influences on management decisions and practices. The third characteristic involves the use of intrinsic or derived measures. Intrinsic variables are used at their natural level in the hierarchy, such as organizational experience in a foreign country influencing future entry mode decisions. Derived variables are collected at one level and used at another. Derived scores can be disaggregated from higher levels to lower, such as culture scores used to explain organizational practices, or aggregated from lower levels to higher, such as data from MBA students and alumni used to represent national levels of psychic distance (Håkanson & Ambos, 2010).

Correct multilevel analysis of CD effects would take into account the sample sizes and relationships between observational units at different levels such as firms within industries and industries within countries. However, standard practice in CD research is to use derived data from the cultural level with intrinsic data at the organizational level to explain such outcomes as firm entry decisions. Kogut and Singh's (1988) seminal study of CD effects with a sample size of 228 firm-level decisions is substantially larger than the sample of 13 regions (11 countries, Scandinavia, and other) that they actually compared. Even Luo's (2001) explicitly multilevel analysis tests

CD effects with firm-level decisions. This approach is likely to produce incorrect estimates of degrees of freedom and significance levels (e.g., Galwey, 2006). This approach also leads to "serious sample imbalances" (Harzing, 2004, p. 91), in which far more observations are available from some countries than others. For example, Kogut and Singh show data on more than 100 entry mode choices for Japan and the United Kingdom versus two for Malaysia and just one for South Africa.

Aggregating lower-level data to make country-level inferences depends on generalizable samples within countries. Unfortunately, CD research often relies on small samples within particular countries. For example, Dow and Ferencikova (2010) hypothesize that increasing CD increases the use of joint ventures over wholly owned subsidiaries and examine patterns of entry modes into Slovakia from 24 different home countries. Analyzing 154 entry decisions gives an average of fewer than 6.5 decisions per country. The 95% confidence intervals around the observed proportions are substantial, with ranges of .50 or greater for 75% of the countries. The high overlap of intervals across countries suggests that any CD effects across countries will be difficult to detect. That is, even if entry decisions are measured perfectly accurately in a specific time period, they may not be reliable indicators of national tendencies. Such random error attenuates relationships between variables and may have contributed to the limited CD effects observed by Dow and Ferencikova (2010). Thus, analyzing small numbers of countries, or small numbers of observations within particular countries, hampers CD research regardless of how many observations are available in total.

Pan's (1996) study on EJVs in China illustrates the effects of testing country-level effects at a lower level of analysis. As noted in Table 3, Hong Kong, Japan, and the United States accounted for 2,732, 338, and 549 Chinese EJVs, respectively. The ventures are summarized as minority, 50%, and majority foreign equity. Pan (1996) hypothesizes that levels of equity ownership are inversely related to CD, such that equity positions should be highest in Hong Kong, lowest in the United States, and intermediate in Japan. Given the ordinal measure of equity levels, a Spearman rank-order correlation is a reasonable way of evaluating the relationship between equity levels and CD. Treating the 3,618 EJVs as independent observations, the correlation is highly significant ($p < .001$) but positive ($r = .096$), contrary to the hypothesis. However, regardless of how many EJVs are considered, the country-level data provide just three tests of distance effects: Hong Kong versus Japan versus the United States. The Japan–US comparison is consistent with Pan's hypothesis, whereas the comparisons of Japan and the United States with Hong Kong are not. Correlating CD with the percent of

Table 3. US, Japan, and Hong Kong Equity Levels in Chinese Joint Ventures[a].

Origin	< 50% Foreign Equity		50% Foreign Equity		> 50% Foreign Equity	
	n	%	*n*	%	*n*	%
US	306	55.7	148	26.9	95	17.4
Japan	151	44.6	122	36.0	65	19.3
Hong Kong	1,795	65.7	462	16.9	475	17.4

[a]Data shown are adapted from Pan (1996). Chi-square (4 d.f.) = 95.2, $p < .001$.

nonminority EJVs produces an r of .81, which may appear impressive but which is nonsignificant ($p > .39$) given the n of just three countries.

Individual-level measures of perceived CD are appropriate for individual-level explanations of international business decisions (Sousa & Bradley, 2006). However, as discussed next, this measurement approach does not necessarily solve the problem of difference scores observed with the CD construct.

IS PSYCHIC DISTANCE THE ANSWER?

Johanson and Vahlne (1977) and some subsequent researchers conceptualize psychic distance in terms of factors that might influence perceptions and behaviors (e.g., Brewer, 2007; Dow & Ferencikova, 2010). More commonly, though, psychic distance is defined in terms of subjective perceptions of differences between countries (Evans & Mavondo, 2002; Håkanson & Ambos, 2010; Sousa & Bradley, 2006). The objective differences, or ''psychic distance stimuli'' (Dow & Karunaratna, 2006), logically influence the perceived differences. However, it is the perceptions that are ''really the key parameter affecting many managerial choices in an IB [international business] context'' (Tung & Verbeke, 2010, p. 1265).

One method of measuring psychic distance is to ask respondents to rate how similar or different other countries are relative to the respondents' own country on five-point or seven-point scales (e.g., Evans & Mavondo, 2002; Katsikeas et al., 2009; Sousa & Bradley, 2006). Another approach is to have respondents rate countries on a 100-point scale, with 0 or 1 for the home country, 100 for the most psychically distant country, and other countries rated relative to the maximum distance (e.g., Ellis, 2007; Håkanson & Ambos, 2010). Respondents can provide an overall score or rate differences on specific dimensions that may be combined into a composite score.

Does this measurement approach avoid the methodological problems discussed earlier? Unfortunately, it does not (Edwards, 1994, 2001). Any overall score conceals the underlying dimensions and their contributions to relationships between antecedents and consequences of distance perceptions, compromising interpretation of psychic distance findings. Comparing home and foreign countries implies an underlying model similar to Eq. (2), even if the baseline and alternative scores are not elicited directly. However, direct measures of perceived differences are not generally interchangeable with differences between the perceptions of the components being compared (e.g., Childress & Crompton, 1997; Parasuraman, Zeithaml, & Berry, 1994). Therefore, ''asking respondents to compare components may invoke cognitive processes other than the simple comparisons presumed in much congruence research'' (Edwards, 2001, p. 269), raising questions about what construct is being measured in psychic-distance research. If respondents explicitly rated home and foreign countries, researchers could statistically compare the implied model (Eq. (3)) with a more general version (Eq. (4)) to test whether the theoretical constraints are supported. The constraints cannot be tested when only a direct distance measure is available.

Peter et al. (1993) present a more positive view of having respondents directly indicate degrees of difference or similarity. They note that ''the direct comparison approach has the advantage of allowing [respondents] to combine their thoughts as they wish rather than have an arbitrary combination rule forced on them'' (p. 661). Unfortunately, researchers do not have access to the respondents' thoughts, so that the nature and effects of psychic distance are masked by the measurement operation. Thus, respondent-generated distance measures ''are prone to the same problems that plague difference scores because these problems do not depend on whether the respondent or the researcher calculates the difference'' (Edwards, 2001, p. 268).

DISCUSSION

Meta-analyses and narrative reviews reveal that CD has limited success in explaining decisions and outcomes of foreign entry modes (e.g., Harzing, 2004; Magnusson et al., 2008; Tihanyi et al., 2005). Regardless of the strength of the reported results, the problematic methodological foundations of CD make findings in this area hard to interpret in terms of differences between national cultures.[5] As a composite of difference scores, CD measures *cannot* contain more information than their separate components (Edwards, 1994; Griffin et al., 1999; Johns, 1981). Difference scores also imply equivalent hypotheses that researchers would rarely be able to justify. Analyzing CD at the wrong level almost inevitably weights some countries in the analysis more than others, provides inaccurate reports of sample sizes, and gives inappropriate indications of statistical significance. Thus, this study provides new arguments to reinforce the conclusion earlier reached by Shenkar (2001, p. 520):

> The appeal of the CD construct is, unfortunately, illusory. It masks serious problems in conceptualization and measurement, from unsupported hidden assumptions to questionable methodological properties, undermining the validity of the construct and challenging its theoretical role and application.

CD is an appealing metaphor in international business research (Smith, 2010) that can serve as an ''envelope concept'' into which other cultural concepts can be usefully folded (Tung & Verbeke, 2010, p. 1272). However, conventional research approaches to studying cultural and psychic distance, if not necessarily the underlying concept, are clearly ''past their due date'' (cf. Stöttinger & Schlegelmilch, 2000). At a minimum, researchers should correctly analyze *distance* measures as *differences* at the appropriate conceptual level. Testing models on individual cultural dimensions such as Eq. (4) allows for nonlinear effects without the unrealistic constraints imposed by difference-score specifications. Models with multiple cultural dimensions such as Eq. (6) indicate whether the effects differ in magnitude and direction, unlike conventional CD models. Theory-based or exploratory examinations of alternative distance specifications, such as nominal, absolute, and squared differences, may provide insights on directional and nonlinear effects of cultural differences. Testing a variety of dimensions, including religion, language, economic development, political systems, and so on, may reveal influences that go beyond the traditional value-based dimensions of Hofstede (1980, 2001) or GLOBE (House et al., 2004) (e.g., Berry et al., 2010; Brewer, 2007). Of course, studies must consider enough countries to provide justifiable generalizations, or else the findings must be interpreted with considerable caution (Franke & Richey, 2010).

Reassessing the extant empirical CD literature in terms of specific national dimensions, rather than general CD, may also suggest explanations for international business practices. Much of the literature uses the Kogut-Singh (1988) formula for calculating CD, which often gives distance scores that are highly correlated with individual cultural dimensions (Table 1). If countries such as China and the United States tend to show opposite patterns of results in CD studies, the differences might be due to the dimensions of power distance or individualism-collectivism, which show opposite signs in their correlations with the two countries' CD scores. Because national wealth tends to be correlated with or cause individualism (Hofstede, 2001, p. 253), countries whose CD scores are highly correlated with individualism may also suggest a role for income levels in explaining apparent CD effects.

Assessing perceptions of psychic distance to explain individual-level decisions avoids some but not all problems with CD research. Psychic distances implicitly or explicitly involve difference judgments that may be difficult to interpret. However, measuring and analyzing individual-level variables may increase effective sample sizes and avoid inference errors from inappropriate statistical tests. Obtaining individual-level data may also provide the opportunity to obtain useful qualitative insights. For example, Dyer and Chu (2011) note that "why" questions in interviewing respondents identified process-based trust as an important influence in international business relationships. Ellis's (2011) interviews with Chinese managers reveal that personal networks can not only generate but also constrain foreign opportunities. Ramsey's (2011) interviews with Brazilian CEOs reveal that in many cases, distance is not considered a relevant factor. According to one CEO, "The need to be in China is so critical that the fact that their cultural distance [from Brazil] is tremendous does not come into play."

If distance perceptions influence organizational practices, then identifying the antecedents of psychic distance could make an important contribution to international business research. For example, Håkanson and Ambos (2010) find that geographic distance is more strongly correlated than CD with country-level averages of psychic distance perceptions, using the Kogut-Singh (1988) formula for CD (with four cultural dimensions for 25 countries and also incorporating LTO for a subset of countries). Thus, the data for these correlations are based on a consistent level of analysis. However, the analysis does not show that the constraints implicit in the CD analysis are justified, suggesting the possibility that specific cultural dimensions might show stronger effects than the composite CD scores or perhaps geographic distance. Theoretical advances may also explain how antecedents of psychic distance may also be moderated at the individual level or the national level (cf. Stöttinger & Schlegelmilch, 1998).

CONCLUSION

Metaphors can clarify concepts or constrain ways of thinking. "Cultural distance" has become a popular metaphor, inspiring hundreds of studies as well as related metaphors such as psychic distance, institutional distance, knowledge distance, and technological distance (Smith, 2010). Unfortunately, CD research has distorted the underlying concept from distance as separation on specific dimensions, such as time or space, to distance as a composite of multiple aspects of separation. CD research has also emphasized one aspect of distance, the discord that differences may generate, with less acknowledgment of the clarity or synergies that may come with distance (Shenkar, 2001).

Operationally, much CD research has emphasized a single perspective on culture (Hofstede, 1980, 2001) and a single problematic method of quantifying distances by aggregating specific differences (Kogut & Singh, 1988). Broader interpretations of culture that include values and attitudes as well as elements such as language, religion, technology, politics, and social structures may better explain the environments that influence business decisions. However, no interpretation of culture or cultural distances is likely to give useful empirical findings when tested with misspecified models, at the wrong level of analysis.

NOTES

1. The term "institutional distance" is also used for objective distance measures, especially when multiple dimensions are used in addition to culture (e.g., Bae & Salomon, 2010; Berry et al., 2010).

2. Even measures of geographic distance call for conceptual justification, considering that the average distance may be very different from the greatest or smallest distance between two countries and that differences in latitude may have different national influences than differences in longitude (e.g., Diamond, 1999).

3. This example is presented as an empirical illustration rather than a substantive investigation of CD effects on tourism. The findings for China did not replicate in tests of travel statistics for 11 other countries.

4. General formulas for constrained and unconstrained versions of alternative distance models are given by Edwards (1993, 1994).

5. Problems of interpretation are also likely with country measures that use composite CD scores as a component, such as Luo and Shenkar's (2011) coefficient of cultural friction.

REFERENCES

Bae, J.-H., & Salomon, R. (2010). Institutional distance in international business research. *Advances in International Management*, *23*, 327–349.

Beckerman, W. (1956). Distance and the pattern of intra-European trade. *The Review of Economics and Statistics*, *38*(1), 31–40.

Berry, H., Guillén, M. F., & Zhou, N. (2010). An institutional approach to cross-national distance. *Journal of International Business Studies*, *41*(9), 1460–1480.

Brewer, P. A. (2007). Operationalizing psychic distance: A revised approach. *Journal of International Marketing*, *15*(1), 44–66.

Brock, D. M., Shenkar, O., Shoham, A., & Siscovick, I. C. (2008). National culture and expatriate deployment. *Journal of International Business Studies*, *39*(8), 1293–1309.

Chang, S., & Rosenzweig, P. W. (2001). The choice of entry mode in sequential foreign direct investment. *Strategic Management Journal*, *22*(8), 747–776.

Childress, R. D., & Crompton, J. L. (1997). A comparison of alternative direct and discrepancy approaches to measuring quality of performance at a festival. *Journal of Travel Research*, *36*(2), 43–57.

Cronbach, L. J., & Gleser, C. G. (1953). Assessing similarity between profiles. *Psychological Bulletin*, *50*(6), 456–473.

Diamond, J. (1999). *Guns, germs, and steel: The fates of human societies*. New York, NY: W.W. Norton & Company.

Dow, D., & Ferencikova, S. (2010). More than just national cultural distance: Testing new distance scales on FDI in Slovakia. *International Business Review*, *19*(1), 46–58.

Dow, D., & Karunaratna, A. (2006). Developing a multidimensional instrument to measure psychic distance stimuli. *Journal of International Business Studies*, *37*(5), 578–602.

Dyer, J., & Chu, W. (2011). The determinants of trust in supplier-automaker relations in the US, Japan, and Korea: A retrospective. *Journal of International Business Studies*, *42*(1), 28–34.

Edwards, J. R. (1993). Problems with the use of profile similarity indices in the study of congruence in organizational research. *Personnel Psychology*, *46*(3), 641–665.

Edwards, J. R. (1994). The study of congruence in organizational behavior research: Critique and a proposed alternative. *Organizational Behavior and Human Decision Processes*, *58*(1), 323–325.

Edwards, J. R. (2001). Ten difference score myths. *Organizational Research Methods*, *4*(3), 265–287.

Ellis, P. D. (2007). Paths to foreign markets: Does distance to market affect firm internationalization? *International Business Review*, *16*(5), 573–593.

Ellis, P. D. (2011). Social ties and international entrepreneurship: Opportunities and constraints affecting firm internationalization. *Journal of International Business Studies*, *42*(1), 99–127.

ETC Market Intelligence Group. (2007). Market insights: China, February. Retrieved from http://www.etc-corporate.org/resources/uploads/ETCProfile_China_6_07.pdf

Evans, J., & Mavondo, F. T. (2002). Psychic distance and organizational performance: An empirical examination of international retailing operations. *Journal of International Business Studies*, *33*(3), 515–532.

Evans, J., Treadgold, A., & Mavondo, F. (2000). Explaining export development through psychic distance. *International Marketing Review*, *17*(2), 164–168.

Franke, G. R., & Richey, R. G., Jr. (2010). Improving generalizations from multi-country comparisons in international business research. *Journal of International Business Studies*, *41*(8), 1275–1293.

Galwey, N. W. (2006). *Introduction to mixed modelling: Beyond regression and analysis of variance*. Chichester, UK: Wiley.

Griffin, D., Murray, S., & Gonzalez, R. (1999). Difference score correlations in relationship research: A conceptual primer. *Personal Relationships*, *6*(4), 505–518.

Griffith, D. A., & Myers, M. B. (2005). The performance implications of strategic fit of relational norm governance strategies in global supply chain relationships. *Journal of International Business Studies*, *36*(3), 254–269.

Grosse, R., & Trevino, L. J. (1996). Foreign direct investment in the United States: An analysis by country of origin. *Journal of International Business Studies*, *27*(1), 139–155.

Håkanson, L., & Ambos, B. (2010). The antecedents of psychic distance. *Journal of International Management*, *16*(3), 195–210.

Harzing, A. W. K. (2004). The role of culture in entry mode studies: From negligence to myopia? *Advances in International Management*, *15*, 75–127.

Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. Beverly Hills, CA: Sage.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations* (2nd ed.). Thousand Oaks, CA: Sage.

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (2004). *Culture, leadership, and organizations: The GLOBE study of 62 societies*. Thousand Oaks, CA: Sage.

Hult, G. T. M., Ketchen, D. J., Jr., Griffith, D. A., Finnegan, C. A., Gonzalez-Padron, T., Harmancioglu, N., & Cavusgil, S. T. (2008). Data equivalence in cross-cultural international business research: Assessment and guidelines. *Journal of International Business Studies*, *39*(6), 1027–1044.

Johanson, J., & Vahlne, J.-E. (1977). The internationalization process of the firm – A model of knowledge development and increasing foreign market commitments. *Journal of International Business Studies*, *8*(1), 23–32.

Johns, G. (1981). Difference score measures of organizational behavior variables: A critique. *Organizational Behavior and Human Performance*, *27*(3), 443–463.

Katsikeas, C., Skarmeas, D., & Bello, D. (2009). Developing successful trust-based international exchange relationships. *Journal of International Business Studies*, *40*(1), 132–155.

Kirkman, B. L., Lowe, K. B., & Gibson, C. B. (2006). A quarter century of *Culture's Consequences*: A review of empirical research incorporating Hofstede's cultural values framework. *Journal of International Business Studies, 37*(3), 285–320.

Kluckhohn, F. R., & Strodtbeck, F. L. (1961). *Variations in value orientations*. New York, NY: Row, Peterson & Company.

Kogut, B., & Singh, H. (1988). The effect of national culture on the choice of entry mode. *Journal of International Business Studies*, *19*(3), 411–432.

Luo, A. (2001). Determinants of entry in an emerging economy: A multilevel approach. *Journal of Management Studies*, *38*(3), 443–472.

Luo, A., & Shenkar, O. (2011). Toward a perspective of cultural friction in international business. *Journal of International Management*, *17*(1), 1–14.

Magnusson, P., Wilson, R., Zdravkovic, S., Zhou, J., & Westjohn, S. (2008). Breaking through the cultural clutter: A comparative assessment of multiple cultural and institutional frameworks. *International Marketing Review*, *25*(2), 183–201.

Mason, C. H., & Perreault, W. D. (1991). Collinearity, power, and interpretation of multiple regression analysis. *Journal of Marketing Research*, *28*(3), 268–280.

Ng, S. I., Lee, J. A., & Soutar, G. N. (2007). Tourists' intention to visit a country: The impact of cultural distance. *Tourism Management*, *28*(6), 1497–1506.

Pan, Y. (1996). Influences on foreign equity ownership level in joint ventures in China. *Journal of International Business Studies*, *27*(1), 1–26.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1994). Alternative scales for measuring service quality: A comparative assessment based on psychometric and diagnostic criteria. *Journal of Retailing*, *70*(3), 201–230.

Peter, J. P., Churchill, G. A., & Brown, T. J. (1993). Caution in the use of difference scores in consumer research. *Journal of Consumer Research*, *19*(1), 655–662.

Ramsey, J. (2011). *Brazilian multinationalization: Views from the top*. Unpublished working paper.

Sarala, R. M., & Vaara, E. (2010). Cultural differences, convergence, and crossvergence as explanations of knowledge transfer in international acquisitions. *Journal of International Business Studies*, *41*(8), 1365–1390.

Shenkar, O. (2001). Cultural distance revisited: Towards a more rigorous conceptualization and measurement of cultural differences. *Journal of International Business Studies*, *32*(3), 519–535.

Smith, B. (2010). Software, distance, friction, and more: A review of lessons and losses in the debate for a better metaphor on culture. *Advances in International Management*, *23*, 213–229.

Sousa, C. M. P., & Bradley, F. (2006). Cultural distance and psychic distance: Two peas in a pod? *Journal of International Marketing*, *14*(1), 49–70.

Stöttinger, B., & Schlegelmilch, B. B. (1998). Explaining export development through psychic distance: Enlightening or elusive? *International Marketing Review*, *15*(5), 357–372.

Stöttinger, B., & Schlegelmilch, B. B. (2000). Psychic distance: A concept past its due date? *International Marketing Review*, *17*(2), 169–173.

The World Factbook. (2009). Washington, DC: Central Intelligence Agency. Retrieved from http://www.cia.gov/library/publications/the-world-factbook/index.html

Tihanyi, L., Griffith, D. A., & Russell, C. J. (2005). The effect of cultural distance on entry mode choice, international diversification, and MNE performance: A meta-analysis. *Journal of International Business Studies*, *36*(3), 270–283.

Tung, R. L., & Verbeke, A. (2010). Beyond Hofstede and GLOBE: Improving the quality of cross-cultural research. *Journal of International Business Studies*, *41*(8), 1259–1274.

Van de Vijver, F. J. R., Van Hemert, D. A., & Poortinga, Y. H. (Eds.). (2008). *Multilevel analysis of individuals and cultures*. Mahwah, NJ: Psychology Press.

Wang, H., & Schaan, J-L. (2008). How much distance do we need? Revisiting the 'National Cultural Distance Paradox'. *Management International Review*, *48*(3), 263–277.

Zhao, H., Luo, Y., & Suh, T. (2004). Transaction cost determinants and ownership-based entry mode choice: A meta-analytical review. *Journal of International Business Studies*, *35*(6), 524–544.

THE ROLE OF RESPONSE FORMATS ON EXTREME RESPONSE STYLE: A CASE OF LIKERT-TYPE VS. SEMANTIC DIFFERENTIAL SCALES

Joseph F. Rocereto, Marina Puzakova,
Rolph E. Anderson and Hyokjin Kwak

ABSTRACT

Purpose – *A major limitation in cross-cultural research continues to be attempts to compare construct measurements across cultures without adequate conceptual and empirical evidence of the equivalency of the measurement scores. Of significant concern in such studies is the presence of various types of response bias that may systematically differ from one culture to another, resulting in a potential violation of the assumption that measurement scores across cultures are equivalent. The focus of this study is to investigate the role of the response format type, extreme response style (ERS). Most studies have investigated response bias styles using Likert-type scales as response formats, yet it has long been argued that these particular formats tend to result in various types of response style bias, especially in cross-cultural research. Would other scaling devices, such as semantic differential (SD), lessen response style bias in pan-cultural studies? To answer this question, our study employs two types of*

Measurement and Research Methods in International Marketing
Advances in International Marketing, Volume 22, 53–71

ISSN: 1474-7979/doi:10.1108/S1474-7979(2011)0000022006

response formats (i.e., Liker-type and SD) to empirically test for the presence of ERS within each response format style.

Methodology/approach – *This chapter takes the form of empirical research using ERS indices to test for the degree of ERS between response formats using samples from a collectivistic culture (i.e., South Korea) and an individualistic culture (i.e., United States).*

Findings – *Results show that samples from both cultures exhibit greater levels of ERS when using Likert-type scales compared to SD scales. Additionally, this study finds that, when using Likert-type scales, ERS is greater for U.S. respondents than for South Korea respondents. Finally, results show that there is no statistically significant difference in ERS between the two cultural groups when using SD response formats.*

Research implications – *Findings show that the use of SD response formats eliminates systematic differences in ERS between a collectivist sample and an individualist sample. Therefore, the use of such response formats in future cross-cultural research can greatly diminish the problematic effects of culturally based ERS and lead to greater confidence in the equivalency of measurement scores across cultures.*

Originality/value of paper – *This study is the first to simultaneously assess culturally based ERS using two types of response formats to investigate the impact of response format on ERS. Furthermore, this study assesses the role of response format on ERS both within and between two distinctly different cultures.*

Keywords: Cross-cultural research; extreme response style; response formats

INTRODUCTION

Investigations of marketing constructs and measures across cultures have multiplied in recent years with the accelerating globalization of markets, the growing power of multinational corporations, and the expansion of international workforces. Consequently, cross-cultural investigations are become increasingly critical to better understanding of marketing phenomena in different cultures. Rating scales are one of the most frequently employed tools in research (Chen, Lee, & Stevenson, 1995). When responding to survey research questionnaires, a person's observed score on the rating scale usually is a function of three components: a true score, systematic measurement error, and random measurement error. Over the years, systematic biases in responding to rating scales have been purported and reported in the literature (Berg, 1967; Cronbach, 1946, 1950; Couch & Keniston, 1960). For example, studies have consistently found that Asian cultures tend to avoid the extreme ends of rating scales as compared to Western cultures (Stevenson et al., 1990). This tendency toward the mid-point of scales is consistent with the Confucian philosophy that one should not stand out from the group, whereas Western cultures tend to more highly value and reward independent thought and action. Response styles are systematic ways of answering questions without directly relating to the question content, but instead exhibiting behavioral characteristics of the respondents themselves (Baumgartner & Steenkamp, 2001, 2006; Oskamp, 1977). Non-content-based forms of responding are usually referred to as response styles or response biases. Responses may often be influenced by such non-content factors as the rating scale itself (Cronbach, 1946; Lentz, 1938). Response styles are a potentially potent source of systematic measurement error and can jeopardize the validity of conclusions drawn from marketing research data (Bearden, Netemeyer, & Teel, 1989).

Major concerns have been expressed over the years about the possible contaminating influences of response biases in cross-national comparisons (Leung, 1989; Leung & Bond, 1989). Cultural differences in the use of rating scales for questionnaires have theoretical as well as methodological implications for cross-cultural research. In questionnaire research, response styles can be a source of contamination by either inflating or deflating respondents' scores on measurement instruments thereby making the validity of conclusions drawn from marketing research data questionable. Response styles can also affect conclusions about the relationships between scales by inflating or deflating the correlation between respondents' scores on measurement instruments. Thus, two types of contaminating effects of response styles must be considered (Bagozzi & Yi, 1993). Response styles may bias the assessment of true scores by inflating or deflating observed scale scores, which, in turn, may also bias the investigation of relationships between constructs by inflating or deflating a scale's correlations with other scales. In light of these potential problems, it is crucial to investigate the extent to which cross-cultural differences in ratings reflect true cultural differences or are simply the result of differences in response styles. In the present study, we investigate stylistic responding as a source of nonrandom measurement error or, more specifically, the systematic effects of particular response styles.

Response styles are a source of concern in both domestic and international marketing research because they threaten the validity of empirical findings by contaminating responses to questions (Craig & Douglas, 2000; Greenleaf, 1992a; Van de Vijver & Leung, 1997). However, response style bias is of particular concern in cross-cultural research because it threatens the validity of empirical findings by tainting evaluations of the constructs being studied (Baumgartner & Steenkamp, 2001; Craig & Douglas, 2000). Contamination of measures can invalidate the assumption of construct equivalency across cultures and lead to erroneous interpretations of attitudes or opinions by respondents from different cultures. International marketing researchers, therefore, are in jeopardy of interpreting findings as verifying significant differences between cultures when in fact there are none or in reporting construct invariance across cultures when differences truly exist.

Researchers have long been aware of varying forms of response style biases that can confound their findings, especially in cross-cultural studies (Cronbach, 1946; Lentz, 1938). More recently, Smith reaffirms that cross-cultural researchers who utilize questionnaires for data collection are cognizant of cultural variations in several types of response bias. In particular, when Likert-type scales are employed, consistent patterns are revealed among respondent categories in choosing points on the scales (Smith, 2004). Paulhus (1991, p. 17) defines a response bias as "a systematic tendency to respond to a range of questionnaire items on some basis other than the specific item content." Response style bias may yield responses that do not reflect the person's true scores, but ones influenced by systematic bias toward the measurement scale used. In more quantitative terms, Van de Vijver and Leung (1997) describe construct equivalence as the invariance of factor loadings of the indicators for the construct in question between groups (i.e., different cultures) under study. When types of response bias are relatively consistent within one culture, but differ in type from that of another culture, then the assumption of construct equivalency is likely to be violated. No matter their particular form, response style biases can result in misinterpretations and misunderstandings when comparing cross-cultural constructs for marketing decision making, especially in segmenting international markets.

Various researchers have investigated approaches to account for response bias when interpreting measurement scores (Chen et al., 1995; Cheung & Rensvold, 2000; Welkenhuysen-Gybels, Billiet, & Cambre, 2003), but few studies have dealt with the impact of response format styles on observed differences in response style bias between cultures. Moreover, in the extant literature, all these studies have used Likert-type response formats (Chen et al., 1995; Clarke III, 2001; Grimm & Church, 1999; Hui & Triandis, 1989; van Herk, Poortinga, & Verhallen, 2004; WelkenHuysen-Gybels et al., 2003). Accordingly, there is a pressing need for investigations into other response format styles, especially the use of the semantic differential (SD). Therefore, the primary purpose of our study is to assess the potential effects of two distinct response format styles (Likert-type and SD) on systematic differences in extreme response style (ERS) between what Hofstede (1980) defines as "collectivist" and "individualist" cultures.

LITERATURE REVIEW

Individualism–Collectivism (IND-COL)

Hofstede's (1980, 1991, 2001) classic nation-level cultural dimensions include five major categories: individualism–collectivism (IND-COL), power distance (PDI), uncertainty avoidance (UAI), masculinity–femininity (MAS), and long-term vs. short-term orientation (LTO). Of interest in this study is the IND-COL index. In individualistic cultures, people are expected to take care of themselves so connections between individuals are relatively loose. Individuals tend to prefer to act independently rather than as members of a group, and they seek autonomy and personal achievement (Oyserman, Coon, & Kemmelmeiser, 2002). Whereas people in collectivistic cultures see themselves as members of a larger cohesive group and tend to place higher priority on the group goals than their individual goals. The values of individualistic cultures tend to be power, achievement, and hedonism, while collectivistic cultures value tradition, conformity, and benevolence. Although Hofstede (1980) felt that these two cultures may be opposites, other researchers have found that people may retain both their independent and interdependent selves, and each may activate in different situations (Markus & Kitayama, 1991).

Response Style

Douglas and Craig (1983, p. 192) found a general "response style" tendency of people in the United States to be more extreme in their responses than Asian (Japanese) respondents. When one cultural group reveals a relatively conservative response style and another group a relatively ERS, direct statistical comparison of the scaling data may not be very meaningful. Ethnographic studies consistently show that Asian cultures place a greater emphasis on emotional moderation than do Western cultures. The desire for emotional moderation can be traced back to Confucian teachings, in which "the fundamental moral idea of moderation, balance, and subtleness" is emphasized (DeBary, Chart, & Watson, 1960, p. 117). Because social relationships are extremely important in Asian cultures, people try to avoid excessive emotions that might threaten, offend, or disturb the stability of existing relationships (Bond & Hwang, 1986; Chiu & Kosinski, 1994; Markus & Kitayama, 1994). Song (1985, p. 53) found that Chinese respondents were "more emotionally reserved, introverted, fond of tranquility, and habituated to self-restraint compared to Westerners." As affirmed by Koo (1976), Asian cultures believe that the ability to control one's emotions promotes personal health and interpersonal harmony (Chiu & Kosinski, 1994). Moreover, emotional control is necessary for harmonious interpersonal and group relationships. "Guns shoot the birds that come out" is a common expression of moderation in Asian cultures. Given the foregoing observations, we would expect to find evidence of moderate emotional behavior or even emotional suppression on the part of the Asian respondents to measurement scales, as compared with respondents from the United States. Extant literature on comparative methodology has rarely addressed the issue of response style bias, and appropriate practices and procedures have not been agreed upon (Aulakh & Kotabe, 1993; Clark III, 2001).

Extreme Response Style

Among the most commonly used scales in behavioral research are attitude scales designed to measure people's opinions. These scales generate numerical scores that are compared across potential markets. On the basis of the number of pertinent articles in the literature, the greatest concern in cross-cultural research appears to be ERS. ERS is the tendency to use the most extreme response categories of rating scales regardless of content (De Jong, Steenkamp, Fox, & Baumgartner, 2008; Greenleaf, 1992b; Johnson, 2003). Paulhus (1991) emphasizes the systematic nature of ERS in that respondents who exhibit ERS tend to do so across both time and stimuli. To illustrate, if high-ERS participants are given a survey using a 7-point Likert-type scale, their responses will tend to be either 1 (strongly agree) or 7 (strongly disagree), regardless of item content. Whereas, if low-ERS participants are given the same survey, their responses will tend to cluster around the mid-point of 4 (neither agree nor disagree). Several studies have documented cross-cultural differences in ERS (Greenleaf, 1992b; Hui & Triandis, 1985; Schaninger & Buss, 1986; Triandis, 1994). Lee and Green (1991), for instance, discerned that Koreans tend to avoid extreme or end-point responses and prefer the midpoints of scales. This preference for the mid-point by low-ERS cultures may reflect their desire to appear modest and nonjudgmental. In contrast, members of high-ERS cultures may wish to demonstrate independence, confidence, and conviction by more readily choosing end-point responses.

ERS differences can have many adverse effects on cross-cultural and international comparisons (Chun, Campbell, & Yoo, 1974). Because ERS affects numerical scores, comparisons of means may not be meaningful. At a more fundamental level, ERS differences produce noninvariant factor loadings and intercepts, leading to the conclusion that the numbers on the response scale mean different things to different groups of people. Additionally, standard deviation and variance are influenced by ERS, impacting the size of item intercorrelations (Hui & Triandis, 1989). Lastly, due to the consistent nature of ERS, it can result in an increase in reliability but a decrease in the validity of the test (Cronbach, 1946).

Extreme Response Style in Cross-Cultural Research

Globalization of markets has spurred growing dependence on cross-cultural research for marketing decision making. Consequently, cross-cultural researchers are more concerned than ever about potential bias in their research methodology. One especially relevant cross-cultural methodological issue is ERS (Samiee & Jeong, 1994). Development of multicultural markets has made understanding and accounting for this potential bias critical for cross-cultural market researchers. Cross-cultural quantitative research typically compares the responses from people in different cultures to various concepts of interest, but caution is needed in interpreting their responses because cultural differences can influence the pattern of responses (Albaum, Alpert, & Alpert, 1987a; Albaum, Golden, Murphy, & Strandskov, 1987b). To illustrate, if people from a designated culture are inclined to respond using only the extreme response categories of a scale (ERS), while other cultures exhibit a conservative response style, comparisons between the two cultures will be distorted. Therefore, a critical methodological query for cross-cultural research is whether or not response formats affect the validity of results. And, if so, which scaling formats are most likely to alleviate response style bias?

Attitude Measurement Scales

Most attitude measurement scales employed by marketing researchers are concerned with obtaining the responses of people (e.g., consumers, purchasing agents, marketing managers) regarding specific stimuli such as alternative products or services, advertisements, package designs, brand names, and the like. The two most often used scaling devices or measurement formats are: Likert-type and SD scales.

Likert-Type

Named after its developer, the Likert (1932) scale is a widely used rating scale that requires respondents to indicate their degree of agreement or disagreement with each of a series of statements expressing either a favorable or an unfavorable attitude toward the concept or object by assigning it a numerical score. Typically, each scale item has five or seven response categories, ranging from strongly disagree to strongly agree. Although Likert-type scales are not assumed to be intervally measured, they are usually treated as such by marketing researchers (Albaum, 1997). The scales purport to measure attitude direction (by agree or disagree) and intensity (by strongly agree or strongly disagree). Usually there is a midpoint response option that allows a neutral or undecided selection by respondents. Although the agree–disagree format is the most common form of Likert scale, other types of response end points also are used (e.g., very unmotivated; very motivated; below average; above average; and so forth). Because of their design, Likert-type formats ask individuals to think along at least two different dimensions – content and intensity. Respondents must decide first whether they agree or disagree with the content of each stated proposition. Then, they must determine their level of intensity regarding the proposition by indicating how strongly they feel (e.g., strongly agree or strongly disagree).

Semantic Differential

Another widely used and versatile scale for marketing research is the SD, developed by Osgood, May, and Miron (1975). This measurement instrument explores the *connotative meaning* or personal meaning of something, as distinct from its actual physical characteristics. SDs can be used to describe the connotative meaning of abstract concepts as people react to stimulus words by assigning ratings on bipolar scales with contrasting adjectives at each end. Using this scaling instrument, marketing researchers are able to measure both the direction and intensity of respondents' perceptions regarding various marketing phenomena like corporate image, advertisements, and brand image. Respondents consider sets of paired adjective antonyms with the extremes of each pair separated by seven intervals that are assumed to be equal.

Contrasting Likert-Type and Semantic Differential Scales

Researchers are in general agreement that the use of Likert-type scales increases the likelihood of various forms of response bias styles (Clarke III, 2001; Hui & Triandis, 1989; Smith, 2004). Specifically, Clarke III (2001) asserts that Likert-type scales have a strong tendency to increase levels of ERS. This may be due partially to the relatively lower level of involvement among respondents when using such scale format. Merely agreeing or disagreeing with a statement is comparatively simpler and less mentally taxing than efforts to rate a construct on various sets of bipolar adjectives as used in SD scales. Therefore, it is expected that the use of Likert-type scale formats will result in respondents thinking less about the content of the items, thus increasing the degree of ERS. Furthermore, while an extreme response on a Likert-type scale reflects an extreme general agreement or disagreement with a statement, an extreme response on a SD scale signifies that a respondent feels that the construct of interest completely possesses a perceptual or descriptive quality, such as wholly "trustworthy" or entirely "untrustworthy." It is unlikely that many respondents view any construct as possessing a particular characteristic in its entirety, so the extreme selections seems less likely on the SD scale.

However, the degree to which ERS increases due to the use of Likert-type scales for each group is not likely to be invariant. For example, differing cultural tendencies may result in a greater degree of ERS bias among individualist respondents than for collectivist respondents. Individualist cultures value independence and sincerity and, thus, have little unease about standing apart from the group. However, members of a collectivist society possess a desire to exhibit modesty and avoid standing out from the group (Cheung & Rensvold, 2000). Therefore, while the use of a Likert-type scale is likely to increase the rate of ERS bias in both samples, it is expected that this increase will be greater for members of an individualist culture such as the United States vis-à-vis a more collectivist society like South Korea.

It has been consistently shown that Asian cultures, including Korea, are more likely to avoid extreme responses and prefer the midpoint of scales compared to their Western counterparts (Cheung & Rensvold, 2000; Lee & Green, 1991). These predilections are likely caused by differing cultural norms that affect response styles (Grimm & Church, 1999; Guptara, Murray, Razak, & Sheehan, 1990; Hui & Triandis, 1989). Individuals in collectivist societies prefer to appear modest and nonjudgmental, while those in individualist cultures readily demonstrate sincerity and conviction (Cheung & Rensvold, 2000). A likely reason for the Asian respondent's desire to exhibit modesty is the Confucian philosophy that discourages standing out from the group (Chen et al., 1995). Additionally, Markus and Kitayama (1991) suggest that the tendency for Asians to prefer the midpoint of scales is related to the perceived importance of conformity and interdependency among individuals. Therefore, it appears that the importance that Asians place on the avoidance of manifest deviation from beliefs of the group is reflected in their response styles. It is this culturally based value system that is likely to increase the systematic manner in which Asians respond to survey items regardless of item content, particularly when exposed to specific types of response formats. Thus, we proffer:

H1a. When using Likert-type scale formats, both U.S. and Korean respondents will exhibit greater levels of ERS bias than when using SD scale formats.

H1b. When using Likert-type scale formats, U.S. respondents will exhibit greater levels of ERS bias than Korean respondents.

A review of the literature reveals that increases in ERS have not been attributed to use of SD scales in marketing research. Perhaps, the structure of the response formats for SD scales is more likely to require respondents to pay closer attention to item content, thus suppressing any automatic response tendency. Having to appraise each set of bipolar adjectives as scale anchors may cause respondents to more carefully consider each response option. Furthermore, these bipolar anchors associated with SD scales do not remain constant (as is the case of scale anchors used in Likert-type scales) but differ with each survey item. This lack of uniformity of anchor terms or phrases is likely to lead to greater involvement among respondents, resulting in more attention to and comprehension of item content. This enhanced involvement and attention to item content also may help suppress ERS bias when using SD scales, thereby leading to equally low levels of ERS bias between U.S. and Korean respondents. Thus, we hypothesize:

H2. When using SD scale formats, there will be no difference in ERS bias between U.S. and Korean respondents.

METHOD

Sample

Survey questionnaires were administered in the United States and South Korea. Questions to measure subjects' degree of consumer ethnocentrism and a foreign advertisement were included, along with typical demographic variables (e.g., age and gender) in the questionnaire. In the United States, a sample of 242 individuals ($N_{male} = 93$, 38.4%) was gathered using U.S.-born business undergraduate students at an American university. In Seoul, Korea, 205 business undergraduate students ($N_{male} = 132$, 64.4%) comprised the sample. The average age of the subjects was comparable in both countries (U.S. = 22; South Korea = 22.5) (see Table 1).

Scales used in the present study were initially translated into Korean by a bilingual Korean marketing expert. Using back-translation, the precision of the translation was reviewed and revised by two other Korean bilingual scholars of advertising and marketing who are also familiar with English-language survey instruments.

Table 1. Summary of Extreme Response Index (ERI).

	The United States ($n = 242$)	South Korea ($n = 205$)	t
Consumer ethnocentrism			
# Items	17	17	
# Responses on 1 and 7	5.50	3.98	
ERI	0.32	0.23	3.40 (445)*
SD	0.99	0.96	
Mean	2.59	2.91	
α	0.94	0.92	
Attitude toward advertisements for foreign products			
# Items	8	8	
# Responses on 1 and 7	0.42	0.33	
ERI	0.05	0.04	0.73 (445)
SD	0.91	0.92	
Mean	4.45	4.37	
α	0.91	0.91	

*Significant at $p < 0.05$.

Measures

To assess the extent of ERS between both culture and response formats, the same two measurement scales were used within each cultural sample. For each sample, one measurement scale used a Likert-type response format, and the second scale used the SD response format. This allowed for not only a comparison of levels of ERS between response formats within each culture but also for an assessment of differing degrees of ERS between cultures when utilizing the same response format.

Consumer Ethnocentrism

To measure the degree of consumer ethnocentrism, Shimp and Sharma's (1987) CETSCALE was adopted. This scale was initially designed to evaluate consumers' ethnocentric tendencies related to purchasing foreign versus American made products. However, the scale has been applied in other cultures (e.g., France, Japan, Russia) and has shown reliability for use in cross-cultural research (Durvasula, Andrews, & Netemeyer, 1997; Netemeyer, Durvasula, & Lichtenstein, 1991). The scale contains 17 Likert-type items anchored by 7 points (1 = "strongly disagree"; 7 = "strongly agree"). All 17 items show acceptable internal consistency levels in the two cultures: Cronbach's alpha of 0.94 in the United States, and 0.92 in Korea. Higher scores represent higher ethnocentricity.

Attitude toward Advertisements for Foreign Products ($A_{foreign-ad}$)

A general advertising attitude scale was used to evaluate consumers' attitudes toward advertisements for foreign products (Beltramini, 1988; Beltramini & Evans, 1985). The scale consists of eight SD items with seven points (e.g., "untrustworthy"–"trustworthy"; "not credible"–"credible"). Subjects were asked to answer the following question: "What is your general attitude toward advertisements for foreign products compared to advertisements for domestic products?" Cronbach's alpha showed acceptable reliability coefficients for the two cultures: 0.91 in the United States, and 0.91 in Korea. Higher scores on the scale reflect consumers' more favorable attitude toward advertisements for foreign products.

RESULTS

To obtain scores for ERS in both consumer ethnocentrism and attitude toward advertisements for foreign products scales, the frequencies of response for the various response categories were calculated. Following

Bachman and O'Malley (1984), we created the ERS index by summing the total number of extreme response (i.e., the total number of responses of 1 and 7). This number was then divided by the total number of items in each sample, resulting in an ERS index ranging between 0.00 and 1.00.

Mean ERS scores from the Likert-type scale were 0.323 for the U.S. group and 0.234 for the Korean group. Mean scores from the SD scale were 0.053 for the U.S. group and 0.042 for the Korean group. H1a stated that, when using a Likert-type scale, ERS bias would be greater for both the U.S. group and the Korean group than when using a SD scale (i.e., within group comparisons). Results support this hypothesis. For the U.S. group, the mean ERS score was statistically greater when using the Likert-type scale than when incorporating the SD scale ($t = 12.92$, $p < 0.01$). Likewise, for the Korean group, the mean ERS score was statistically greater when using the Likert-type scale than when using the SD scale ($t = 10.28$, $p < 0.01$). H1b expected that, when using a Likert-type scale, ERS bias would be greater for the U.S. group than for the Korean group (i.e., between group comparison). Results support this hypothesis, as well. ERS bias was statistically greater for the U.S. group than for the Korean group ($t = 3.43$, $p < 0.01$). H2 stated that, when using a SD scale, there will be no difference in ERS bias between the U.S. group and the Korean group (i.e., between group comparison). Results provide support for this hypothesis ($t = 0.74$, NS).

DISCUSSION

In today's global marketing environment, researchers cannot afford to ignore ERS in cross-cultural research because their findings may be contaminated by ERS effects. ERS varies between cultures and is a serious concern in cross-cultural marketing research, especially if corporate executives are to have sufficient confident in research findings to make critical decisions affecting their companies' marketing strategies, including selection of overseas target markets.

Marketing researchers have long suspected and substantiated that Likert-type scales are prone to extreme response bias across cultures, and this study provides further support for this conclusion. We found that both Korean and American respondents tended to make extreme responses on seven-point Likert-type scales, but the U.S. ERS level was significantly higher than the level for Korean participants. It seems likely that the end-points (strongly agree and strongly disagree) on the Likert scale caused Korean respondents to pull back from the extremes to their cultural comfort zone of moderation, irrespective of the question content. Moreover, perhaps Likert-type scales result in lower involvement among respondents as they pay less attention to item content in carrying out the somewhat mechanical process of simply "agreeing" or "disagreeing" to a series of statements. Respondents may be more apt to merely check strongly agree or strongly disagree in response to statements that have repetitive scale categories and anchors.

In general, it was known beforehand that Asian respondents tend to have a lower tendency to use extreme response categories than do Western respondents. We offered an explanation for these findings by referring to Hofstede's definitions of collectivistic (Korean) versus individualistic (U.S.) societies. In collectivistic cultures, people tend to be group-oriented and find no particular value in standing apart from the group. Instead, they prefer harmonious, stable, and supportive relationships as they are believed to contribute to mental and physical health. Belief in the value of moderation, conformity, and loyalty to the group encourages Asian respondents to choose more centralist or neutral positions than do U.S. respondents on Liker-type scales. In contrast to Asian cultures, Western cultures, exemplified by the United States, tend to more highly value independence, self-assertiveness, and standing out from the crowd. Western culture values confidence, conviction, and independence.

Hofstede's other four behavioral dimensions also provide insights as to why Asian and Western cultures may differ in responding to various scale formats. Cultures with high power distance (PDI), such as Korea, feel pressure to meet the expectations of other group members (Hui, Xie, & Zhou, 2008) and avoid expressing divergent opinions. Higher uncertainty avoidance (UAI) cultures, like Korea, strive to minimize ambiguity and risks, preferring the structure and stability of following group norms; whereas, lower UAI cultures (U.S.) are more at ease with ambiguity and accepting risks outside group norms. Cultures higher in masculine qualities, like the U.S., are more competitive and strive for individual achievement and success, so they may more readily select extreme responses on scales. However, more feminine cultures, like Korea, are more modest and less likely to choose extreme points in responding to questions through a scale. Turning to long-run versus short-run orientations, Western cultures are more apt to be future or long-run oriented, while Asian cultures are more focused on the past and present that may contribute to an aggressive versus a conservative approach to scale formats. As seen in the foregoing summary, all five of Hofstede's constructs suggest that Asian cultures will respond more conservatively to questions than would Western cultures as our results for the Likert-type format showed.

A major finding from our study is that the SD measurement scale appears to reduce extreme response bias, so that the two cultures (Asian and Western) were found not to be significantly different in terms of their attitudes toward marketing phenomena. More so than the SD, Likert-type scales force respondents to reveal the direction (agree or disagree) and intensity (strongly agree or strongly disagree) of their personal opinion or attitude toward a person, object, or concept. U.S. respondents, who tend to exhibit more individualistic behaviors and attitudes are more willing to state strong opinions, either positive or negative. Whereas, Korean respondents, who tend to exhibit more collectivistic behaviors and attitudes, are less willing to take strong positions on the Likert-type scale, thereby suggesting response style bias. However, when switching to the SD scale, no significant differences were found between the Korean and the U.S. responses. Given the more connotative, perceptual, and descriptive nature of SD scales as compared to Likert-type scales, Korean respondents appear more willing to give their individual perceptions of the objects or constructs being studied without fearing that they are expressing extreme attitudes. Moreover, because SD scales ask respondents to consider contrasting descriptive adjectives, like attractive–unattractive, people in all cultures recognize that nothing is totally or wholly so. Thus, they perceive that no product, service, or person possesses an unlimited amount of a quality, nor is it totally devoid of that quality. Merely providing one's perception of something as on a SD scale seems to be less threatening to the person's group membership and loyalty than is expressing his or her individual level of agreement with a proposition as required by Likert-type scales. On the basis of our results, it appears that cross-cultural marketing researches would yield more valid results if the SD were used in pan-cultural studies rather than Likert-type scales.

Our research has some limitations that offer exciting opportunities for future scholarly work. Since only one collectivist culture (Korea) and one individualist culture (U.S.) were investigated here, it would be valuable to learn how other Asian cultures (e.g., Japan, Singapore, or India) respond to the two Likert-type and SD scale formats. Other Western countries (e.g., Germany, France, England, Sweden) might be included in these future studies. Also, it would be insightful to see whether Eastern European countries (e.g., Czech Republic, Hungary, Poland, Russia, Ukraine) vary in their responses to different scale formats from those of Western and Asian countries. African countries might also be included in future studies.

Although a student sample was appropriate for the purposes of the current investigation, there may be unresolved generalizability issues. Therefore, future samples might be drawn from businesspeople or more representative consumers in the countries of interest. Potential research might consider other popular response formats, such as Stapel scales, graphic rating scales, itemized rating scales, Thurstone scales, or comparative rating sales. It would be interesting to compare the levels of ERS as the scale format changes along with the culture. Of course, providing greater explanation as to ''why'' these differences occur across cultures would help researchers develop ways to adjust to or avoid the bias elements of designated response styles across cultures.

Finally, this study used Likert-type scales and SD scales that measured two different constructs to test for relative differences in ERS between scale types. While specific item content is not believed to influence ERS, we cannot entirely rule out the possibility that a portion of the differences in scores may be construct specific in nature. To address this potential, future investigations into the role of response format on ERS might use several different scale types that measure the same construct, thereby eliminating any differences in scores that might be influenced by item content. Overall, the use of different scale formats in multiple cross-cultural studies is a fertile area for future scholarly research.

REFERENCES

Albaum, G. (1997). The Likert scale revisited: An alternate version. *Journal of the Market Research Society*, *39*(2), 331–348.

Albaum, G., Alpert, M., & Alpert, J. (1987). Response set bias and cross-cultural measures of attribute importance. In: C. F. Keown and A. G. Woodside (Eds.), *Proceedings of the 2nd symposium on cross-cultural consumer and business studies*.

Albaum, G., Golden, I., Murphy, B., & Strandskov, J. (1987). Likert scale and semantic differential: Issues relevant to cross-cultural research. In: C. F. Keown and A. G. Woodside (Eds.), *Proceedings of second symposium on cross-cultural consumer and business studies*.

Aulakh, P., & Kotabe, M. (1993). An assessment of theoretical and methodological development in international marketing: 1980–1990. *Journal of International Marketing*, *1*(2), 5–28.

Bachman, J. G., & O'Malley, P. M. (1984). Yea-saying, nay-saying, and going to extremes: Black-white differences in response styles. *Public Opinion Quarterly*, *48*(2), 491–509.

Bagozzi, R. P., & Yi, Y. (1993). Multitrait-multimethod matrices in consumer research: Critique and new developments. *Journal of Consumer Psychology*, *2*, 143–170.

Baumgartner, H., & Steenkamp, J. B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, *38*(May), 143–156.

Baumgartner, H., & Steenkamp, J. B. E. M. (2006). An extended paradigm for measurement analysis of marketing constructs applicable to panel data. *Journal of Marketing Research*, *43*(August), 431–442.

Bearden, W. O., Netemeyer, R. G., & Teel, J. E. (1989). Measurement of consumer susceptibility to interpersonal influence. *Journal of Consumer Research*, *15*(4), 473–481.

Beltramini, R. F. (1988). Perceived believability of warning label information presented in cigarette advertising. *Journal of Advertising*, *17*(1), 26–32.

Beltramini, R. F., & Evans, K. R. (1985). Perceived believability of research results information in advertising. *Journal of Advertising*, *14*(3), 18–24.

Berg, I. A. (1967). *Response set in personality assessment*. Chicago, IL: Aldine.

Bond, M. H., & Hwang, K. K. (1986). The social psychology of Chinese people. In: M. H. Bond (Ed.), *The psychology of Chinese people* (pp. 213–266). New York: Oxford University Press.

Chen, C., Lee, S., & Stevenson, H. W. (1995). Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *American Psychological Society*, *6*(3), 170–175.

Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-Cultural Psychology*, *31*(2), 187–212.

Chiu, R. K., & Kosinski, F. A. (1994). Is Chinese conflict-handling behavior influenced by Chinese values? *Social Behavior and Personality*, *22*, 81–90.

Cronbach, L. J. (1946). Response set and test validity. *Educational and Psychological Measurement*, *6*(Winter), 475–494.

Cronbach, L. J. (1950). Further evidence on response sets and test design. *Educational and Psychological Measurement*, *10*, 3–31.

Chun, K. T., Campbell, J. B., & Yoo, J. H. (1974). Extreme response style in cross-cultural research: A reminder. *Journal of Cross-Cultural Psychology*, *5*, 465–480.

Clarke, I., III. (2001). Extreme response style in cross-cultural research. *International Marketing Review*, *18*(3), 301–324.

Couch, A. S., & Keniston, K. (1960). Yeasayers and naysayers: Agreeing response set as a personality variable. *Journal of Abnormal and Social Psychology*, *60*, 151–174.

Craig, C. S., & Douglas, S. P. (2000). *International marketing research* (2nd ed.). New York, NY: Wiley.

DeBary, W. T., Chart, W. T., & Watson, B. (1960). *Sources of Chinese tradition*. New York, NY: Columbia University Press.

De Jong, M. G., Steenkamp, J. B. E. M., Fox, J. P., & Baumgartner, H. (2008). Using item response theory to measure extreme response style in marketing research: A global investigation. *Journal of Marketing Research*, *45*(February), 104–115.

Douglas, S. P., & Craig, C. S. (1983). *International marketing research with S.P. Douglas*. Englewood Cliffs, NJ: Prentice Hall.

Durvasula, S. J., Andrews, C., & Netemeyer, R. G. (1997). A cross-cultural comparison of consumer ethnocentrism in the United States and Russia. *Journal of International Consumer Marketing*, *9*(4), 73–93.

Greenleaf, E. A. (1992a). Improving rating scales measures by detecting and correcting bias components in some response styles. *Journal of Marketing Research*, *29*(May), 176–188.

Greenleaf, E. A. (1992b). Measuring extreme response style. *Public Opinion Quarterly*, *56*, 328–351.

Grimm, S. D., & Church, A. T. (1999). A cross-cultural investigation of response biases in personality measures. *Journal of Research in Personality*, *33*, 415–441.

Guptara, P., Murray, K., Razak, B., & Sheehan, T. (1990). The art of training abroad. *Training & Development Journal*, *44*, 13–18.

Hofstede, G. H. (1980). *Culture's consequences: International differences in work-related values*. Beverly Hill, CA: Sage.

Hofstede, G. H. (1991). *Cultures and organizations: Software of the mind*. London: McGraw Hill.

Hofstede, G. H. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Hui, C. H., & Triandis, H. C. (1985). Measurement in cross-cultural psychology: A review and comparison of strategies. *Journal of Cross-Cultural Psychology*, *16*, 131–152.

Hui, C. H., & Triandis, H. C. (1989). Effects of culture and response format on extreme response style. *Journal of Cross-Cultural Psychology*, *20*, 296–309.

Hui, C. H., Xie, C., & Zhou, J. (2008). The effects of country-of-origin on Chinese consumers' wine purchasing behavior. *Journal of Technology Management in China*, *3*(3), 292–306.

Johnson, T. R. (2003). On the use of heterogeneous thresholds ordinal regression models to account for individual differences in extreme response style. *Psychometrika*, *68*(4), 563–583.

Koo, L. C. L. (1976). *Nourishment of life: The culture of health in traditional Chinese society*. Unpublished dissertation thesis, University of California, Berkeley.

Lee, C., & Green, R. T. (1991). Cross-cultural examination of the Fishbein behavioral intentions model. *Journal of International Business Studies*, *22*, 289–305.

Lentz, T. F. (1938). Acquiescence as a factor in the measurement of personality. *Psychological Bulletin*, *35*(November), 659.

Leung, K. (1989). Cross-cultural differences: Individual-level vs. culture-level analysis. *International Journal of Psychology*, *24*, 703–719.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *140*, 44–53.

Leung, K., & Bond, M. H. (1989). On the empirical identification of dimensions for cross-cultural comparisons. *Journal of Cross-Cultural Psychology*, *20*, 133–151.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Bulletin*, *98*(2), 224–253.

Markus, H. R., & Kitayama, S. (1994). The cultural construction of self and emotion: Implications for social behavior. In: H. R. Markus (Ed.), *Emotion and culture: Empirical studies of mutual influence* (pp. 89–130). Washington, DC: American Psychological Association.

Netemeyer, R. G., Durvasula, S., & Lichtenstein, D. R. (1991). A cross-national assessment of the reliability and validity of the CETSCALE. *Journal of Marketing Research*, *28*, 320–327.

Osgood, C. E., May, W. H., & Miron, M. S. (1975). *Cross-cultural universals of affective meaning*. Urbana, IL: University of Illinois Press.

Oskamp, S. (1977). *Attitudes and opinions*. Englewood Cliffs, NJ: Prentice Hall.

Oyserman, D., Coon, H. M., & Kemmelmeiser, M. (2002). Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin*, *128*(1), 3–72.

Paulhus, D. L. (1991). Measurement and control of response bias. In: J. P. Robinson, P. R. Shaver & L. S. Wright (Eds.), *Measures of personality and social psychological attitudes* (pp. 17–59). San Diego, CA: Academic Press.

Schaninger, C. M., & Buss, W. C. (1986). Removing response-style effects attribute-determinance ratings to identify market segments. *Journal of Business Research*, *14*, 237–252.

Samiee, S., & Jeong, I. (1994). Cross-cultural research in advertising: An assessment of methodologies. *Journal of the Academy of Marketing Science*, *22*(3), 205–217.

Shimp, T. A., & Sharma, S. (1987). Consumer ethnocentrism: Construction and validation of the CETSCALE. *Journal of Marketing Research*, *24*(3), 280–289.

Smith, P. B. (2004). Acquiescent response bias as an aspect of cultural communication style. *Journal of Cross-Cultural Psychology*, *35*(1), 50–61.

Song, W. (1985). A preliminary study of the character traits of the Chinese. In: W. S. Tseng & D. Y. H. Wu (Eds.), *Chinese culture and mental health* (pp. 47–55). Orlando, FL: Academic Press.

Stevenson, H. W., Lee, S., Chen, C., Stigler, J. W., Hsu, C., & Kitamura, S. (1990). Contexts of achievement: A study of American, Chinese, and Japanese children. Monographs of the Society for Research in Child Development, Vol. 55 (Serial No. 221).

Triandis, H. C. (1994). Cross-cultural industrial and organizational psychology. In: H. C. Triandis, et al. (Eds), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 14, pp. 103–172). Menlo Park, CA: Consulting Psychologist Press.

Van de Vijver, F. J. R., & Leung, K. (1997). Methods and data analysis of comparative research. In: J. W. Berry, Y. H. Poortinga & J. Pandey (Eds.), *Handbook of cross-cultural psychology, Volume 1: Theory and Method* (pp. 257–300). Boston, MA: Allyn and Bacon.

van Herk, H., Poortinga, Y. H., & Verhallen, T. M. M. (2004). Response styles in rating scales: Evidence of method bias in data from six eu countries. *Journal of Cross-Cultural Psychology*, *35*(3), 346–360.

Welkenhuysen-Gybels, J., Billiet, J., & Cambre, B. (2003). Adjustment for acquiescence in the assessment of the construct equivalence of likert-type score items. *Journal of Cross-Cultural Psychology*, *34*(6), 702–722.

A MULTICOUNTRY ADVERTISING RESEARCH FRAMEWORK: LESSONS LEARNED FROM TESTING GLOBAL CONSUMER CULTURE POSITIONING

Shintaro Okazaki, Barbara Mueller
and Sandra Diehl

ABSTRACT

Purpose – *This chapter presents a framework useful in conducting multicountry marketing and advertising research. For the purpose of illustrating the series of steps involved in conducting such investigations, a six-country study examining global consumer culture positioning (GCCP) is presented. The suggested steps are relevant for the exploration of a wide variety of marketing- and advertising-related topics.*

Methodology/approach – *Steps essential to a well-planned research design are addressed in detail, including: theory identification, stimuli selection, hypotheses formulation, measurement development, country selection, fictitious ad development, survey design, cross-national data equivalence, and hypotheses testing. Particular attention is given to construct specification (in this case for soft-sell and hard-sell advertising*

Measurement and Research Methods in International Marketing
Advances in International Marketing, Volume 22, 73–96

ISSN: 1474-7979/doi:10.1108/S1474-7979(2011)0000022007

appeals) and fictitious ad development. General consumers in six countries responded to the ads. Specific procedures for validating formative constructs and testing their cross-country equivalency are suggested.

Findings – *The chapter provides practical recommendations for conducting cross-cultural research. These recommendations are likely to prove useful to both researchers conducting multicountry investigations, and to instructors teaching graduate-level courses in international marketing and advertising research.*

Originality/value of paper – *Multicountry research requires a series of challenging decisions. Although a well-planned research design is particularly essential in a cross-cultural setting, little attention has been given in providing researchers and instructors with methodological recommendations. This chapter is intended to be a useful reference for these audiences.*

Keywords: Multicountry research; research framework; construct specification; fictitious ad development; validation of formative constructs; cross-country equivalency; global advertising strategy; global consumer culture positioning; soft-sell/hard-sell appeals

OBJECTIVE OF THE CHAPTER

Although all marketing and advertising research involves rigorous procedures, a well-planned research design is particularly essential when conducting investigations in a multicountry setting. The research process in the international environment is significantly more complex, requiring a series of challenging decisions. Construct definition, specification, and validation exemplify such complexity, as researchers typically must apply proposed constructs and theory in more than one country. In addition, attention should be paid to how comparability and equivalence of results in different countries can be established. All too often, marketing and advertising researchers are confronted with reinventing the wheel when conducting such cross-cultural investigations. Clearly, they would benefit from a framework which outlines steps essential to executing a multicountry investigation. Such a framework would also be invaluable to instructors teaching graduate-level courses in international marketing or advertising research. Both master and PhD candidates require extensive guidance if they are to employ a multicountry investigation in their efforts to contribute to the academic literature.

The objective of this chapter is to outline a framework for conducting a multicountry research investigation. A six-country study is employed for the purpose of illustrating the series of steps involved in conducting such an investigation. This study tests one of the emerging global branding theories: global consumer culture positioning (GCCP). We believe that the study serves as a useful case, given that it involved all phases of a typical international marketing or advertising research investigation.

SUGGESTED MULTICOUNTRY RESEARCH FRAMEWORK

The phases incorporated in conducting our international advertising investigation are outlined in Table 1. In what follows, we attempt to describe a step-by-step procedure for executing a multicountry study according to this framework.

THEORY IDENTIFICATION

Any researcher hoping to make a significant contribution to the literature must begin with a solid theory and a strong rationale supporting the topic under investigation. Research advances in the area of international marketing, and in particular, efforts focusing on global advertising, have been characterized as sluggish in recent years (Ford, Mueller, & Taylor, 2011; Taylor, 2005). All research should be based on a carefully selected underlying – and hopefully innovative – theory.

For our investigation, the underlying theory employed was that of GCCP. GCCP is one of three branding strategies suggested by Alden, Steenkamp, and Batra (1999). In GCCP, a brand is associated with a widely understood and recognized set of symbols believed to constitute emerging global consumer cultures (e.g., Nike's "Just do it" international campaign). Local consumer culture positioning (LCCP) is defined as a strategy that associates the brand with local cultural meanings, reflects the local culture's norms and identities, is portrayed as consumed by local people in the national culture, and/or is depicted as locally produced for local people (e.g., McDonald's chicken teriyaki ads in Japan). In contrast, foreign consumer culture

Table 1. Multicountry Research Framework.

Steps	Suggested Decision Criteria
1. Theory identification	Literature review
2. Stimuli selection	Focal point of research
3. Hypotheses formulation	Theoretical propositions
4. Measurement development	
• Specification issues	Reflective vs. formative measures
• Item-generation	Literature review; qualitative inquiry; Q-sorting
• Validation	Multiple validations; student as well as general consumer sample; nomological net
5. Country selection	Cultural grouping, collaborator availability
6. Fictitious ad development	
• Stimuli development	Content analysis; focus groups
• Pre-survey examination of the stimuli	
7. Survey design and implementation	Type and size of sample
• Questionnaire development	
• Manipulation check	
8. Cross-national data equivalence	Structural equation modeling
• Cross-national invariance of reflective measures	
• Cross-national invariance of formative measures	
• Common method bias	
9. Hypotheses testing	Comparative statistical analysis
• Mean difference	
• Structural path difference	
10. Lessons learned	Theoretical as well as practical implications

positioning (FCCP) positions the brand as symbolic of a specific foreign consumer culture (e.g., Spain's Sangre de Toro ads in countries other than Spain). Alden et al. (1999) content analyzed a total of 1,267 national-brand television advertisements from 7 countries. Coders were asked to determine whether the overall sales appeal of the advertisement should be labeled as soft-sell/image (image-oriented content that does not emphasize reasons to buy, but rather general associations with the brand) or hard-sell/direct (sales-oriented, verbal, strong message arguments, comparative content). The researchers found that ads employing GCCP indeed utilized soft-sell messages more frequently than hard-sell ones. In fact, 56.3% of the GCCP ads employed a soft-sell approach, whereas just 43.7% employed the hard-sell approach. This difference is significant and in the direction predicted by the researchers. Zhou and Belk (2004) applied the GCCP framework in their examination of reader responses toward globally versus locally positioned Chinese advertisements. They also found global advertisements used less literal or "softer" appeals, and portrayed the image of cosmopolitan sophistication. Employing GCCP in a case study approach, Amine, Chao, and Arnold (2005) examined Taiwan's country-image advertising campaign. Their study lent further support to Alden et al.'s (1999) finding that global ads employ soft- over hard-selling tactics, as Taiwan's ads portrayed an affective approach using images of culture and quality of life.

STIMULI SELECTION

Alden et al. (1999) argue that GCCP is the optimal strategy for global brands, and theorize that a soft-sell approach (indirect and image based) is more suitable than a hard-sell approach (direct and information-based) for GCCP. However, they did not examine consumers' perceptions of soft- versus hard-sell appeals in a global context. They note that the features of the soft-sell approach make it more suitable than the hard-sell approach for a global consumer culture positioning strategy. A core argument of GCCP is that such image-based content is more likely to be uniformly accepted across borders than commercial messages containing direct, explicit content. Further, it is proposed that soft-sell advertising appeals will elicit more favorable reactions from consumers in multiple markets. The reasoning here is that soft-sell advertising appeals deliver more visual imagery and are more subtle and ambiguous than hard-sell advertising appeals, which are primarily based on more feature-oriented informational content. As a result, image-based soft-sell appeals tend to evoke more implicit and abstract responses (Messaris, 1997), whose interpretation may require less culturally specific cues. And as Alden et al. (1999) note, "because global consumer culture is an emerging and rapidly changing phenomenon, with differing sets of signs in differing global segments, advertising using this positioning should be more effective if it communicates in a subtle, indirect and abstract fashion. A more direct and tangible approach runs a greater risk of misspecifying the symbols that are reflective of GCCP" (p. 79).

However, while the investigations noted earlier explored the prevalence of GCCP ads and documented that the content of such ads tend to employ an image-based, soft-sell approach, consumer preferences for soft- versus hard-sell approaches in advertising had not been explored. Thus, soft- versus hard-sell advertising appeals became the focal point of our investigation.

HYPOTHESES FORMULATION

A core argument of GCCP is that image-based content is more likely to be uniformly accepted across borders than commercial messages containing direct, explicit content, due to the association between brands and the imagined membership in a global consumer segment (Appadurai, 1990). This suggests that soft-sell advertising appeals will elicit more favorable reactions from consumers in multiple markets. Our investigation extended previous GCCP research by examining whether brands that do employ an image-based, soft-sell approach in standardized advertisements are perceived more favorably by consumers in six different markets than brands using hard-sell appeals. Variables used in this study are ad credibility, ad irritation, attitude toward the ad, and intention to purchase the advertised product, all of which have been regarded as established measures of advertising effectiveness. In light of prior research, there was good reason to believe that soft-sell appeals would lead to greater credibility, less irritation, and more favorable attitudes toward the ad (Okazaki, Mueller, & Taylor, 2010). The following hypotheses were proposed:

H1. Advertisements will be perceived as more credible when standardized soft-sell appeals are employed, than when standardized hard-sell appeals are employed, in all markets examined.

H2. Less irritation will be perceived when standardized soft-sell appeals are employed, than when standardized hard-sell appeals are employed, in all markets examined.

Prior research indicated that implicit and mood-based advertising appeals are likely to elicit more favorable attitudes toward the ad, compared with explicit and argumentational advertising appeals (Pae, Samiee, & Tai, 2002). If the perception of soft-sell appeals is uniformly more favorable across six countries, there is a greater likelihood that this appeal might be effectively employed in GCCP strategies. We expected the contrary to be true in case of the ads employing hard-sell advertising appeals. Thus, we hypothesized the following relationships:

H3. Attitudes toward the advertisement will be more favorable when standardized soft-sell appeals are employed, than when standardized hard-sell appeals are employed, in all markets examined.

H4. Purchase intention of the advertised brand will be greater when standardized soft-sell appeals are employed, than when standardized hard-sell appeals are employed, in all markets examined.

Our next set of hypotheses went beyond mere attitudinal differences by addressing structural relationships among the variables.

H5. Standardized soft-sell appeals will lead to stronger credibility, compared with standardized hard-sell appeals.

H6. Standardized soft-sell appeals will lead to less irritation, compared with standardized hard-sell appeals.

In line with H3, it seemed logical to posit a causal relationship between advertising appeals and attitude toward the ad. That is, soft-sell approach in standardized advertisements would cause favorable attitude toward the ad, due to the less culturally specific cues based on image and emotions (Alden et al., 1999). Therefore:

H7. Standardized soft-sell appeals will lead to more favorable attitudes toward the ad, compared with standardized hard-sell appeals.

Finally, we completed our model by adding purchase intention as a final dependent variable. Attitude toward the ad is known to be a strong determinant of ultimate purchase intent, which has widely been documented in the literature on behavioral research (Ajzen, & Fishbein, 1980). Hence, we formulated the final hypothesis:

H8. Attitude toward the ad will lead to stronger purchase intention when standardized soft-sell appeals are employed, than when standardized hard-sell appeals are employed.

MEASUREMENT DEVELOPMENT

The next step is the development of the measurements used in a study. Here, researchers should recognize that while the psychometric method (Churchill, 1979) is considered the traditional approach, is has recently come under some criticism (Rossiter, 2002). Researchers should focus not only on sender-oriented perspectives, but also consider receiver-oriented theories, given that many practitioners and academics alike have shifted their advertising paradigm from a behaviorist to a constructivist approach. This approach assumes that ‘‘message recipients treat ‘stimuli’ as problems to be understood and solved rather than as overpowering shots from a cannon against which no defense but surrender was possible’’ (Mendelsohn, 1990). With regard to our study, based on the adoption of a constructivist approach and supported by an extensive literature review, we identified three contrasting dimensions that have proven useful in defining the essence of soft- and hard-sell advertising appeals: (1) feeling versus thinking, (2) implicitness versus explicitness, and (3) mood versus fact.

Specification Issues

In our analysis of consumer responses to soft- versus hard-sell appeals, we were confronted with how to specify the relationship between the above noted contrasting dimensions and the latent construct. We found careful review of measurement model specification guidelines and recommendations in research conducted by Diamantopoulos, Riefler, and Zeugner-Roth (2008), Diamantopoulos and Siguaw (2006), Jarvis, MacKenzie, and Podsakoff (2003), Rossiter (2002), and Diamantopoulos and Winklhofer (2001) to be of great benefit. Investigators should be aware that there is significant debate over the issue of possible misspecification regarding the direction of causality between a construct and its measures. Observed indicators can be treated as reflective or formative. In the case of the former, underlying constructs are hypothesized to cause changes in the indicators, while in the latter, changes in the indicators are hypothesized to cause changes in the underlying construct (Jarvis et al., 2003). All too often, researchers tend to automatically turn to reflective indicators, as ‘‘virtually all progress in the assessment of constructs and their measures has been based on classical test theory and the assumptions it makes about the relationships between a construct and its indicators’’ (p. 199). Yet, as Hulland (1999) points out, ‘‘the choice between using formative or reflective indicators for a particular construct can at times be a difficult one to make’’ (p. 201). Jarvis et al. (2003) explain that ‘‘some potentially serious consequences of measurement model misspecification exist, and researchers need to think carefully about the direction of causality between constructs and their measures.’’ Indeed, a number of the studies cited earlier provide examples of constructs which have previously (and erroneously) been operationalized by means of reflective multi-item scales, although a formative perspective would have been theoretically appropriate.

In light of these concerns, we determined that our concepts of soft- and hard-sell advertising appeals should be multidimensional composite constructs, where second-order factors have first-order factors as formative indicators and the first-order factors themselves also have formative indictors. More specifically, with regard to our soft-sell appeal measurement instrument, the three dimensions (feeling, implicitness, and mood) are relatively independent sources of appeal that, together, all share the characteristic of being "soft sell." Thus, they are conceptualized as the first-order factors that determine or form the second-order factor "soft-sell" appeals. Each of the first-order factors was measured by a series of formative indicators. Similarly, we conceptualized the hard-sell appeal measurement instrument as a second-order factor, whose formative indicators (thinking, explicitness, and fact), are measured by a series of formative measures.

This conceptualization is in line with Jarvis et al.'s (2003) discussion of overall similarity construct, proposed by Crosby, Kenneth, and Cowles (1990) as being a function of appearance, lifestyle, and status similarity. Additional examples of prior research employing second-order constructs, with formative indicators for both the first- and second-order factors include Bruhn, Georgi, and Hadwich (2008), and Ulga and Eggert (2006). As a result of this conceptualization, our measurement instruments were constructed and validated as formative measures, rather than reflective measures.

Item Generation

The next step in our construct development was to propose measurement items that could be used in determining the three components of soft- (feeling, implicitness, and mood) and hard-sell appeals (thinking, explicitness, and fact). A thorough review of the literature related to these appeal categories, combined with consultations with marketing academics, generated a total of 30 adjective items, 15 items each for soft- and hard-sell appeals. In an attempt to expand this list of items, a content analysis of 899 print ads was conducted. All 2005 issues of four nationally circulated US magazines (news, women's, sports, and business) were collected. Employing Lin's (2001) methodology, six ads representative of soft- and hard-sell advertising appeals were selected (one soft- and one hard-sell ad for each of the following product categories: autos, cell phones, and alcoholic beverages). These product categories were intended to represent low, medium, and high involvement goods. The appropriateness of the ads for this investigation was assessed by a panel of four advertising and marketing scholars. All sample advertisements were considered suitable for the purpose of the study. Next, nine focus group sessions were conducted, each consisting of five to six participants. The session moderator presented participants with the six ads in rotating order. Subjects were asked to verbally describe their impressions of the ads. Sessions were tape-recorded and key terms identified.

To ensure content validity of the formative measures, Pette, Straub, and Rai (2007), recommend Q-sorting and expert panels. In Q-sorting, persons with no prior knowledge of the study are asked to examine a series of cards containing the measures that will be used for the construct and to categorize each card into a specific construct. For this investigation, we conducted a free-association task with 109 student subjects. A free-association task can serve as an even more rigorous alternative to Q-sorting, given the larger sample size and the fact that the setting is less artificial. In our investigation, after exposure to the six ads, subjects were asked to write down their impressions and perceptions in terms of adjectives. This exercise produced a list of 27 nonredundant adjectives (items) for the soft-sell appeal measurement instrument and 27 nonredundant adjectives (items) for the hard-sell appeal measurement instrument. Seven advertising/marketing experts were recruited to assess the items using a three-point scale: appropriate, indifferent, and inappropriate. In light of the generally accepted definitions of soft- and hard-sell, all measurement items were deemed to be suitable.

Validation

Resulting measurement instruments should ideally be validated via multiple sets of data. For our investigation, we conducted two separate tests. The first validation was performed with undergraduate students at a large American university located in the southwest. A structured questionnaire was developed and the six advertisements (three matched ads each for soft- and hard-sell appeals) culled from the previously noted content analysis were employed in the instrument validation. Per Rossiter (2002), the use of multiple ad samples improves the generalizability of the results. Following exposure to either the soft- or hard-sell ad, 220 subjects were asked to indicate which of the 54 items (27 soft- and 27 hard-sell items) applied to the ad on a seven-point scale ("not at all applicable" to "fully applicable"). A within-subject design was employed to extract maximum and consistent perceptions of the appeals from the sample. All responses were pooled and randomly divided into an equal number of two subsamples.

Focusing on the first subsample, a principal component analysis was conducted for the soft-sell appeal measure. The items converged into the three proposed components, explaining 76% of total variance. The same procedure was repeated for the hard-sell appeal measure; and here the three proposed components captured more than 70% of the variance. Next, the model was tested via partial least squares (PLS). PLS was preferred over covariance-based structural equation modeling, as it functions with measurement models consisting of formative indicators (Chin, 1998). PLS also offers greater flexibility than covariance-based methods, given that it employs a least-squares estimation procedure, thereby avoiding many restrictive assumptions such as multivariate normality and residual distributions (Falk & Miller, 1992). Furthermore, PLS was considered more appropriate for this investigation as it is primarily intended for predictive analyses in which (1) the explored problems are complex, (2) theoretical knowledge is scarce, and (3) sample size is small (Chin, 1998; Hulland, 1999). In this model, formative first-order indicators were optimally weighted and combined using the PLS algorithm to create latent variable scores (Chin, 1998). The hierarchical component model suggested by Wold (1982) was employed to examine the second-order model. In this method, the indicators that measure second-order factors are used twice: once for measuring the first-order factors, and again for measuring the second-order latent construct. The software used was SmartPLS version 2.0 M3 (Ringle, Wende, & Will, 2005) which applies a bootstrapping method (500 cases, sample size 220).

The model fit in PLS should be assessed in light of (1) estimation of individual item reliability, (2) individual item weight, (3) correlations between the items, and (4) indicator multicollinearity. Individual formative-item reliability was examined according to the loadings of the items on their respective constructs. Following Hair, Ringle, and Sarstedt's (2011) generally accepted recommendation, attempts were made to retain items higher than 0.70. This ensures that there is more shared variance between the construct and its measure than error variance (Diamantopoulos & Winklhofer, 2001). In our investigation, several of the items did not meet this criterion. Regarding item weights, negative figures were found in some of the model results, which may have been caused by the existence of multicollinearity between items. An analysis of both Pearson product–moment correlations and variance inflation factor (VIF) revealed that a number of the items were highly correlated, which complicated assessment of the indicators' validity. Also, indicators with almost perfect linear combinations of others contained redundant information, thus inflating the estimation results (Bruhn et al., 2008). It should be noted that an excessive number of highly correlated indicators is undesirable ''because of both the data collection demands it imposes and the increase in the number of parameters when the construct is embedded within a broader structural model'' (Diamantopoulos & Winklhofer, 2001, p. 272). These items were thus eliminated, reducing the number of items soft-sell appeals to 15, and for hard-sell appeals to 18. Internal consistency examinations (e.g., Cronbach's alpha) are not appropriate for formative indicators (Hair et al., 2011). The same validation procedure was repeated using the remaining subsample. The model fit in terms of the loadings and weights was very similar to the preceding validation.

Regarding measurement validation, in order to increase the generalizability of the results, at least one study should employ a general consumer sample. In our investigation, 195 nonstudent respondents participated in the second measurement validation. Trained interviewers approached subjects on city streets and on public transit, as well as in cafes, and explained that the investigation dealt with consumer responses to advertising. Participants examined each of the six ads (the same ads used in the previous validation with the student sample) and responded to a short questionnaire. A within-subject design was again employed to extract maximum and consistent perceptions of the appeals from the sample.

As a next step, the validated scales were tested in a multicountry setting utilizing fictitious ads employing hard- and soft-sell appeals.

COUNTRY SELECTION

Central to multicountry research is the selection of specific markets to be analyzed. Though rigorous justification is required regarding the particular countries chosen, in practice, those selected often depend on the availability of collaborators. Nevertheless, researchers should make every attempt to balance such pragmatic concerns with the more important theoretical rationale behind the countries included in the investigation. For example, this study examined consumer responses to soft- and hard-sell appeals in six countries, including those reflecting different communication styles (low versus high context), as well as those belonging to different country clusters per GLOBE (House, Hanges, Javidan, Dorfman, & Gupta, 2004). France, Italy, and Spain belong to the Latin Europe cluster, Germany to the Germanic cluster, Japan to the Confucian Asian cluster, and the USA to the Anglo cluster. All six countries are important industrial nations.

FICTITIOUS AD DEVELOPMENT

Stimuli Development

To test the hypotheses outlined in our investigation, stimulus ads were developed and pretested. Ad development consisted of three qualitative procedures: content analysis, focus group discussions, and free association tasks. The first step was to conduct a content analysis of European and US magazines. A total of 54 students at a large Austrian University conducted the analysis, examining the magazines for product categories employing standardized ad campaigns. Six product categories were found to commonly employ a standardized approach: sportswear, coffee, wristwatches, mobile phones, cosmetics, and alcoholic beverages. This was followed by a brainstorming session, which revealed that wristwatches were considered by the subjects to be the most appropriate category for a standardized campaign, as watches are (1) equally important to both male and female consumers, regardless of age; (2) commonly found in magazine ads; and (3) the use of an unknown fictitious brand name would appear credible. Further, wristwatches are considered by consumers in many markets as a means of demonstrating identity and lifestyle, and as such, can be considered global cultural symbols.

Students were then instructed to search for wristwatch ads in international publications. This exercise resulted in the collection of 79 unique magazine ads for watches. Among these ads, sports (such as skiing, cycling, and swimming) were found to be the most frequently employed ad theme, along with celebrity endorsements. Sports can be considered part of global culture; sport activities have a positive association in virtually all countries, and are viewed as both important and desirable. In terms of stimulus ad development, use of a sport theme was deemed equally suitable for messages employing soft- and hard-sell appeals. The collected ads were next analyzed for features generally associated with wristwatches. In order to identify relevant product features related to wristwatches, students performed free association tasks. Based on this exercise, the following product features were identified as most important: "design," "quality," "endurance," "material," and "price." A sheer frequency count revealed endurance to be the most commonly highlighted feature in international wristwatch ads.

A series of focus groups were then conducted with the same students in order to identify a sports theme for the stimulus ads. These sessions revealed that, if endurance were to be emphasized in the ads, the featured sport should ideally be associated with both risk and adventure, and that reliable materials should be highlighted. A number of sports activities were singled out as potential candidates: hang gliding, surfing, diving, parachute jumping, mountain biking, and skiing. Several were considered as less than appropriate or even unappealing for some of the countries to be surveyed, resulting in the selection of skiing and mountain biking. In order to reflect the endurance capabilities of the wristwatch, two sets of visuals were crafted: for the soft-sell ads, a skier was portrayed alone in one ad, and a sole mountain biker was portrayed in the other; for the hard-sell ads, several skiers competing in a race were presented in one version, whereas several mountain bikers participating in a competition were portrayed in the other. General reactions toward these ads were tested in a series of eight focus groups, which deemed the visuals associated with skiing most appropriate. During these same sessions, three possible slogans were also generated: "Wherever you go, whatever you do," "Stands the test of time," and "Always progressive." Finally, a fictitious brand name had to be selected. The focus group sessions, along with extensive research of existing brands, produced three potential brand names: "Chronier," "Viventure," and "Iving."

PRE-SURVEY EXAMINATION OF THE STIMULI

The aforementioned ad components were rigorously pretested. Sixty-six students provided their opinions regarding the appropriateness of (1) the product category, (2) the main product feature, and (3) the fictitious brand name, via a semi-structured survey instrument. Respondents were offered five different choices and asked to rate them on a seven-point scale. Results indicated that wristwatches are indeed a commonly advertised product category, with which most respondents were relatively familiar. Both visuals and slogans were also perceived as very suitable for the product category. Endurance was the second-most frequently mentioned valued product feature (following design), confirming this as a relevant feature for the stimulus ads. Of the three brand names tested, "Chronier" received the highest evaluation ($M = 4.30$). Skiing was perceived as a more appropriate sports theme than biking, as the latter elicited some negative associations, such as doping, drugs, enormous stress, and so on.

Free association tasks were employed to confirm the appropriateness of the slogans. The slogan "Always progressive" appeared to give subjects the impression that the wristwatch might be imprecise or run too fast, and thus, was dropped. The two remaining slogans, "Wherever you are, whatever you do" and "Stands the test of time," both elicited very positive associations. In terms of visuals, respondents associated both the lone skier and the participation in a race with words such as "adventure," "dynamic," "appealing," and "exciting." As anticipated, the soft-sell visual was not associated with competition, whereas the hard-sell visual was.

The final stimulus ads were professionally developed by the creative department of an Austrian advertising agency. The advertisements were designed in black/white to avoid the influence of color preferences. To ensure that the fictitious ads possessed the appeals intended, a manipulation check was performed with 140 general consumers. Subjects evaluated the soft-sell appeal version of the stimulus ad using both hard- and soft-sell formative measurement items developed by Okazaki et al. (2010), in which a soft-sell appeal is conceptualized as a second-order formative model consisting of three first-order constructs: feeling (creative, instinctive, imaginative, and abstract), implicitness (insinuation, appealing, subjective, and expressive), and image (entertaining, interpretive, playful, and impression based). Similarly, the hard-sell appeal consists of three first-order constructs: thinking (rational, logical, analytic, factual, and concrete), explicitness (precise, explanation, convincing, persuasion, and instructive), and fact (educational, descriptive, realistic, informative, and evidence based). For simplicity's sake, mean values of first-order constructs were calculated assuming an equal weight of each loading. For the soft-sell appeal ad, a t-test indicated that the mean value of the summed soft-sell appeal scale items was significantly greater than that of the summed hard-sell appeal scale items ($t = 37.73$, $p < 0.001$). Thus, respondents indeed perceived the ad as the study intended. The same procedure was repeated for the hard-sell appeal version of the stimulus ad. Resulting data suggest that the mean value of the summed hard-sell appeal was significantly greater than that of the soft-sell appeal ($t = 15.72$, $p < 0.001$). Thus, this ad was indeed perceived as a hard-sell appeal ad. Based on the above, the manipulation check was considered to be a success.

SURVEY DESIGN AND IMPLEMENTATION

Questionnaire Development

A two-part structured questionnaire was developed. All the construct items were included in the first part, whereas the second part requested demographic information. The questionnaire was translated for each country using the translation-back-translation procedure suggested by Craig and Douglas (2005), to ensure cross-cultural content equivalency. All interviewers received extensive training, and were provided with a standardized text to employ in approaching subjects in order to ensure comparable interviewing situations in all countries. A general consumer sample participated in the main investigation. Subjects were approached along public streets, at the entrance to walking/jogging parks, on public transit, as well as in cafés. Subjects were told that the investigation explored consumer responses to advertising. Given that the majority of respondents completed the questionnaire independently, interviewer influence was minimized. To analyze differences in responses to the hard- and soft-sell appeal ads, a ''between-subject'' design was employed, so that in each country different subjects were shown the soft- and hard-sell appeal ads. The final sample size employed in this investigation was 392 in the USA, 374 in Japan, 424 in Germany, 381 in France, 479 in Spain, and 400 in Italy.

Manipulation Check

A thorough manipulation check was performed to ensure that respondents in each country perceived the stimulus ads as intended. For each dataset, a mean value for the soft- and hard-sell appeal ads was calculated (Okazaki et al., 2010). Results confirmed that, in all countries, both the soft- and hard-sell ads were perceived as intended. Thus, the manipulation was deemed successful for all countries.

CROSS-NATIONAL DATA EQUIVALENCE

The importance of data equivalence in cross-cultural business research has been widely documented (Craig & Douglas, 2005). In particular, measurement equivalence in terms of metric equivalence for reflective measures has been a focus of debate in prior research, whereas little attention has been paid to data equivalence for formative measures. Because our investigation employed both types of measures, this section offers comprehensive evaluation of each case.

Reflective Measures

To examine the reliability and validity of the reflective measures, a confirmatory factor analysis (CFA) was performed, using AMOS 18.0 with the maximum likelihood estimation. Specifically, we estimated a full-sample measurement model with ad credibility, ad irritation, attitude toward the ad, and purchase intention. The full sample model resulted in an acceptable fit: $\chi^2_{71} = 567.93$ ($p < 0.001$), CFI = 0.95, IFI = 0.95 for the soft-sell model; $\chi^2_{71} = 523.95$ ($p < 0.001$), CFI = 0.95, IFI = 0.95 for the hard-sell model. All items loaded significantly onto respective constructs at $p < 0.0001$. Then, we tested cross-country measurement invariance of this baseline model across six countries, using the method that Steenkamp and Baumgartner (1998) suggest. The results provide partial metric invariance for the soft-sell model ($p = 0.10$), and the hard-sell model ($p = 0.64$) as two loadings per construct were invariate in all countries. On this basis, we calculated the composite reliability and average variance extracted, for both pooled and separate country datasets. All figures exceed the recommended levels of 0.70 and 0.50, respectively (Hair, Black, Babin, Anderson, & Tatham, 2006).

Formative Measures

Next, we performed an invariance test for the formative measures, that is, the soft- and hard-sell advertising appeals. We followed the guidelines suggested by Diamantopoulos and Papadopoulos (2010) using the multiple indicators-multiple causes (MIMIC) model (Diamantopoulos & Winklhofer, 2001). Diamantopoulos and Papadopoulos (2010) propose the notion of slope invariance which refers to the degree to which the formative measure is influenced to the same extent by a given indicator in each country. If slope invariance is supported, then it is legitimate to compare structural relationships involving the focus constructs across countries. If not, "the theoretical importance of the indicators is not stable across countries" (p. 362). However, in practice, full measurement invariance is extremely difficult to achieve, thus researchers should ensure that at least partial measurement invariance holds (Steenkamp & Baumgartner, 1998). If some of the indicators in a formative measure are invariant across countries, partial slope invariance can be ascertained. Diamantopoulos and Papadopoulos (2010) proposed a three-step procedure: (1) testing for metric invariance of the reflective indicators; (2) estimating a baseline MIMIC model; and (3) introducing equality constraints on the parameters of the formative indicators.

The MIMIC model is a latent construct combining formative and reflective indicators (Diamantopoulos & Winklhofer, 2001; MacKenzie, Podsakoff, & Jarvis, 2005). Following this procedure, we first tested metric invariance of content-valid indicators for informational and transformational appeals adapted from Puto and Wells (1984). The goal of informational appeals is to focus directly on factual, relevant brand data in a clear and logical manner, whereas transformational appeals emphasize the experience of brand usage with a unique set of psychological characteristics. Although the concepts of informational/transformational do not capture the full meaning of hard- and soft-sell as employed in this investigation, there does seem to exist some overlap. Two items are "The ad evokes positive emotions" and "Purchasing this brand would make me feel good about myself." These indicators were subsequently used as reflective items in the MIMIC model. Because a two-indicator single-factor model is unidentified, we added one more item, "This ad stimulates my imagination," as a reference indicator by fixing its loading as 1. A multigroup CFA model across six countries produced a good fit, supporting the metric invariance ($\chi^2_6 = 11.84$, CFI = 1.0, RMSEA = 0.020) with a nonsignificant χ^2 value ($p = 0.05$). Then, using these two reflective indicators, along with the formative indicators of soft-sell appeals (means of each dimensions), the baseline MIMIC model was tested (M1), which resulted in highly acceptable fit statistics. Next, we tested full slope invariance by constraining all coefficients of the formative measure to be equal across countries (M2). The difference in χ^2 value between M1 and M2 was statistically significant ($p < 0.001$), rejecting full slope invariance hypothesis. A careful examination of associated parameter changes indicated that the significant χ^2 value was due to the coefficient of image in three countries, namely Spain, Italy, and France. Thus, this path was freed in these countries. The fit statistics of the resulting model were not significantly worse than those of the baseline (i.e., fully unconstrained) model. Thus, partial slope invariance was established.

The same procedure was repeated for hard-sell appeals. To test metric invariance of the reflective indicators, we added "This ad made me think about important features related to watches" and "This ad seemed to be speaking directly to me," with "This ad explained to me clearly the watch brand I should purchase" (Puto & Wells, 1984) being a reference indicator. A multigroup CFA yielded a good fit ($\chi^2_6 = 9.01$, CFI = 1.0, RMSEA = 0.043) with nonsignificant χ^2 value ($p = .17$). Thus, subsequent MIMIC model was tested for full slope invariance. The difference in χ^2 value between the baseline and the constrained model was statistically nonsignificant ($p = .14$), thus supporting full slope invariance hypothesis.

Common Method Bias

Most researchers agree that common method variance is a potentially serious biasing threat in behavioral research, when estimating models that use same-source surveys (Chang, van Witteloostuijn, & Eden, 2010). Podsakoff, MacKenzie, Lee, and Podsakoff (2003) suggest a confirmatory factor-analytic approach, instead of the Harman one-factor test, which is considered an insensitive test. In this method, a worse fit for the one-factor model would suggest that common method variance does not pose a serious threat. The one-factor model of all the reflective measures yielded a statistically significant χ^2 difference, compared with the original CFA model, whereas other fit indexes worsened considerably. This suggests that common method bias was not a serious threat in this study.

HYPOTHESES TESTING

Mean Difference (H1–H4)

To test our hypotheses, we computed the mean values for each construct and applied *t*-tests between standardized soft-sell versus standardized hard-sell ads in each country. H1 assumed that ad credibility would be greater for the soft-sell appeal ad than for its hard-sell counterpart. ANOVA found no significant differences between the two samples in any of the countries with the exception of Italy, where the soft-sell ad was perceived as significantly more credible than the hard-sell ad ($M = 4.83$ vs. 4.34 for the soft- and hard-sell ads, respectively; $p < 0.001$). Thus, H1 was not supported. Next, in H2, we posited that the soft-sell ad would be less irritating, compared with its hard-sell counterpart. In fact, this turned out to be true in Italy (2.80 vs. 3.15), Germany (2.78 vs. 3.12), Spain (2.19 vs. 2.66), Japan (2.19 vs. 2.61), and France (.2.44 vs. 3.14), with the USA (3.12 vs. 2.73) being the only exception. Thus, H2 was reasonably supported.

H3 suggested that more favorable attitudes toward the ad would be found for the soft-sell ad than for the hard-sell. Such differences were found in Italy (5.11 vs. 4.69), Japan (4.71 vs. 4.48), and France (4.97 vs. 4.55) in the hypothesized direction. No other statistical significance was found. Thus, H3 was partially supported. Finally, for H4, purchase intention was posited to be significantly stronger for the soft-sell ad, compared with the hard-sell. This was true in the USA (3.33 vs. 2.84), but the contrary was the case in Spain (3.41 vs. 3.79). Such contradictory findings resulted in the rejection of H4.

Structural Difference (H5–H8)

Beyond the bivariate analysis, hypotheses H5–H8 further our theoretical propositions, by hypothesizing differences in the strength of structural relationships among the proposed constructs. Our model consisted of soft- and hard-sell appeal measures, ad credibility, ad irritation, attitude toward the ad, and purchase intention. The basic premise was that if a standardized soft-sell appeal provoked a series of perceptual chain reactions, and if the strength of the relationships was more stable and stronger, as compared with those of a hard-sell appeal, we could safely say that soft-sell appeals are more suitable for GCCP. To this end, we performed structural equation modeling.

First, we tested the baseline full-sample model using the maximum likelihood method with AMOS 18.0. Results show an acceptable fit: $\chi^2_{1095} = 3761.77$, CFI = 0.90, RMSEA = 0.045 for soft-sell model and $\chi^2_{1095} = 2187.87$, CFI = 0.91, RMSEA = 0.036 for hard-sell. Next, we performed an individual estimation of the baseline model for each country. The results were fairly consistent across countries.

The structural path from soft-sell appeals to attitude toward the ad was significant at $p < 0.001$ in all countries, with the exception of Germany ($p < 0.10$). In addition, statistically significant unstandardized betas were consistently high across countries, ranging between 0.56 (the USA) and 0.95 (Spain). In contrast, with regard to hard-sell appeals, this path was significant in only two countries (Japan and the USA), with rather modest unstandardized betas.

Finally, the path from soft-sell appeals to attitude toward the ad was statistically significant in all countries, except Germany (which was marginally significant at $p < 0.10$). In addition, the magnitude of coefficient betas was greater than 0.50 in France, Spain, and Japan. In contrast, the path from hard-sell appeals to attitude toward the ad was significant only in two countries. In addition, regarding these two countries, the standardized coefficients of the path from hard-sell appeals to attitude toward the ad were notably lower than those of the path from soft-sell appeals to attitude toward the ad. This appeared consistent with our *t*-tests results in that soft-sell appeals had a more important impact on attitude formation, compared with hard-sell appeals. Thus, H7 was reasonably supported by our data. With regard to the path from advertising appeals to credibility, standardized coefficients seemed to be more solid in the hard-sell model than in the soft-sell. As for the other structural paths, results were less clear. Thus, the remaining hypotheses, namely H5, H6 and H8, were not supported.

LESSONS LEARNED

This chapter places a special emphasis on construct definition, as well as specification and validation issues as critical phases in designing multi-country research. Unless the construct is well defined, it cannot be validated in more than a single country, and as a result, all efforts will have been wasted. In terms of construct development, we propose a structured methodology consisting of a series of essential steps. Decisions related to reflective versus formative indicators have been identified as particularly important, given that an ever increasing number of international marketing researchers fail to comply with basic guidelines regarding the use of such indicators (Diamantopoulos and Papadopoulos, 2010).

As a practical lesson, a relatively simple study design is recommendable so that it can easily be replicated in foreign markets. If an experiment is incorporated, investigators should decide early on whether to adopt a within- or between-subject design. Pros and cons are associated with each approach, and relate to sample size, demand craft, response bias, etc. Further, researchers must determine the number and type of products to be included in the study. One approach for selecting products is to conduct a content analysis of magazine ads published in the countries under examination. Items promoted in such ads can then serve as a pool of potential product categories.

Another challenging task is the stimulus development. Here, a particularly challenging issue relates to the role of verbal versus visual components. With regard to our examination of advertising appeals, we found that the potential existed for soft-sell visual images to "mask" the effects of more hard-sell textual messages. Thus, the combination of visual and verbal or textual content should be measured by an appropriate instrument. Further, investigators must consider any and all additional elements incorporated in the advertisement, such as background images, slogans employed, use of color, gender and nationality of models, and so on. Rigorous focus group sessions are recommended, along with additional qualitative techniques to uncover the influence of such elements. Both realism and manipulation checks are recommended to ensure the validity of the study.

Methodological errors can be minimized during the data collection phase by engaging in the translation-back-translation technique for any research instruments employed in the investigation and via the use of homogeneous survey methods. Before conducting any comparative analysis (e.g., ANOVA, ANCOVA, *t*-test, MANOVA, etc.), statistical treatments should thoroughly examine the invariance structure for both reflective and formative measurements. Regarding the former, Steenkamp and Baumgartner's (1998) method is widely recommended. As to the latter, recently, a method outlined by Diamantopoulos and Papadopoulos (2010) was proposed as an innovative technique to establish invariance of formative measures. Over the past few decades, there has been continued growth in international marketing and advertising research. However, multimarket investigations are fraught with pitfalls. It is hoped that the framework outlined in the preceding pages can help both current and future researchers avoid at least some of them.

ACKNOWLEDGMENTS

The preparation of this chapter has been facilitated by a grant from the Spanish Ministry of Science and Innovation (National Plan for Research, Development and Innovation EC02008-01557), and the Yoshida Hideo Memorial Foundation (Tokyo, Japan).

REFERENCES

Alden, D., Steenkamp, J. B. E. M., & Batra, R. (1999). Brand positioning through advertising in Asia, North America and Europe: The role of global consumer culture. *Journal of Marketing*, *63*(1), 75–87.

Amine, L. S., Chao, M. C. H., & Arnold, M. J. (2005). Exploring the practical effects of country of origin, animosity, and price-quality issues: Two case studies of Taiwan and Acer in China. *Journal of International Marketing*, *13*(2), 114–150.

Appadurai, A. (1990). Disjuncture and difference in the global economy. In: M. Featherstone (Ed.), *Global culture: Nationalism, globalization and modernity* (pp. 295–310). London: Sage Publications.

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behaviour*. Englewood Cliffs, NJ: Prentice-Hall.

Bruhn, M., Georgi, D., & Hadwich, K. (2008). Customer equity management as formative second-order construct. *Journal of Business Research*, *61*(12), 1292–1301.

Chang, S. J., van Witteloostuijn, A., & Eden, L. (Eds.). (2010). From the editors: Common method variance in international business research. *Journal of International Business Studies*, *41*, 178–184.

Chin, W. W. (1998). The partial least squares approach for structural equation modeling. In: G. A. Marcoulides (Ed.), *Modern methods for business research* (pp. 295–336). Hillsdale, NJ: Lawrence Erlbaum Associates.

Churchill, G. A. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, *16*, 64–73.

Craig, C. S., & Douglas, S. P. (2005). *International marketing research*. Chichester, UK: Wiley.

Crosby, L. A., Kenneth, R. E., & Cowles, D. (1990). Relationship quality in services selling: An interpersonal influence perspective. *Journal of Marketing*, *54*(July), 68–81.

Diamantopoulos, A., & Papadopoulos, N. (2010). Assessing the cross-national invariance of formative measures: Guidelines for international business researchers. *Journal of International Business Studies*, *41*(February/March), 360–370.

Diamantopoulos, A., Riefler, P., & Zeugner-Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research*, *61*(12), 1203–1218.

Diamantopoulos, A., & Siguaw, J. A. (2006). Formative versus reflective indicators in organizational measure development: A comparison and empirical illustration. *British Journal of Management*, *17*(4), 263–282.

Diamantopoulos, A., & Winklhofer, H. M. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research*, *38*(May), 269–277.

Falk, R. F., & Miller, N. B. (1992). *A primer for soft modeling*. Akron, OH: University of Akron Press.

Ford, J., Mueller, B., & Taylor, C. R. (2011). The tension between strategy and execution: Challenges in international advertising research: Globalization is much more than global branding. *Journal of Advertising Research*, *51*(1), 27–41.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* (6th ed.). Upper Saddle River, NJ: Prentice Hall.

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM. Indeed a silver bullet. *Journal of Marketing Theory & Practice*, forthcoming.

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (Eds.). (2004). *Culture, leadership, and organizations*. Thousand Oaks, CA: Sage.

Hulland, J. (1999). The use of partial least squares (PLS) in strategic management research: A review of four recent studies. *Strategic Management Journal*, *20*(2), 195–204.

Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research*, *30*(2), 199–218.

Lin, C. A. (2001). Cultural values reflected in Chinese and American television advertising. *Journal of Advertising*, *30*(4), 83–95.

MacKenzie, S. B., Podsakoff, P. M., & Jarvis, C. B. (2005). The problem of measurement model misspecification in behavioral and organizational research and some recommended solutions. *Journal of Applied Psychology*, *90*(4), 710–730.

Mendelsohn, H. (1990). Mind, affect, and action: Construction theory and the media effects dialectic. In: S. Kraus (Ed.), *Mass communication and political information processing* (pp. 37–45). Hillsdale, NJ: Lawrence Erlbaum Associates.

Messaris, P. (1997). *Visual persuasion*. Thousand Oaks, CA: Sage.

Okazaki, S., Mueller, B., & Taylor, C. R. (2010). Measuring hard-sell and soft-sell advertising appeals. *Journal of Advertising*, *39*(2), 5–20.

Pae, J. H., Samiee, S., & Tai, S. (2002). Global advertising strategy: The moderating role of brand familiarity and execution style. *International Marketing Review*, *19*(2), 176–189.

Petter, S., Straub, D. W., & Rai, A. (2007). Specifying formative constructs in information systems research. *MIS Quarterly*, *31*(4), 623–656.

Podsakoff, P. M., MacKenzie, S. B., Lee, J., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *20*(5), 879–903.

Puto, C., & Wells, W. D. (1984). Informational and transformational advertising. The differential effects of time. *Advances in Consumer Research*, *11*, 638–643.

Ringle, C. M., Wende, S., & Will, A. (2005). SmartPLS – Version 2.0 (beta), http://www.smartpls.de. Hamburg, Germany: University of Hamburg.

Rossiter, J. R. (2002). The C-OAR-SE procedure for scale development in marketing. *International Journal of Research in Marketing*, *19*(4), 305–335.

Steenkamp, J. B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*(1), 78–90.

Taylor, C. R. (2005). Moving international advertising research forward. *Journal of Advertising*, *34*(1), 7–16.

Ulga, W., & Eggert, A. (2006). Value-based differentiation in business relationships: Gaining and sustaining key supplier status. *Journal of Marketing*, *70*(January), 119–136.

Wold, H. (1982). Soft modeling: The basic design and some extensions. In: K. G. Joreskog & H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction. Part II* (pp. 1–54). Amsterdam: North Holland.

Zhou, N., & Belk, R. (2004). Chinese consumer readings of global and local advertising appeals. *Journal of Advertising*, *33*(3), 63–76.

STOCHASTIC FRONTIER ESTIMATION IN INTERNATIONAL MARKETING RESEARCH: EXPLORING UNTAPPED OPPORTUNITIES

Matthew E. Sarkees and Ryan Luchs

ABSTRACT

Purpose – *This chapter explores the basic characteristics of stochastic frontier estimation, discusses advantages of the method that make it conducive to research in international marketing, and provides an application to demonstrate its use. Potential applications in international marketing research are also discussed.*

Methodology – *Stochastic frontier estimation.*

Findings – *Stochastic frontier estimation models, prevalent in other fields, are very limited in the international marketing literature. Many potential opportunities exist for its use in the context of international marketing.*

Originality/value of paper – *The intent of this chapter is to show that stochastic frontier estimation is a potentially valuable tool for*

Measurement and Research Methods in International Marketing
Advances in International Marketing, Volume 22, 99–114

ISSN: 1474-7979/doi:10.1108/S1474-7979(2011)0000022008

international marketing research. We show this by demonstrating the use of the tool and by providing examples of potential research studies.

Keywords: Stochastic frontier estimation; international marketing; efficiency; input–output models

Various innovative analytical methods enjoy widespread application throughout academic marketing journals. Yet, one potentially valuable analytical method, stochastic frontier estimation (SFE), is seemingly under-utilized in marketing, and in particular, in the international marketing literature. Other disciplines, such as economics, draw on SFE to explain issues that are critical to managers, government officials, and other key stakeholders. For example, a scan of recent years in the *Business Source Premier* database produces more than 100 articles that use SFE. Alternatively, a scan of marketing journals produces a very limited number of papers that draw on SFE (e.g., Dutta, Narasimhan, & Rajiv, 1999; Haughton, Haughton, Kelly-Hawke, & Moriarty, 2000; Luo & Donthu, 2005; Narasimhan, Rajiv, & Dutta, 2006; Yeh, 2010). SFE papers with an international marketing focus are even more scant.

SFE is an econometric method that captures the efficiency with which inputs are converted into an output (Battese & Coelli, 1988, 1995; Greene, 2003a; Kumbhakar & Lovell, 2000). This input–output effect estimates efficiency levels in the context of ‘‘competitors’’ in the same panel of data (Aigner et al., 1977). The results provide a ‘‘frontier’’ of observations so that managers or key decision makers can understand which ones are more and less efficient.

Traditionally, SFE research looks at efficiency in maximizing production or minimizing cost in areas such as agriculture and manufacturing, among others. For example, public policy makers and managers may explore which farms are most efficient in emerging economies, which manufacturing inputs generate a desired output (most cars, defect free cars, etc.), or which individual unit among many (farm animal, piece of equipment, employee, etc.) produces the most.

The practical use of SFE at various levels of analysis for marketing managers is untapped but readily apparent. SFE can measure firm-level actions, such as how combinations of resource investments generate a particular output. For example, a manager can analyze how certain marketing-related inputs generate sales. However, the input–output effect also has a number of other applications at business unit, functional, group or individual levels. SFE is not a cure-all analytical method for every issue.

The analytical method should always be in line with the research question and data (Cadogan, 2010). However, where cross-national or regional comparisons of the impact of activities within or across firms are important, SFE can be a viable option in an international marketing context.

This chapter provides a background on some of the key characteristics of SFE as well as an illustration of its use. The context of the illustration is marketing-related investments of U.S. and international publicly traded firms to highlight its potential in international marketing research. Following the illustration, some potential theoretical considerations and international research issues where SFE can be a helpful analytical method are discussed. The goal of this chapter is not to advocate that SFE is "best" but to spark interest in SFE as an appropriate analytical method when researchers are considering certain international marketing issues and relationships related to efficiency.

STOCHASTIC FRONTIER ESTIMATION

SFE: Characteristics

SFE is a parametric method that uses maximum likelihood estimation. It has certain econometric methodological advantages that make it conducive to international marketing research. SFE can be better suited for certain types of data sets than other econometric methods such as data envelopment analysis (DEA).[1] Aside from the advantages listed below, other choices, depending on the research questions, may become as important and should be considered before proceeding.

First, as a parametric method, SFE is useful with both small and large data sets that require researcher assumptions and probability distributions. The distribution decision is non-trivial. Distribution assumptions such as half-normal, truncated normal, exponential, or gamma can be used in SFE. This can provide increased power over non-parametric methods (DEA), but caution is suggested as the researcher is responsible for determining the distribution (Greene, 1990,2003a, 2003b).

Second, SFE can accommodate different types of data as inputs and outputs. For example, a researcher can integrate financial data (income, expenses, price, etc.), count data (number of alliances, employees, consumers, age, etc.), and dummy variables (gender, product type) in an SFE model. This allows for great flexibility in modeling as long as researchers have correctly specified the model. For international marketing researchers, struggling with various data types or with how to bring data types together for analysis, this is a key aspect of SFE.

Third, SFE allows for a two-part error term that captures both inefficiency in the firm and inherent randomness. The analysis of group, business unit, or firm issues in particular lend to SFE given the two-part error term. Parsing out the error not due to randomness allows for explicit understanding of that component of deviation from the "frontier" of more efficient competitors. Thus, managers are better informed due to the separation of the two-part error term. The two-part error term provides an advantage over DEA, which is a deterministic method and does not provide for this level of detail (Kooreman, 1994).

Fourth, SFE is well-suited for panel data, and in particular unbalanced panel data (Battese & Coelli, 1992). The specification of Battese and Coelli (1992) allows for time-variant estimates of technical efficiency. DEA can deal with panel data but estimates either need to be calculated separately for each year (e.g., Reinhard, Lovell, & Thijssen, 2000) or as a sequential frontier where all observations up to a given year are used to calculate the technical efficiency (e.g., Hjalmarsson, Kumbhakar, & Heshmati, 1996). Thus, SFE can explicitly handle panel data by allowing for differences in technical efficiencies across years whereas DEA accommodates panel data by estimating each year separately or by using all previous years' of data until the year of interest.

Finally, researchers can employ SFE as either a minimization or a maximization model. This allows researchers to investigate different types of outputs that are related to key questions of interest. As an illustration, the maximization (e.g., such as a model with the goal of understanding what factors maximize sales or units produced) of an objective function takes the following form (Battese & Coelli, 1992):

$$Y_{it} = \int (X_{it}, \alpha) + \varepsilon_{it} - \eta_{it} \tag{1}$$

where Y_{it} is the output for the ith firm in the tth time period, X_{it} is the vector of inputs (explanatory variables), and α is the vector of coefficients for the associated input variables.[2] The two-part error term, $\varepsilon_{it} - \eta_{it}$, respectively represents vectors of stochastic error (random shocks outside of management control that influence the variables) and inefficiency error (omitted variables). Again, the composite error term is unique and advantageous to SFE as it helps to remove potential bias of the more general, singular error term. The random error component, ε_{it}, is assumed to be independent and identically distributed with a mean 0 and variance $\sigma_\eta^2 \sim N(0, \sigma_\eta^2)$. The inefficiency error component, η_{it}, is assumed to be non-negative, independent and identically distributed with a mean μ and variance $\sigma_\eta^2 \sim N(0, \sigma_\eta^2)$ with a half-normal distribution. The error terms are also assumed to be independent of each other as well as of the independent variables.

Using Eq. (1), a maximum likelihood estimate for each observation (i) in a period (t) can then be obtained through the Cobb–Douglas formula (Battese & Coelli, 1988; Dutta et al., 1999):

$$\mathrm{y} = \mathrm{e}^{\alpha}\left(\prod_{i=1}^{k}\prod_{t=1}^{k} x_{it}\alpha_{it}\right)\mathrm{e}^{\varepsilon}\mathrm{e}^{-\eta} \tag{2}$$

Rearranging Eq. (2) results in the following input–output capability model, which captures the essence of the SFE approach across a set of observations:

$$\text{Efficiency} = \frac{\mathrm{y}}{\mathrm{e}^{\alpha}\left(\prod_{i=1}^{k}\prod_{t=1}^{k} x_{it}\alpha_{it}\right)\mathrm{e}^{\varepsilon}} = \mathrm{e}^{-\eta} \tag{3}$$

Eq. (3) is a ratio of inputs to outputs such that the resulting efficiency levels can only have values between 0 and 1. Higher frontier estimates that trend toward "1" represent higher efficiencies. SFE compares the entire set of efficiency estimates for a given period, creating a "frontier" of efficiency of all observations. It provides an evaluation of competition across firms, given a suitable dataset. In marketing terms, one might picture marketing-related investments and how they impact a particular output, such as sales, comparing the efficiency scores across brands, sales teams, firms, or industries.

There are a number of variations and judgments that researchers must make when using SFE. Furthermore, the capabilities and extensions of SFE continue to grow. An incomplete list of references that discuss the intricacies of SFE is shown in Table 1.[3] A few of the seminal papers on SFE are listed (e.g., Aigner et al., 1977) as well as those that discuss enhancements to the SFE model that may be important to international marketing researchers. For example, Kumbhakar and Wang (2005) discuss a model that allows for country-level heterogeneity. Papers that discuss dealing with panel data (e.g., Battese & Coelli, 1992; Orea & Kumbhakar, 2004), which is prevalent in the marketing literature, are also shown.

Table 1. Examples of Articles Discussing Specifications and Extensions of the SFE Model.

Aigner, Lovell, and Schmidt (1977)	Derives an SFE model that has an error term that is the sum of symmetrical normal and half-normal (negative) random variables.
Battese and Coelli (1992)	Develops an SFE model for panel data. The model conceptualizes firm effects as an exponential function of time.
Battese and Coelli (1995)	Develops an SFE model for panel data in which the non-negative technical inefficiency effects are assumed to be a function of firm-specific variables and time.
Bera and Sharma (1999)	Proves that when a firm reaches its most efficient level, it also has the least production uncertainty.
Coelli et al. (2005)	Reference text that discusses various efficiency analyses including SFE as well as DEA (data envelopment analysis).
Greene (2003b)	Extends the original formulation of the SFE model that was based on a normal-exponential framework to a normal-gamma framework and provides a new estimation technique for this model.
Greene (2004)	Develops a panel data SFE model that allows for non-linearity in time-variant effects.
Greene (2005)	Considers a special case of the random parameters model that produces a random effects model that preserves the central feature of the stochastic frontier model and accommodates heterogeneity.
Haughton et al. (2000)	Apply SFE to a direct marketing problem and find that frontier estimation is particularly valuable when the data are heteroscedastic.
Huang (2004)	Develops an SFE model that allows for the possibility that firms within a dataset adopt different technologies.
Kim and Schmidt (2000)	Compare specifications of both classical and Bayesian SFE models on the same dataset. Fixed effects models generally perform poorly and that Bayesian and classical procedures perform similarly.
Kumbhakar and Wang (2005)	Estimate SFE models taking country heterogeneity into account.
Kumbhakar and Lovell (2000)	Book that discusses many aspects of SFE in detail.
Orea and Kumbhakar (2004)	Estimate a latent class SFE model which accounts for heterogeneity in a panel data setting.

SFE: An Application in an International Marketing Context

The following example of SFE is drawn from the pharmaceutical industry. The objective is to compare the efficiency of U.S. and non-U.S.

pharmaceutical firms in terms of generating an output, which in this example is sales revenue. The data set is from U.S. publicly traded pharmaceutical companies (SIC Code 2834) for the 2005 fiscal year. For ease of understanding, one year of data is used. If there were multiple years under consideration, time dummies for each year could be utilized. Also, if comparing outputs within a period across multiple periods, separate SFE estimates can be generated for each period. So, if there were 10 years of data, a researcher would have 10 SFE calculations, each providing a competitive landscape for a particular year.

A maximization model is used as the output of interest is sales revenues given a level of resource investments. Sales (SALES) is the output variable and is expressed in U.S. dollars for each firm in the sample for the 2005 year. Two measures for the key inputs that potentially influence the output, firm sales, are: (1) selling, general and administrative, and (2) marketing-oriented alliances. The two measures, one financial and one count data, are used to illustrate SFE and not necessarily to define the only relevant inputs to generate sales revenues. First, selling, general, and administrative (SGA) expenses (SGA) include a firm's costs to maintain its sales force that serves its current customer base. SGA expense is drawn from the Computstat financial database from each firm's income statement. In the pharmaceutical industry, sales representatives are a particularly large cost as they are tasked with directly interacting with physicians and hospital representatives who then prescribe the drugs for end user customers, the patients. As a result, the sales force is a significant driver of sales. Also included in SGA are advertising expenses. Research indicates that strong advertising investments promote sales (e.g., Leone, 1995). The pharmaceutical industry often relies heavily on direct-to-consumer marketing of their products to raise interest such that consumers ask healthcare providers about them.[4]

Second, marketing-oriented alliances (MKTGA) allow firms to extend their geographic reach, speed products to market or further penetrate existing customers with the primarily benefit of increased revenues. These arrangements also provide for information flows and customer feedback. The SDC Platinum database using the 2834 SIC code (pharmaceutical preparations) provides the data necessary to develop a count of marketing-oriented alliances for each firm for each year. The SDC Platinum database categorizes alliances by type, making it easy to separate these arrangements, thus reducing arbitrary decision making for the researcher.

Following Dutta et al. (1999), the Cobb–Douglas formulation is used to specify the frontier model for i firms in t years, taking the logarithm of both sides:[5]

$$\ln(\mathrm{SALES}_{it}) = \alpha_0 + \alpha_1 \ln(\mathrm{SGA}_{it}) + \alpha_2 \ln(\mathrm{MKTGA}_{it}) + \varepsilon_{it} - \eta_{it} \tag{4}$$

Drawing on Eq. (3) and rearranging Eq. (4), the model results in the following input–output model:

$$Y_{it} = \frac{\ln(\mathrm{SALES}_{it})}{\alpha_0 + \alpha_1 \ln(\mathrm{SGA}_{it}) + \alpha_2 \ln(\mathrm{MKTGA}_{it}) + \varepsilon_{it} - \eta_{it}} \tag{5}$$

In this illustration, the *frontier* command in STATA 10 was used. The frontier command produces, among other information, coefficients, standard errors, and significance levels for each input as well as information on the two-part error term (Table 2). For the coefficients for each input in the model, selling, general, and administrative expenditures had a significant effect on sales while marketing-oriented alliances did not. The model also demonstrated significant overall fit. This is an important outcome as there is the potential for lack of fit in maximum likelihood estimation models. Models that are not significant overall may have a lack of variability in the data and thus comparisons of efficiency become meaningless.

The frontier estimation process in this illustration produces individual firm observations of efficiency ranging between 0 and 1. Higher estimates correspond to firms that are more efficient at generating higher levels of

Table 2. Results.

Variable	Coefficient	S.E.
Ln(SGA)	1.29***	.09
Ln(Marketing Alliances)	−.02	.06
Constant	1.20*	.54
Parameters for compound error		
v (random error)	.46*	.21
μ (inefficiency error)	2.81***	.13
sigma_v	1.26	.13
sigma_μ	4.08	.27
sigma square ($\sigma_\mu^2 + \sigma_v^2$)	18.28	2.13
lambda (σ_μ^2/σ_v^2)	3.25	.33
Wald χ^2 (2)	214.93***	

*$p<.05$; **$p<.01$; ***$p<.001$.

sales given their marketing resource investments as compared to other firms in the given year. The average estimate across the 220 observations is .20 (s.d. = .15). For U.S. firms, the average efficiency score is $M = .20$ (s.d. = .15). For International firms, $M = .28$ (s.d. = .17). Overall, international firms demonstrate higher efficiency in this competitive environment.

There is sufficient variation across observations, but the overall efficiency is low. Either managers can generate more sales given the current level of inputs or reduce expenses and marketing-related alliances to maintain sales. Either path allows for increased efficiency. For example, the most efficient firm came in at .81 (which was an international firm). Given the mean of the sample, this implies that an average firm has the *opportunity* to increase sales up to 75% [1 – (.20/.81)] from the same marketing investments.

The output in Table 2 allows for the calculation of the Gamma $(\gamma) = [\sigma_\mu^2/(\sigma_\mu^2 + \sigma_v^2)]$, which is .913, indicating that most of the residual variation is derived from management inefficiency (μ) while the random error (v) accounts for just 8.7% (Jondrow, Lovell, Materov, & Schmidt, 1982; Yeh, 2010). Thus, the benefit of the two-part error term in SFE for decision-makers is that they can better understand what is (and is not) within their control.

SFE RESEARCH OPPORTUNITIES IN INTERNATIONAL MARKETING

Theoretical Perspectives

SFE is focused on the efficiency of transforming inputs into output(s), which lends itself most naturally to the theoretical perspectives of the resource-based view. The resource-based view suggests that firm-specific resources and capabilities account for the differences across firms (Barney, 1991; Wernerfelt, 1984). Consistent with the use of SFE as a measure of transformational efficiency, deployed resources in combination create firm-specific capabilities (Amit & Schoemaker, 1993). Thus, some researchers have taken the perspective that SFE is a measure of the strength of certain firm capabilities (Dutta et al., 1999). Under this perspective, researchers could provide insight into the resource allocation strategy of firms by examining firms of national versus global scope to see if different capabilities are important.

Using a similar perspective, a firm's capacity to absorb information or knowledge, is another theoretical lens for examining relationships using SFE (Cohen & Levinthal, 1990; Narasimhan et al., 2006). Both the resource-based view and absorptive capacity provide ample opportunity for marketing researchers to explore relationships in an international context.

Other theoretical perspectives have not been utilized in marketing but could be appropriate. For example, transaction cost economics theories explore the concept of output monitoring for suppliers. Output monitoring involves measuring the visible consequences of a partner's actions, such as a supplier's delivery time, order accuracy, and product quality (Anderson & Oliver, 1987). The same output monitoring concept can be applied to distributors, employee sales teams as well as third-party sales partners. For example, pharmaceutical firms who do not have the reach or resources to hire sales employees can compare the outputs based on a set of inputs of third-party partner teams across global geographies. Consumer behavior theories also support SFE especially when the goal is to examine the overall efficiency of inputs that generate increases in consumer action such as purchasing volume or revenues or word-of-mouth. The efficiency with which word of mouth spreads may be an interesting study to carry out across cultures.

Research Issues

There are a number of different research opportunities in an international marketing context that lend themselves to examination using the SFE methodology. This methodology is interesting because the 0 to 1 ranking of efficiency of one unit against another provides marketing decision makers with an instant picture of how they are faring against competitors. The units used in the SFE analysis could be firms or within firm units such as customers, retail outlets or members of the sales force. The following discussion represents examples of the types of opportunities that exist for using SFE in an international marketing context.

Resource theory suggests that how firms acquire and deploy resources in ways that are difficult to imitate is critical to competitive advantage (e.g., Barney, 1991). Dynamic capabilities theory asserts that in the long term it is not the simple possession of valuable resources that explains superior business performance, but the ability to create or acquire new resources, to adapt existing resources, and to release unneeded ones (e.g., Teece, Pisano, & Shuen, 1997). In this context, research on marketing capabilities is scarce

(Krasnikov & Jayachandran, 2008). There is an opportunity for international marketing researchers to explore marketing capabilities using SFE. Two marketing capabilities in particular, brand management and customer relationship management, are critical to firm success (e.g., Ambler, 2004; Srivastava, Shervani, & Fahey, 1998). Marketing-related (e.g., sales force level data such as number of employees, advertising expenditures, product data, and marketing collateral), CRM-related (revenues per brand, revenues per customers, cross-selling wins, customer wins, customer, profitability, etc.) or product-related (number of units sold) inputs and outputs are all relationships of interest in this context. Examining capabilities such as these will allow researchers to answer questions such as, ‘‘Can increased investment in the sales force generate greater revenue per customer and how much can international firms benefit from more efficient investment in the sales force as compared to U.S. firms?’’ Doing this in an international context is particularly compelling as the similarities and differences in the drivers of efficiency across markets will offer compelling insights to managers. Other related areas, such as new product development and innovation, also provide opportunities for fresh research using SFE. As SFE accommodates various types of inputs (e.g., financial and count data), studying these topics in an international context becomes particularly fruitful.

SFE also presents opportunities to study within firm effects. Given the input–output nature and the resulting frontier of efficiency, there are a number of marketing-related areas of interest. For example, drawing on sales force level data as inputs and number of units sold as outputs, researchers can examine brand success or sales force effectiveness within firms on a global basis. In the context of direct marketing, cross-sectional units (e.g., households, service providers) can be studied to identify which units will be most receptive to a particular campaign (Haughton et al., 2000).

Another area for studying within firm effects is retail chains. Many retailers are expanding their operations to multiple countries and studying the efficiency of retail outlets may provide insights into how to expand further internationally. Store-level outputs, such as retail sales or customer complaints, can be evaluated through SFE. For example, suppose management wants to minimize customer complaints, a cost minimization function in SFE. Using customer complaints at the store level as the output, managers can specify certain inputs that they believe may minimize those events. Number of store employees, customer service training expenditures for each store, number of customer transactions, and sales volume, among other factors, could be used as inputs. Furthermore, if retail sales are the output, firms may wish to detect which outlets are the most efficient at producing sales give inputs including square footage, ancillary services provided by the outlet, loyalty programs, or store configurations. For example, is having a loyalty program associated with higher sales output and are outlets in one nation more efficient at producing sales revenue than outlets in another nation?

International marketing research is keenly focused on market entry including timing and related issues (e.g., Gielens & Dekimpe, 2007). For example, developing firm–foreign retailer relationships are critical for successful market entry (Kumar, 1997). Assume a firm wants to build relationships and lay the foundation for launching products in China. A firm hires many teams to work regional areas. The firm uses a given threshold to define a relationship as having been developed. Using the number of developed relationships as an output, employee expenses, cost of marketing collateral, cost of samples, employee travel costs, population density, and miles of paved roadways can serve as inputs. Which team is most efficient at maximizing relationships given the level of inputs? In this hypothetical example, financial data, marketing data, and geographic data can be combined in SFE analysis so that managers can properly evaluate important levers of efficiency.

The growth of online and mobile marketing is ripe for SFE as the marketing research in general is relatively new and the dynamics of future innovations in this space are as yet unknown. A firm's actions online are now key components of the marketing strategy and a primary mechanism for finding, developing, and serving customers. It is the place where markets are made and competitors stage large-scale battles. Online presence can also go terribly wrong, forcing the firm to think in ways that integrated marketing communications never did before.

Online ad impressions were up 22% between 2009 and 2010 (ComScore, 2010). Internet-advertising revenues for the first half of 2010 were approximately $12.1 billion, up 11.3% over that same period in 2009 (The Internet Advertising Bureau, 2010). How are firms "measuring up" against competitors in these new, dynamic spaces? Theoretical perspectives in the context of marketing in social mediums are suitable for SFE, particularly in an international context, as they often measure multiple outputs such as number of "friends," website hits, click-through rates, or "followers" as well as more traditional outputs of sales volume and revenue. What drives increases in these outcomes? It is advertising in other mediums such as TV or radio, third-party partners that feature the firm, search engine optimization expenditures, free samples giveaways in stores, guerilla marketing, employee expenses, or the quality of the web site? All these factors can serve as inputs in an SFE model, depending on the ultimate research question.

SFE has long informed public policy in an economic context. The use of maximization and minimization models to understand degrees of efficiency (and inefficiency) in agriculture and manufacturing in emerging economies is evident in the economics literature and public policy position papers. As marketing finds its way in an international context, SFE can contribute to public policy decision making in many areas such as advertising, consumer financial decision making, consumption, and healthcare marketing, among others. Papers that address minimizing consumption could be particularly interesting, especially if they apply to children or disadvantaged groups. Papers that address maximizing and minimizing healthcare marketing or consumer consumption of healthcare are also of interest. In an international context, these types of papers from a marketing perspective can influence firms as well as public policy decision makers as they proceed in emerging markets. Overall, the frontier model has promise in public policy as the key concept of making efficient use of resources is at the top of mind of decision makers in these areas.

CONCLUSION

To date, the utilization of SFE in marketing, and in particular international marketing, research is limited. Alternative analytical methodologies, such as DEA, have been used increasingly in the literature. However, depending on the research question and the underlying assumptions, researchers may be missing an opportunity to draw on SFE in an international marketing context. International marketing research provides ample fodder for the exploration of questions of efficiency at various levels of analysis. In this context, SFE allows researchers to model efficiency issues using combinations of inputs and outputs. Key to the SFE methodology is that the results provide a picture of competitive outputs given levels of certain inputs. With this picture of efficient and inefficient observations, managers and key stakeholders can see which inputs are contributing to the desired outputs thus informing decision making. Furthermore, SFE appears to have many applications in international marketing research. Among other things, SFE has potential in comparing the effects of marketing capabilities across regions, examining the efficiency of sales force members across cultures, or comparing retail outlets in one geographic setting versus another. Marketing researchers should strongly consider SFE to analyze issues of interest and where appropriate this technique should be applied more liberally in the international marketing context.

NOTES

1. The goal of this chapter is to introduce SFE as an option for studying interesting international marketing phenomena. Alternative options, such as data envelopment analysis (a non-parametric method) have been used in the literature to ''benchmark'' outputs against a best in class outcome. Detailed comparisons of data envelopment analysis and SFE are not discussed here; see Coelli, Rao, O'Donnell, and Battese (2002) for further discussion as well as Coelli and Perleman (1999) for an example of this comparison. Some researchers suggest that SFE and DEA should be viewed as complementary methods, each with their own strengths and weaknesses (Kooreman, 1994). This chapter in no way proposes that one method is better than another.
2. A minimization (cost efficiency) model would change the error term to $\varepsilon_{it}+\eta_{it}$.
3. For those interested in an extensive primer on SFE, its extensions and implications, see Professor William H. Greene's website http://pages.stern.nyu.edu/~wgreene/ for articles, materials, and coursework on SFE.
4. Compustat does not provide a full data set for advertising in the pharmaceutical industry. However, advertising is captured in SGA line item. A review of firm annual reports suggests that a large number of pharmaceutical companies in this data set do not separately present these expenditures.
5. Natural log transformation of variables on both sides of the model is normal so that the inefficiency term can be interpreted as the percentage deviation of observed performance.

ACKNOWLEDGMENTS

The authors thank Charles R. Taylor, Marko Sarstedt, and John Hulland for their helpful guidance.

REFERENCES

Aigner, D. J., Lovell, C. A. K., & Schmidt, P. (1977). Formulation and estimation of stochastic production frontier models. *Journal of Econometrics*, *6*(1), 21–37.

Ambler, T. (2004). *Marketing and the bottom line*. Upper Saddle River, NJ: Pearson.

Amit, R., & Schoemaker, P. (1993). Strategic assets and organizational rent. *Strategic Management Journal*, *14*(1), 33–46.

Anderson, E., & Oliver, R. L. (1987). Perspectives on behavior-based versus outcome-based salesforce control systems. *Journal of Marketing*, *51*(4), 76–88.

Barney, J. B. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, *17*(1), 99–120.

Battese, G. E., & Coelli, T. J. (1988). Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *Journal of Econometrics*, *38*(3), 387–399.

Battese, G. E., & Coelli, T. J. (1992). Frontier production functions, technical efficiency and panel data: With application to paddy farmers in India. *The Journal of Productivity Analysis*, *3*(1–2), 153–169.

Battese, G. E., & Coelli, T. J. (1995). A model for technical efficiency effects in a stochastic frontier production function for panel data. *Empirical Economics*, *20*(2), 325–332.

Bera, A., & Sharma, S. (1999). Estimating production uncertainty in stochastic frontier production function models. *The Journal of Productivity Analysis*, *12*(3), 187–210.

Cadogan, J. (2010). Comparative, cross-cultural, and cross-national research: A comment on good and bad practice. *International Marketing Review*, *27*(6), 601–605.

Coelli, T. J., Rao, D. S. P., O'Donnell, C. J., & Battese, G. E. (2005). *An introduction to efficiency and productivity analysis*. New York, NY: Springer.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, *35*(1), 128–152.

ComScore. (2010). U.S. online display advertising market delivers 22 percent increase in impressions vs. year ago. Retrieved from http://www.comscore.com/Press_Events/Press_Releases/2010/11/U.S._Online_Display_Advertising_Market_Delivers_22_Percent_Increase_in_Impressions

Dutta, S., Narasimhan, O., & Rajiv, S. (1999). Success in high-technology markets: Is marketing capability critical? *Marketing Science*, *18*(4), 547–568.

Gielens, K., & Dekimpe, M. (2007). The entry strategy of retail firms into transition economies. *Journal of Marketing*, *71*(2), 196–212.

Greene, W. H. (1990). A gamma distributed stochastic frontier model. *Journal of Econometrics*, *46*, 141–163.

Greene, W. H. (2003a). *Econometric analysis*. Upper Saddle River, NJ: Pearson-Prentice Hall.

Greene, W. H. (2003b). Simulated likelihood estimation of the normal-gamma stochastic frontier function. *Journal of Productivity Analysis*, *19*(2–3), 179–190.

Greene, W. H. (2004). The behavior of the fixed effects estimator in nonlinear models. *The Econometrics Journal*, *7*(1), 98–119.

Greene, W. H. (2005). Fixed and random effects in stochastic frontier models. *Journal of Productivity Analysis*, *23*(1), 7–32.

Haughton, D., Haughton, J., Kelly-Hawk, A., & Moriarty, T. (2000). The use of frontier estimation in direct marketing. *Journal of Interactive Marketing*, *14*(2), 35–42.

Hjalmarsson, L., Kumbhakar, S., & Heshmati, J. (1996). DEA, DFA and SFA: A comparison. *The Journal of Productivity Analysis*, *7*(2–3), 303–327.

Huang, H. (2004). Estimation of technical inefficiencies with heterogeneous technologies. *The Journal of Productivity Analysis*, *21*(3), 277–296.

Internet Advertising Bureau. (2010). Internet ad revenues break records, climb to more than $12 billion for first half of '10' [Press Release]. Retrieved from http://www.iab.net/about_the_iab/recent_press_releases/press_release_archive/press_release/pr-101210

Jondrow, J., Lovell, C. A. K., Materov, I. S., & Schmidt, P. (1982). On estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics*, *19*(2–3), 233–238.

Kim, Y., & Schmidt, P. (2000). A review and empirical comparison of Bayesian and classical approaches to inference on efficiency levels in stochastic frontier models with panel data. *The Journal of Productivity Analysis*, *14*(2), 91–118.

Kooreman, P. (1994). Data Envelopment analysis and parametric frontier estimation: Complementary tools. *Journal of Health Economics*, *13*, 345–346.

Krasnikov, A., & Jayachandran, S. (2008). The relative impact of marketing, research-and-development, and operations capabilities on firm performance. *Journal of Marketing*, *72*(4), 1–11.

Kumar, N. (1997). The revolution in retailing: From market driven to market driving. *Long Range Planning*, *30*(6), 830–835.

Kumbhakar, S., & Lovell, K. (2000). *Stochastic frontier analysis*. Cambridge, UK: Cambridge University Press.

Kumbhakar, S., & Wang, H. (2005). Estimation of growth convergence using a stochastic production frontier approach. *Economic Letters, 88*(3), 300–305.

Leone, R. P. (1995). Generalizing what is known about temporal aggregation and advertising carryover. *Marketing Science*, *14*(3), 141–150.

Luo, X., & Donthu, N. (2005). Assessing advertising media spending inefficiencies in generating sales. *Journal of Business Research*, *58*(1), 28–36.

Narasimhan, O., Rajiv, S., & Dutta, S. (2006). Absorptive capacity in high-technology markets: The competitive advantage of the haves. *Marketing Science*, *25*(5), 510–524.

Orea, L., & Kumbhakar, S. (2004). Efficiency measurement using a latent class stochastic frontier model. *Empirical Economics*, *29*(1), 169–183.

Reinhard, S., Lovell, C., & Thijssen, G. (2000). Environmental efficiency with multiple environmentally detrimental variables: Estimated with SFA and DEA. *European Journal of Operations Research*, *121*(2), 287–303.

Srivastava, R. K., Shervani, T. A., & Fahey, L. (1998). Market-based assets and shareholder value: A framework for analysis. *Journal of Marketing*, *62*(1), 2–18.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, *18*(7), 509–533.

Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic Management Journal*, *5*(2), 171–180.

Yeh, T. (2010). Bank loan loss provision decisions: Empirical analysis of Taiwanese banks. *Journal of Financial Services Marketing*, *14*(4), 278–289.

MARKETING ACCOUNTABILITY: APPLYING DATA ENVELOPMENT ANALYSIS TO ASSESS THE IMPACT OF ADVERTISING EFFICIENCY ON SHAREHOLDER VALUE

Sascha Raithel, Sebastian Scharf, Charles R. Taylor, Manfred Schwaiger and Lorenz Zimmermann

ABSTRACT

Purpose – *Marketers are under increasing pressure to demonstrate the financial return associated with marketing expenditures. Concurrently, more attempts at measuring return on investment from marketing as well as achieving other long-term goals such as building brand equity and increasing shareholder value have been made. As a result of this emphasis, the degree to which advertising budgets are spent efficiently and the impact of these expenditures on the bottom line are an important topic to study.*

Methodology/approach – *This study applies data envelopment analysis (DEA) to a group of large firms to assess the degree to which companies spend advertising dollars efficiently and to examine the impact of advertising efficiency on investor behavior and, ultimately, stock prices.*

Measurement and Research Methods in International Marketing
Advances in International Marketing, Volume 22, 115–139

ISSN: 1474-7979/doi:10.1108/S1474-7979(2011)0000022009

Findings – *The analysis reveals that firms that advertise more efficiently are rewarded by investors by positive stock returns.*

Research limitations/implications (if applicable) – *The study is limited to large enterprises with strong brands within a time frame of only four years.*

Practical implications (if applicable) – *The results imply that it is advisable for marketing managers not to limit their focus to increasing market-based assets at any cost. The efficiency of their efforts can send a positive signal to investors and contribute to shareholder value enhancement.*

Originality/value of the chapter – *The chapter finds investors to pay attention not only to the effectiveness of advertising activities but also to their efficiency. The study also demonstrates how DEA and stock return response modeling can be combined to investigate the link between advertising efficiency and investor behavior.*

Keywords: Advertising efficiency; firm value; stock returns; data envelopment analysis; stock return response modeling

INTRODUCTION

Current and future customers are usually key addressees of a company's advertising efforts. Traditionally, related market place outcomes, such as brand awareness, brand recognition, sales, or market share, were the dominating metrics when chief executive officers (CEOs) and chief financial officers (CFOs) asked the question for marketing performance (Lehmann, 2004). CEOs and CFOs, however, speak the language of "return on investment," "financial performance," and "shareholder value" pressuring the marketing profession to provide evidence of its accountability. For some time now, marketing managers have found themselves needing to document the return that will come from advertising and other marketing expenditures. A special section of Advertising Age titled "ROI, the Marketer's Obsession" called attention to this pressing issue (Neff, 2005). Subsequently, there have been calls for more academic research on ROI as companies and consulting firms have made efforts to develop better metrics for measuring ROI from marketing expenditures (e.g., Taylor, 2010).

Providing evidence of return on investment, though, is not trivial as marketing outcomes were traditionally conceptualized as being predominantly of non-financial nature. This level of performance measurement exhibits a few shortcomings. These measurements do not allow any inferring as to whether the achievements fulfill the equity holders' expectations and hence create shareholder value. Beyond that, it is important to understand how marketing outcomes can be transformed into metrics containing information relevant to financial markets and showing incremental value relevance (Hanssens, Rust, & Srivastava, 2009). Therefore, marketers face the challenge to go beyond traditional marketing analysis, to focus on new stakeholder groups, to manage their so-called market-based assets in a different way, and to apply "new-to-marketing" methods to measure marketing performance (e.g., Srinivasan & Hanssens, 2009; Srivastava, Shervani, & Fahey, 1998). As illustrated by Keller and Lehmann (2003), marketing investments are ultimately designed to create shareholder value by changing the customer mind-set in a way that enhances market performance. Thus, efforts to quantify the impact of marketing expenditures and especially measures such as advertising efficiency that assesses the degree to which a company *generally* makes effective expenditures are worthwhile.

Following the signaling and spillover hypothesis (e.g., Joshi & Hanssens, 2010), investors should anticipate and value advertising effects. Fittingly, research has provided ample evidence of the effects of advertising intensity on investor behavior (e.g., Eng & Keh, 2007; Fosfuri & Giarratana, 2009; Joshi & Hanssens, 2010; Luo & de Jong, forthcoming; McAlister, Srinivasan, & Kim, 2007; Rao, Agarwal, & Dahlhoff, 2004; Simpson, 2008; Srinivasan, Pauwels, Silva-Risso, & Hanssens, 2009). However, leveraging these effects is subject to restrictions. The firm's resources are limited while excessive advertising spending would negatively affect the firm's bottom line. Nevertheless, there is a paucity of research on whether investors also monitor the economic use of resources when generating these advertising effects (Luo & Donthu, 2006). Therefore, it is less clear to what extent the ability to advertise products and services economically affects the valuation of a company. Do company outsiders distinguish between different levels of advertising efficiency? Do investors reward companies that use their resources economically? Or, in the words of Hanssens et al.(2009, p. 115), "If marketing resources are used well, will that trickle down to the capital markets?".

The aim of this study is to introduce and to apply the conceptual framework and the empirical methods to answer these questions. For that purpose, we present a two-step approach linking advertising measures to investor behavior. First, we relate advertising spending to its most notable outcomes – brand value, sales, and sales growth – thereby estimating each company's ability to advertise efficiently. In the second step of the analysis, we use this efficiency estimate to predict each company's stock market performance as an unbiased measure of investor behavior. Unlike competing ways of operationalizing attitudes and behavior, stock returns are free of measurement errors and circumvent scaling problems. They also present an aggregate measure on a company across the borders of the country markets it is active in.

The research methods we apply in our analyses are twofold: to scrutinize the efficiency of advertising activities, we apply data envelopment analysis (DEA), which has become an increasingly visible method for efficiency analyses (e.g., Grewal, Gopalkrishnan, Kamakura, Mehrotra, & Sharma, 2009; Luo, 2008; Luo & Homburg, 2007) and has also been used in the context of advertising research (e.g., Luo & Donthu, 2006; Pergelova, Prior, & Rialp, 2010). An additional purpose of the chapter is to illustrate how DEA can be used for the purpose of assessing the efficiency of marketing expenditures. DEA allows us to transform multiple inputs and outputs – measured on different scales (e.g., monetary vs. nonmonetary values) – into a single performance measure (Bhargava, Dubelaar, & Ramaswami, 1994; Donthu, Hershberger, & Osmonbekov, 2005; Luo & Donthu, 2001). To study stock market performance, we draw on the widely accepted Fama–French financial benchmark model that recognizes the random walk nature of stock prices and, unlike raw stock returns, provides a suitable basis for evaluating marketing performance effects over time (Srinivasan & Hanssens, 2009).

Theoretical Background

Conceptual Model

The evaluation of the economic use of advertising budgets (i.e., advertising efficiency) is a nontrivial task, because it entails considering short- and long-term effects, as advertising can have both immediate and lagged effects. As Ambler (2008, p. 7) notes, "[…] efficiency is indeed important and less productive activities should be dropped in favor of more productive, but here too we need to consider the longer term as well as the immediate." Accordingly, we define advertising efficiency as the ratio of short- and long-term marketing outputs (specifically sales, sales growth, and brand value) to advertising spending (specifically television, radio, magazine, newspaper, and outdoor advertising expenditures). Furthermore, relating effects to any type of marketing communication activities requires considering (mediated) interdependencies in a complex system of causes and effects. To describe these effects, the marketing literature commonly uses a four-stage functional model as discussed by Rust, Ambler, Carpenter, Kumar, and Srivastava (2004); Srivastava et al. (1998); and Wang, Zhang, and Ouyang (2009).

The initial stage (stage 1, Fig. 1) involves investments in communication activities. These investments act as stimuli, which the respondents subsequently process (stage 2, Fig. 1) in a way that reflects the customer mind-set. Typical elements of the customer mind-set are perceptions, beliefs, and attitudes (e.g., Vakratsas & Ambler, 1999). From a company perspective, these elements resemble market-based assets such as brand equity or corporate reputation. Market-based assets positively affect product market outcomes (stage 3, Fig. 1) by increasing cash flows (e.g., Srivastava et al., 1998), for example, through price premiums (e.g., Gruca & Rego, 2005) or higher retention rates (e.g., Luo & Donthu, 2001). Finally, product market outcomes translate into shareholder value (stage 4, Fig. 1) – the final target of marketing activities (e.g., Srinivasan & Hanssens, 2009). Numerous studies have examined the relationship between specific elements of the effect chain

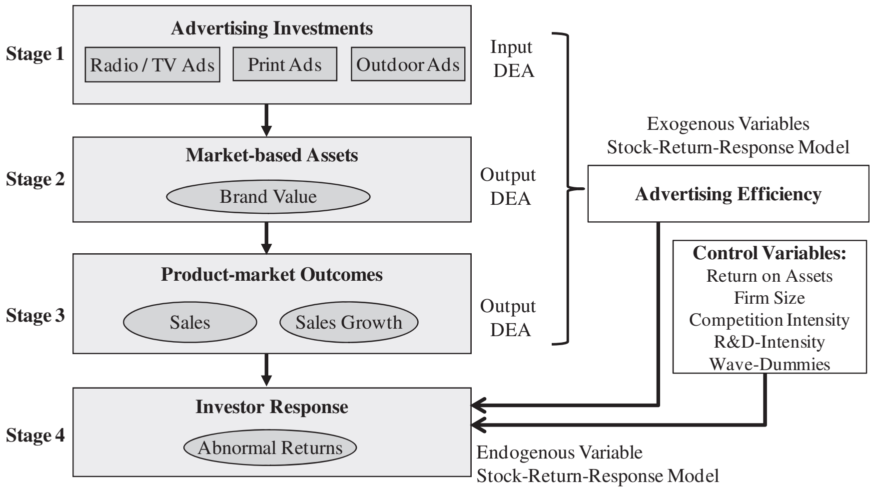

Fig. 1. The Advertising Chain of Effects as a Research Framework.

and have provided support for the model's central tenets (e.g., Fornell, Mithas, Morgeson, & Krishnan, 2006; Gruca & Rego, 2005; Jacobson & Mizik, 2009; Kerin & Sethuraman, 1998; Luo & Homburg, 2007; Madden, Fehle, & Fournier, 2006; Mizik & Jacobson, 2008; Morgan & Rego, 2009; Raithel, Sarstedt, Scharf, & Schwaiger, 2011).

Extending this perspective, we provide a holistic consideration of the model's four stages by using efficiency measures – obtained by means of DEA – as input for stock return response modeling (Fig. 1). The DEA considers stage 1 as the input factor and stages 2 and 3 as output factors. This allows the simultaneous examination of attitudinal (stage 2) and behavioral outcomes (stage 3) as postulated by Vakratsas and Ambler (1999). The result of this primary analysis is an estimation of advertising efficiency. Subsequent analyses will then test its influence on shareholder value (stage 4) by means of stock return response modeling.

Linking Advertising Efficiency to Stock Returns

According to Srivastava et al. (1998), shareholder value – defined as the net present value of future cash flows – is primarily driven by (1) higher cash flows, (2) acceleration of cash flows, (3) reduced volatility and sensitivity of cash flows, and (4) higher residual value of cash flows occurring in the distant future.

These four financial drivers can be interpreted as being preceded by, inter alia, advertising efficiency (Rao et al., 2004) because efficiency gains make it possible to meet the same goal criteria with a smaller amount of resources (i.e., advertising expenditures). Consequently, managers can use spare liquid assets to make other value-generating investments (Luo & Donthu, 2006). Alternatively, companies can use efficiency gains to generate more valuable market-based assets, triggering multiplier effects and, thus, improving future advertising efficiency (Mittal, Anderson, Sayrak, & Tadikamalla, 2005).

On the basis of this framework, literature discusses two basic mechanisms that explain why advertising affects stock prices directly: signaling and spillover effects. According to signaling theory, a company's advertising is a signal of financial well-being and a promise of growth opportunities to potential investors (e.g., Joshi & Hanssens, 2010; Karrh, 2004). As investors look for signals that reduce their uncertainty about the firm's future financial performance and expected stock returns, they try to reduce their personal risk by buying and holding stocks with higher perceived "quality." Assuming investors are at least partially aware of the aforementioned relation between advertising efficiency and financial performance, increases in advertising efficiency can be seen as an improvement in company-specific capabilities according to the resource-based view (RBV) and thus an improvement in a company's "quality." Research in the field of the RBV argues that competitive advantage does not arise from resources directly but rather comes from a company's processes and abilities (i.e., capabilities) to capitalize on those resources (Barney, 1991; Sirmon, Hitt, & Ireland, 2007). Accordingly, the ability to transform resources by means of communicational capabilities is more important than driving performance through absolute resource levels directly (DeSarbo, Di Benedetto, & Song, 2007; Vorhies, Morgan, & Autry, 2009).

Spillover effects arise because advertising activities – even if directed at a company's consumer – also influence investors as they are consumers themselves and simply cannot elude basic psychological phenomena. Assuming advertising triggers the same psychological effects as with consumers and causes cognitive as well as affective information processing (Vakratsas & Ambler, 1999); advertising may also create attention, awareness, and preference for a stock even if this is not done intentionally. Consequently, it is plausible to assume that this spillover effect is even more pronounced when a firm advertises efficiently. This spillover effect is important because investors are also "informed through actual exposure" (Joshi & Hanssens, 2010, p. 27) because they frequently do not have insights into actual advertising budgeting, advertising strategies, and advertising campaign execution plans.

As stock prices always fully reflect all available information (Fama, 1970), we expect investors to react to changes in advertising efficiency: they update their expectations about future cash flows in response to positive (negative) changes in advertising efficiency and buy or hold (do not buy or sell) stocks. It is this notion that the following empirical study seeks to explore.

Methodology and Model Formulation

Modeling Advertising Efficiency

To measure advertising efficiency, we use DEA, which has become an increasingly visible approach in marketing and related disciplines to model efficiency (e.g., Kamakura, Lenartowicz, & Ratchford, 1996; Mittal et al., 2005). Originally developed by Charnes, Cooper, and Rhodes (1978), DEA is a linear programming approach. It identifies a piecewise-linear Pareto frontier, defining the most efficient transformation of inputs into outputs. This allows the efficiency of the units (here, firms) to be measured in terms of this efficiency frontier (Boles, Donthu, & Lohtia, 1995). The efficiency frontier constitutes linear combinations of multiple observations. Consequently, a firm's activities are efficient if there is no linear combination of other firms that generates a higher output with a given input (Charnes et al., 1978). The efficiency measure then takes the value of 1. The degree of inefficiency is determined by measuring the distance of a firm to the Pareto frontier, which can range between 0 and 1 (Cooper, Seiford, & Zhu, 2004).

The setup of the DEA's model requires deciding on two generic modeling options, namely, constant vs. variable return-to-scales property, as well as output vs. input-based modeling. Owing to the constant returns-to-scale assumption not being tenable in advertising contexts, we apply Banker, Charnes, and Cooper's (1984) variable returns-to-scales model (Büschken, 2007). Moreover, we have to choose between input- and output-based models: while input-based models seek to minimize inputs for a fixed output, the opposite is true of optimization using output-based models (Cooper et al., 2004). Usually, choosing between these model types depends on whether the model's variables can be influenced by decision-makers (Coelli, Rao, O'Donnell, & Battese, 2005). In the present study, decision-makers are firms responsible for advertising budgets, that is, they can influence all the input variables, namely, the advertising spending levels, directly. By determining the budgets, they seek to influence the output variables positively. Consequently, the application of an output-based model is deemed appropriate.

As we aim to analyze advertising efficiency over time (Grifell-Tatjé & Lovell, 1997), we extend the initial cross-sectional DEA. Specifically, to adjust for time-varying effects – for example, shifts in the "best-practice frontier" – we apply the Malmquist approach (Caves, Christensen, & Diewert, 1982; Luo & Donthu, 2006). Originally, Malmquist (1953) developed dynamic models to assess general economic activities' total factor productivity over time. The Malmquist index allows advertising efficiency to be decomposed into technology changes – that is, movements of the efficiency frontier – and firm-specific efficiency changes (Tone, 2004). Consequently, the composite score of advertising efficiency (AE) for firm i at period $t+1$ can be expressed as follows (Chen & Ali, 2004):

$$
\begin{aligned}
\mathrm{AE}_i^{t+1}(\boldsymbol{y}^{t+1},\boldsymbol{x}^{t+1},\boldsymbol{y}^t,\boldsymbol{x}^t) &= \left[\frac{D_i^t(\boldsymbol{y}^{t+1},\boldsymbol{x}^{t+1})}{D_i^t(\boldsymbol{y}^t,\boldsymbol{x}^t)}\cdot\frac{D_i^{t+1}(\boldsymbol{y}^{t+1},\boldsymbol{x}^{t+1})}{D_i^{t+1}(\boldsymbol{y}^t,\boldsymbol{x}^t)}\right]^{1/2} \\
&= \underbrace{\frac{D_i^{t+1}(\boldsymbol{y}^{t+1},\boldsymbol{x}^{t+1})}{D_i^t(\boldsymbol{y}^t,\boldsymbol{x}^t)}}_{\text{change of efficiency}}\cdot\underbrace{\left[\frac{D_i^t(\boldsymbol{y}^{t+1},\boldsymbol{x}^{t+1})}{D_i^{t+1}(\boldsymbol{y}^{t+1},\boldsymbol{x}^{t+1})}\cdot\frac{D_i^t(\boldsymbol{y}^t,\boldsymbol{x}^t)}{D_i^{t+1}(\boldsymbol{y}^t,\boldsymbol{x}^t)}\right]^{1/2}}_{\text{change of technology}} \qquad (1)
\end{aligned}
$$

where $D_i^t(\boldsymbol{y}^{t+1}, \boldsymbol{x}^{t+1})$, $D_i^{t+1}(\boldsymbol{y}^{t+1}, \boldsymbol{x}^{t+1})$, $D_i^t(\boldsymbol{y}^t, \boldsymbol{x}^t)$, and $D_i^{t+1}(\boldsymbol{y}^t, \boldsymbol{x}^t)$ are relative efficiencies estimates (distance measures) by four linear programming models ($\boldsymbol{x}$ represents the input vector and $\boldsymbol{y}$ represents the output vector). An improvement (deterioration) is indicated by values greater (smaller) than 1. In the next step, this aggregated efficiency measure is used as explanatory variable in the stock return response model.

Modeling Financial Performance Effects

Using the results of the Malmquist approach (i.e., advertising efficiency, hereafter AE), we apply stock return response modeling (Srinivasan & Hanssens, 2009) in the next step of the analysis (stage 4, Fig. 1). This modeling approach allows us to examine if improvements (or deteriorations) in advertising efficiency are incrementally – that is, beyond current accounting performance – value relevant and lead to higher (lower) abnormal stock returns.

Stock return models are based on changes in stock prices, which provide an unbiased estimate of shareholder value in terms of expected future cash flows (Mizik & Jacobson, 2003). To assure comparability of stock returns across different firms, we furthermore have to account for a multitude of risk factors underlying stock price movements. Neglecting these potential sources of risk would misguide investors, leading them to prefer riskier stocks that do not achieve abnormal returns (Aksoy, Cooil, Groening, Keiningham, & Yalcin, 2008).

Recently, marketing research has considered the efforts made in the finance discipline to capture these risk differences across stocks (Srinivasan & Hanssens, 2009). More precisely, the classical capital asset pricing model (CAPM) has been extended to include two additional risk factors for stock returns, which are estimated as follows (E.F. Fama & French, 1993):

$$E(\mathrm{R}_{it}) = \mathrm{R}_{rf,t} + \beta_{\mathrm{i}} \cdot \mathrm{RMRF}_t + s_i \cdot \mathrm{SMB}_t + h_i \cdot \mathrm{HML}_t + \varepsilon_{it} \tag{2}$$

where $E(\mathrm{R}_{it})$ represents the expected return of firm i in period t; $\mathrm{R}_{rf,t}$ is the risk-free rate of return in period t; RMRF_t is calculated as the difference between the market rate of return and the risk-free rate of return in period t, which captures the variability of a firm's stock compared to the total market's average variability (Sharpe, 1964). In addition, this factor compensates for fluctuations and varying economic market conditions such as exchange rates and changes in energy prices (Brealey, Myers, & Allen, 2008). SMB_t (small minus big) is a size factor correcting the differences between returns of small stocks (those with low market capitalization) and returns of big stocks (those with high market capitalization) in period t. HML_t (high minus low) corrects the difference between high book-to-market stocks' return on a value-weighted portfolio and low book-to-market stocks' value-weighted portfolio in period t.[1]

To consider the stocks' above-mentioned risk differences in our focal model, we adjust raw returns by expected returns, which are based on the Fama–French risk factors. The resulting performance measure is the abnormal return, which is defined as the difference between the raw return and the expected return:

$$\mathrm{AR}_{it} = \mathrm{R}_{it} - E(\mathrm{R}_{it}) \tag{3}$$

As we use monthly returns to estimate expected returns, the next step is to aggregate these returns to an annual level, given that the advertising efficiency measures are available annually. To date, there is no consensus on how to aggregate monthly returns to annual returns. Finance literature proposes different options for aggregation (e.g., Barber & Lyon, 1997). Following Jacobson and Mizik's (2009) notion, we base our analysis on the commonly used buy-and-hold abnormal returns (BHAR_{it}), as they reflect the abnormal return an investor would earn by holding the stock for that period:[2]

$$\mathrm{BHAR}_{it} = \mathrm{R}_{it} - E(\mathrm{R}_{it}) \text{ with } \mathrm{R}_{it} = \prod\nolimits_{m=1}^{12}(1 + \mathrm{R}_{itm}) \text{ and } E(\mathrm{R}_{it}) = \prod\nolimits_{m=1}^{12}(1 + E(\mathrm{R}_{itm})) \tag{4}$$

Model Formulation

To assess the value relevance of advertising efficiency, we specify the following valuation model and augment it, in line with accounting, finance, and marketing literature, with several control variables:

$$\begin{aligned}\mathrm{AR}_{it} = \alpha + \beta_1 \cdot \Delta\mathrm{AE}_{it} + \beta_2 \cdot \Delta\mathrm{RoA}_{it} + \beta_3 \cdot \Delta\mathrm{R\&D}_{it} + \beta_4 \cdot \Delta\mathrm{Size}_{it} \\ + \beta_5 \cdot \Delta\mathrm{HHI}_{it} + \sum_{w=1}^{2} \gamma_w \cdot \mathrm{Wave}_w + \varepsilon_{it}\end{aligned} \tag{5}$$

AR_{it} represents the abnormal returns of firm i in period t (see Eq. (3)); $\Delta\mathrm{AE}_{it}$ is the change in the advertising efficiency of firm i between period $t-1$ and period t; and $\Delta\mathrm{RoA}_{it}$, $\Delta\mathrm{R\&D}_{it}$, $\Delta\mathrm{Size}_{it}$, $\Delta\mathrm{HHI}_{it}$, and Wave_w are considered control variables that are well known to impact abnormal returns. Controlling for changes in these factors between $t-1$ and t allows us to calibrate the extent to which changes in advertising efficiency contribute new information that explains stock returns. To operationalize the control variables, we revert to established measures that have been routinely used in prior research (Table 1). This includes the consideration of wave dummies (Wave) to account for potential time period effects in our variables.

Table 1. Operationalizations of Control Variables.

Variable	Description	Measurement	References
RoA	Return on assets	Ratio of the operating income (i.e., income before extraordinary items) to the total assets	Ferreira and Laux (2007); Jacobson and Mizik (2009); Raithel et al. (2011)
R&D	Research and development intensity	Ratio of R&D expenditures to the firm's total assets	Kim and Lyn (1990); Lev and Sougiannis (1996); Raithel et al. (2011); Sorescu et al. (2007); Srinivasan et al. (2009)
Size	Firm size	Differences in the natural logarithm of the number of employees	Luo and Bhattacharya (2006); Luo and Donthu (2006); Luo and Homburg (2007); Mizik and Jacobson (2008); Morgan and Rego (2009); Raithel et al. (2011)
HHI	Market competition intensity	Herfindahl–Hirschmann index	Curry and George (1983); Montgomery and Wernerfelt (1988); Rao et al. (2004)

As findings from empirical capital market research suggest that relations between stock returns and explanatory variables are often nonlinear in nature (e.g., Hiemstra & Jones, 1994; Ng, 2005), we also evaluate nonlinear functional relations to assess the model's robustness.

DATA AND MEASURES

In our analysis, we consider firms listed in the Interbrand rankings as "best global brands" between 2004 and 2007 (Interbrand, 2009), selecting those covered during the entire examination period. Furthermore, to ensure comparability, we excluded any outliers, that is, companies with abnormally low advertising budgets and companies that primarily focus on nontraditional communication channels (e.g., social media and online banner advertisements) as related information is not captured in our data. The resulting dataset comprises 61 companies representing 20 different industries according to the Standard Industrial Classification (SIC).

Using data from TNS Media Intelligence's AdSpender, we considered three general traditional types of advertising expenditures (Büschken, 2007; Luo & Donthu, 2005) as input for the DEA: broadcast, print, and outdoor. Specifically, a firm's advertising expenditures on broadcast media equal the sum of the television and radio expenditures, which include network television spots on ABC, CBS, FOX, NBC, PAX, MNTV, and CW, television spots on 600 plus stations in the top 100 U.S. markets, and radio spot spending on 4,000 stations in 225 U.S. markets. The advertising spending on print media equals the sum of newspaper, magazine, and other expenditures, which include expenditure on advertising in more than 350 consumer magazines and three national newspapers: the *New York Times*, *USA Today*, and the *Wall Street Journal*. Expenditures on outdoor media comprise the total expenditures on more than 200 U.S. markets.

Cobb-Walgren, Ruble, and Donthu (1995) have shown that, as an aggregated measure of communication activities, a larger advertising budget positively influences consumers' brand attitude. As advertising substantially contributes to the formation of brand knowledge, familiarity, recognition, and attitudes (e.g., Campbell & Keller, 2003), we use brand value as first output variable in the DEA (e.g., Madden et al., 2006; McAlister et al., 2007; Rust et al., 2004; Wang et al., 2009). In line with previous studies, we apply the Interbrand Brand Value index (e.g., Barth, Clement, Foster, & Kasznik, 1998; Chu & Keh, 2006; De Beijer, Dekimpe, Dutordoir, & Verbeeten, 2008; Madden et al., 2006). Interbrand uses a two-stage approach to estimate brand value: a combination of psychographic factors (brand strength) and projected profits results in the brand value (see Kerin and Sethuraman, 1998, for a detailed overview). To account for potential endogeneity issues, we computed idiosyncratic brand values by adjusting for lagged brand values, net income, advertising spending, and research and development intensity, using the procedure suggested by Chu and Keh (2006).

The same procedure was used to adjust sales (i.e., the second output variable in the DEA) by carryover effects and relevant explanatory variables such as lagged advertising expenditures, research and development intensity, firm size, focus of the firm, and market competition intensity. Data on sales, research and development spending, firm size, the focus of the firm, and industry competition were obtained from Datastream. We supplemented missing values based on annual reports.

The third output variable in the DEA is sales growth, which we computed as the compounded annual sales growth rate over the previous three years based on information in Datastream (Luo & Donthu, 2006; Rao et al., 2004).

Finally, we obtained financial data (capital market data as well as accounting data) and industry-specific data for the considered firms from Datastream. Additionally, we used annual reports of the considered firms to complete any missing data. The data source for the calculation of expected returns generated by means of Fama–French model was Kenneth French's web site.[3]

Data Analysis and Results

Advertising Efficiency by Means of DEA

In the first step of the analysis, we carried out a DEA using the EMS software package (Scheel, 2004). Before determining benchmark units for the DEA, we carried out an outlier analysis based on the super efficiencies method (Banker & Chang, 2006). As a result, we had to exclude five companies from the analysis, leaving us with a sufficiently large database of 56 companies, which is clearly above the minimum required sample size (Dyson et al., 2001). We examined the resulting database by means of sensitivity analyses to test the efficiency values' robustness regarding different input–output combinations. These analyses clearly provide support for the results' robustness.

Table 2 summarizes static and dynamic efficiency coefficients based on mean-centered values. In total, between 11 and 17 companies were efficient in the period under consideration. The mean efficiency values (MEV) range between 42.8% (2005) and 50.0% (2004), which implies that the average company's outputs could have been realized at a level of 42.8–50.0% of advertising budgets by an efficient reference company.

Table 2. Descriptive Statistics of Advertising Efficiency (Static and Dynamic).

	2004			2005			2006			2007		
	MIN	MAX	MEV	MIN	MAX	MEV	MIN	MAX	MEV	MIN	MAX	MEV
Static efficiency in %	1.5	100.0	50.0	1.1	100.0	42.8	2.3	100.0	44.0	1.9	100.0	43.2
AE (dynamic)	–	–	–	0.1	2.7	1.0	0.1	2.5	1.2	0.2	3.3	1.0
N		56			56			56			56	
among them efficient		17			17			14			11	

Owing to the wide range of industries considered, the efficiency gaps might be inflated compared to an industry-specific analysis. We addressed this drawback of the static efficiency analysis by transforming static measures into dynamic ones, using the Malmquist index. Consequently, the values of advertising efficiency now mirror relative improvements over time rather than on absolute efficiency levels (Luo & Donthu, 2006).

Results of the Stock Return Response Model

Next, we present the estimation results of the variables included in the model described in Eq. (3). We illustrate results based on the operationalization of financial performance by means of BHAR according to the Fama–French three-factor model. Table 3 provides the descriptives and correlations of the variables used in the analysis.

To assess advertising efficiency's incremental information content and value relevance, we regress (risk-adjusted) abnormal returns on ΔAE, ΔRoA, ΔR&D, ΔSize, ΔHHI, as well as two wave dummies (Wave 04/05 and Wave 05/06). For model estimation, we use generalized linear modeling because this procedure allows for a flexible modulation of error structure and for efficient parameter estimation (Greene, 2008).[4] We rule out potential autocorrelation problems as we model all variables in changes rather than in levels (Mizik & Jacobson, 2009). Accordingly, the Durbin–Watson test value (1.99) indicates a negligible level of autocorrelation. To ensure that we can estimate the model with a pooled data set, we apply Chow's (1960) poolability test, which reveals that the analysis of pooled data is unproblematic. Lastly, multicollinearity is not a problem, as the variance inflation factors clearly range below the commonly suggested threshold with values between 1.01 and 1.39. Table 4 illustrates the results for BHAR, which are based on the three-factor model as well as linear and nonlinear (quadratic) relations, including a baseline model with omitted ΔAE.[5]

Table 3. Descriptives and Correlations of Considered Variables.

	Mean	Std. Dev.	BHAR	ΔAE	ΔRoA	ΔR&D	ΔSize	ΔHHI
BHAR	0.033	0.180	1.000					
ΔAE	1.154	0.893	0.227**	1.000				
ΔRoA	−0.616	4.285	0.265**	0.066	1.000			
ΔR&D	0.007	0.078	−0.081	0.041	0.031	1.000		
ΔSize	4.917	0.469	−0.145	0.076	0.000	−0.090	1.000	
ΔHHI	0.026	0.014	0.150	0.010	0.186*	0.008	−0.018	1.000

*Significance at 5% level. **Significance at 1% level.

Table 4. Summary of Results for Two Types of Relations ($N = 168$).

Dependent variable: Buy-and-Hold Abnormal Stock Return (BHAR)	Baseline Model		Linear Model		Quadratic Model	
	Est. [SE][a]	Wald-χ^2	Est. [SE]	Wald-χ^2	Est. [SE]	Wald-χ^2
Intercept	−0.197 [0.126]	2.443	−0.156 [0.128]	1.478	−0.162 [0.123]	1.712
Wave 04/05	0.297 [0.162]	3.374	0.325 [0.166]	3.846*	0.304 [0.167]	3.326
Wave 05/06	0.340 [0.199]	2.926	0.183 [0.190]	0.927	0.221 [0.183]	1.453
ΔAE	–	–	0.247 [0.117]	4.435*	−0.115 [0.185]	0.390
ΔAE × ΔAE	–	–	–	–	0.384 [0.169]	5.175*
ΔRoA	0.149 [0.049]	9.106**	0.155 [0.047]	10.965**	0.146 [0.047]	9.633**
ΔR&D	−0.111 [0.072]	2.380	−0.123 [0.064]	3.696	−0.130 [0.067]	3.755
ΔSize	−0.163 [0.102]	2.575	−0.182 [0.102]	3.194	−0.182 [0.102]	3.202
ΔHHI	0.115 [0.093]	1.549	0.112 [0.095]	1.417	0.118 [0.094]	1.585
Overall model fit						
−2lnL[b,c]	441.554		431.164		426.876	
Likelihood ratio-χ^2	144.801**		135.805**		132.258**	
AIC[c,d]	457.554		449.163		446.875	
AICc[c,e]	458.495		450.348		448.332	
BIC[c,f]	482.255		476.952		477.751	
CAIC[c,g]	490.255		485.952		487.751	
Adj. R^2	0.072		0.124		0.143	
Comparison vs. baseline model						
Adj. R^2 change			0.052		0.071	
Δ2lnL statistic			10.390**		14.678**	
Comparison vs. linear model						
Adj. R^2 change					0.019	
Δ2lnL statistic					4.288*	

*Significance at 5% level. **Significance at 1% level.
[a]Est., Estimate; SE, Standard error.
[b]2lnL, -2*log-likelihood.
[c]Smaller-is-better format.
[d]AIC, Akaike Information Criterion.
[e]AICc, corrected Akaike Information Criterion.
[f]BIC, Bayes Information Criterion.
[g]CAIC, Consistent Akaike Information Criterion.

The results show that the focal model's adjusted R^2 values are 0.124 (linear model) and 0.143 (quadratic model), respectively. Thus, compared with the baseline model, both models exhibit increased values, underlining advertising efficiency's role in predicting abnormal returns. The significant log-likelihood differences (baseline vs. focal models) support this notion. The comparison of the model selection criteria reveals inconsistent results, as the log-likelihood difference, the Akaike Information Criterion (AIC), and the corrected AIC (AICc) favor the nonlinear relationship, whereas consistent AIC (CAIC) and Bayes Information Criterion (BIC) indicate that the linear model is more appropriate. This indicates that the goodness-of-fit is only moderately higher in the quadratic model than in the linear one.

Most importantly, the model results clearly support this study's main theme, as advertising efficiency is found to exert a significant positive influence on abnormal stock returns across the two models. ΔAE's influence is pronounced, showing that investors clearly value the *efficient* creation of market-based assets through advertising activities.

Regarding the control variables, ΔRoA correlates positively with abnormal returns, indicating that information related to accounting measures is relevant for future cash flows. Moreover, the results show that there is a negative, albeit nonsignificant, relation between changes in R&D intensity (ΔR&D) and abnormal returns. At first sight, this result appears surprising. However, investors might only value changes in R&D intensity with a time lag (Dinner, Mizik, & Lehmann, 2009). In addition, the sample structure, which is quite heterogeneous regarding industry sectors, could serve as another possible explanation because R&D intensity varies considerably between the sectors. Moreover, we find a negative, albeit nonsignificant, relation between firm size and financial performance, which is in line with Morgan and Rego's (2009) findings. Lastly, as expected, increasing market competition intensity is accompanied by a positive performance effect.[6]

DISCUSSION

The answer to the opening question: "If marketing resources are used well, will that trickle down to the capital markets?" (Hanssens et al., 2009, p. 115) appears to be "yes," as this study has shown that capital market players regard the efficient use of advertising budgets as valuable information. An input–output relation superior to that of one's competitors is not only important for resource allocation decisions but is also rewarded by the capital market in terms of abnormal returns.

Our study leads to an array of implications regarding the analysis of advertising efficiency: even though all companies evaluated are among the ''best global brands'' and may therefore be considered top performers in the market, DEA results revealed considerable differences in their efficiency levels. Of the 56 companies studied, between 39 and 45 were inefficient in any given year. Looking at the static efficiency values of the years observed, there is an upward potential of up to 57% (year 2005). Given an average budget of USD 184 million for the three focal media in the U.S. market, the potential improvement could be as high as USD 105 million per year and company. In accordance with the relevant literature, we regard this as overspending (Luo & Donthu, 2005, 2006), pointing to shortcomings in the creation, execution, or appropriate media allocation. This underlines the importance of optimizing media selection and advertising design as well as optimizing the allocation of advertising budgets. It is advisable to assess not only the effectiveness of any marketing activity but also its efficiency. Simultaneously, our results show that there is room for efficiently advertising companies to differentiate themselves and to stand out from the majority of companies. Companies should therefore communicate their efficiency gains toward investors. The transparency generated should assist investors to even better process relevant information, therefore helping individuals regard marketing measures as value-enhancing investments rather than as cost factors.

Given our results, continuous tracking of advertising effectiveness, and particularly of advertising efficiency, seems necessary. Our analyses also present the adequate method to do so: DEA is a viable, scientifically well-grounded method of monitoring a company's advertising efficiency in relation to that of its key competitors. Fast moving consumer goods companies could benefit greatly from integrating DEA results into their marketing dashboards, as companies spend large amounts on advertising in this industry sector. This would allow for a constant assessment of marketing's contribution to business performance instead of solely relying on intermediate consumer-based measures such as brand awareness and image, which are weakly related to business success (Binet & Field, 2007a). However, this requires marketers being cognizant of financial outcomes and developing a better understanding of marketing-related levers of shareholder value (McDonald & Mouncey, 2009) instead of merely thinking in terms of awareness measures (Binet & Field, 2007b). Considering that recent research shows that the capabilities of marketing departments usually relate positively to business performance (Verhoef & Leeflang, 2009), increasing levels of financial literacy among marketers significantly contribute to an organization's bottom line.

Although we have found support for our main research theme, we have to keep in mind that investors cannot react immediately to changes in advertising efficiency. Capital market participants have to process information indicating an improvement in a company's competencies and have to adapt their estimations continuously. This underlines the relevance of econometric modeling in general and stock return response modeling in particular, which allows for the consideration of mid-term and time-lagged capital market reactions – unlike, for example, event studies.

Our study implies that managers should not limit their tactics to increasing market-based assets at any cost and raising advertising budgets if they wish to send a positive signal to investors and contribute to value enhancement. They should refrain from merely employing such methods because company specifics play a role in the creation of market-based assets: "[M]easures based solely on expenditures do not capture differences in success rates across firms and thus are suitable only for establishing the average degree of value relevance for a class of intangible investments but not for exploring firm-specific differences in the success of intangible investments" (Kimbrough & McAlister, 2009, p. 315). Keeping Binet and Field's (2007b) recent criticism in mind, we address this shortcoming by comparing investments in market-based assets and the corresponding outcomes to those of a firm's competitors.

Limitations and Further Research

Like any research work, this study is subject to limitations. First, we used a sample consisting of very large companies with strong brands calling for further research projects focusing on smaller companies. We would recommend extending the time horizon of future studies, as our data allowed an analysis within a period of only four years. Moreover, were the study conducted based on data from 2010 and beyond, it would be necessary to include Internet advertising expenditures as this medium began growing at a rapid rate beginning in about 2005 (Taylor, 2010).

The DEA in this study relied on brand value measures published by Interbrand. Although the reported drawbacks of this measurement model could be eliminated almost completely using regression-based reassessment, the results require validation with measures based on alternative measurement models. Recent studies, for example, rely on data from the Brand Asset Valuator by Young & Rubicam (e.g., Mizik & Jacobson, 2008) and from Harris Interactive's Equitrend (e.g., Rego, Billet, & Morgan, 2009).

Researchers might focus on other market-based assets (like corporate reputation, customer satisfaction, etc.) and simultaneously integrate different measures on the input and output sides, which DEA could handle without major problems. Examining the fit of a model considering brand value and, for instance, customer satisfaction simultaneously is very interesting because firms have usually multiple marketing goals. As brand value mirrors customer acquisition success, and customer satisfaction is an antecedent measure of customer loyalty and retention, this approach would be intriguing from a conceptual perspective as well. However, such a study would have to surmount the challenge of data availability.

NOTES

1. Carhart (1997) extends this model to a four-factor model by including a momentum factor, which adjusts the returns by ascertaining the difference between the average return on the two high-prior-return portfolios and the average return on the two low-prior-return portfolios in period t. However, this model has vastly been criticized and the influence of this momentum factor is still ambiguous (e.g., Bollerslev & Zhang, 2003; Subrahmanyam, 2005; Srinivasan & Hanssens, 2009).
2. We likewise carried out all analyses of cumulative abnormal returns (CAR) as well as compounded abnormal returns (CCAR) (Barber & Lyon, 1997; Fama & French, 1998; Jacobson & Mizik, 2009), which yielded similar results.
3. http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html
4. Following Stewart's (2009) notion and to check for robustness, we also estimated the model using linear mixed models as well as OLS with heteroskedasticity-consistent standard errors. These analyses did not yield different results.
5. We also estimated the model with cubic terms of advertising efficiency, but as this term does not explain any additional variance compared to the quadratic term, results remain almost identical. Therefore, we prescind from displaying these results in Table 4.
6. Additional analyses based on different schemes to calculate abnormal returns (the three-factor vs. the four-factor model as well as BHAR vs. CAR vs. CCAR) revealed extremely high correlations between the dependent measures, thus supporting the results' robustness.

ACKNOWLEDGMENT

The authors are grateful to Marko Sarstedt (Ludwig-Maximilians-University, Munich) for his helpful comments on earlier drafts of this chapter.

REFERENCES

Aksoy, L., Cooil, B., Groening, C., Keiningham, T. L., & Yalcin, A. (2008). The long-term stock market valuation of customer satisfaction. *Journal of Marketing*, *72*(4), 105–122.

Ambler, T. (2008). How important is marketing efficiency? *Marketing Review St. Gallen*, *25*(1), 4–7.

Banker, R. D., & Chang, H. (2006). The super efficiency procedure is for outlier identification, not for ranking efficient units. *European Journal of Operational Research*, *175*(1), 1311–1320.

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, *30*(9), 1078–1092.

Barber, B. M., & Lyon, J. D. (1997). Detecting long-run abnormal stock returns: The empirical power and specification of test statistics. *Journal of Financial Economics*, *43*(3), 341–372.

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, *17*(1), 99–120.

Barth, M. E., Clement, M. B., Foster, G., & Kasznik, R. (1998). Brand values and capital market valuation. *Review of Accounting Studies*, *3*(1/2), 41–68.

Bhargava, M., Dubelaar, C., & Ramaswami, S. (1994). Reconciling diverse measures of performance: A conceptual framework and test of a methodology. *Journal of Business Research*, *34*(2/3), 235–246.

Binet, L., & Field, P. (2007a). *Marketing in the era of accountability*. Henley-on-Thames: WARC.

Binet, L., & Field, P. (2007b). Measuring the right things. *International Journal of Market Research*, *48*(5), 545–546.

Boles, J. S., Donthu, N., & Lohtia, R. (1995). Salesperson evaluation using relative performance efficiency: The application of data envelopment analysis. *Journal of Personal Selling & Sales Management*, *15*(3), 31–49.

Bollerslev, T., & Zhang, B. Y. B. (2003). Measuring and modeling systematic risk in factor pricing models using high-frequency data. *Journal of Empirical Finance*, *10*(5), 533–558.

Brealey, R. A., Myers, S. C., & Allen, F. (2008). *Principles of corporate finance*. Boston, MA: McGraw-Hill.

Büschken, J. (2007). Determinants of brand advertising inefficiency – Evidence from the German car market. *The Journal of Advertising*, *36*(3), 51–73.

Campbell, M. C., & Keller, K. L. (2003). Brand familiarity and advertising repetition effects. *Journal of Consumer Research*, *30*(2), 292–304.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, *52*(1), 57–82.

Caves, D. W., Christensen, L. R., & Diewert, W. E. (1982). The economic theory of index numbers and the measurement of input, output and productivity. *Econometrica*, *50*(6), 1393–1414.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*(6), 429–444.

Chen, Y., & Ali, A. (2004). DEA Malmquist productivity measure: New insights with an application to computer industry. *European Journal of Operational Research*, *159*(1), 239–249.

Chow, G. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, *28*, 591–605.

Chu, S., & Keh, H. T. (2006). Brand value creation: Analysis of the interbrand-business week brand value ranking. *Marketing Letters*, *17*(4), 323–331.

Cobb-Walgren, C. J., Ruble, C. A., & Donthu, N. (1995). Brand equity, brand preference, and purchase intent. *Journal of Advertising*, *24*(3), 25–39.

Coelli, T. J., Rao, D. S. P., O'Donnell, C. J., & Battese, G. E. (2005). *An introduction to efficiency and productivity analysis*. New York, NY: Springer.

Cooper, W. W., Seiford, L. M., & Zhu, J. (2004). *Handbook on data envelopment analysis*. Boston, MA: Kluwer.

Curry, B., & George, K. D. (1983). Industrial concentration: A survey. *The Journal of Industrial Economics*, *31*(3), 203–255.

De Beijer, D., Dekimpe, M. G., Dutordoir, M., & Verbeeten, F. H. M. (2008). *The impact of brand value announcements on firm value*. SSRN Working Paper Series. Retrieved from http://ssrn.com/abstract=1344535. Accessed on 27 October 2010.

DeSarbo, W. S., Di Benedetto, C. A., & Song, M. (2007). A heterogenous resource based view for exploring relationships between firm performance and capabilities. *Journal of Modeling in Management*, *2*(2), 103–130.

Dinner, I. M., Mizik, N., & Lehmann, D. R. (2009). *The (unappreciated) value of marketing*. SSRN Working Paper Series. Retrieved from http://ssrn.com/abstract=1324815. Accessed on 27 October 2010.

Donthu, N., Hershberger, E., & Osmonbekov, T. (2005). Benchmarking marketing productivity using data envelopment analysis. *Journal of Business Research*, *58*(11), 1474–1482.

Dyson, R. G., Allen, R., Camanho, A. S., Podinovski, V. V., Sarrico, C. S., & Shale, E. A. (2001). Pitfalls and protocols in DEA. *European Journal of Operational Research*, *132*(2), 245–259.

Eng, L. L., & Keh, H. T. (2007). The effects of advertising and brand value on future operating and market performance. *Journal of Advertising*, *36*(4), 91–100.

Fama, E. F., & French, K. (1998). Market efficiency, long-term stock returns, and behavioral finance. *Journal of Financial Economics*, *49*(1), 283–306.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383–417.

Fama, E. F., & French, K. (1993). Common risk factors in the returns on stock and bonds. *Journal of Financial Economics*, *33*(1), 3–56.

Ferreira, M. A., & Laux, P. A. (2007). Corporate governance, idiosyncratic risk and information flow. *The Journal of Finance*, *62*(4), 951–989.

Fornell, C., Mithas, S., Morgeson, F. V., & Krishnan, M. S. (2006). Customer satisfaction and stock prices: High returns, low risk. *Journal of Marketing*, *70*(1), 3–14.

Fosfuri, A., & Giarratana, M. S. (2009). Masters of war: Rival's product innovation and new advertising in mature product markets. *Management Science*, *55*(2), 181–191.

Greene, W. H. (2008). *Econometric analysis*. Upper Saddle River, NJ: Pearson Prentice Hall.

Grewal, D., Gopalkrishnan, R. I., Kamakura, W. A., Mehrotra, A., & Sharma, A. (2009). Evaluation of subsidiary marketing performance: Combining process and outcome performance metrics. *Journal of the Academy of Marketing Science*, *37*(2), 117–129.

Grifell-Tatjé, E., & Lovell, C. A. K. (1997). A DEA-based analysis of productivity change and intertemporal managerial performance. *Annals of Operations Research*, *73*(1–4), 177–189.

Gruca, T. G., & Rego, L. L. (2005). Customer satisfaction, cash flow, and shareholder value. *Journal of Marketing*, *69*(3), 115–130.

Hanssens, D. M., Rust, R. T., & Srivastava, R. K. (2009). Marketing strategy and Wall Street: Nailing down marketing's impact. *Journal of Marketing*, *73*(4), 115–118.

Hiemstra, C., & Jones, J. (1994). Testing for linear and nonlinear granger causality in the stock price-volume relation. *The Journal of Finance*, *49*(12), 1639–1664.

Interbrand. (2009). Best global brands. Retrieved from http://www.interbrand.com/en/best-global-brands/best-global-brands-2008/best-global-brands-2008.aspx. Accessed on 27 October 2010.

Jacobson, R., & Mizik, N. (2009). *Assessing the value-relevance of customer satisfaction*. SSRN Working Paper Series. Retrieved from http://ssrn.com/abstract=990783. Accessed on 27 October 2010.

Joshi, A., & Hanssens, D. M. (2010). The direct and indirect effects of advertising spending on firm value. *Journal of Marketing*, *74*(1), 20–33.

Kamakura, W. A., Lenartowicz, T., & Ratchford, B. T. (1996). Productivity assessment of multiple retail outlets. *Journal of Retailing*, *72*(4), 333–356.

Karrh, J. A. (2004). Does advertising influence investors? Evidence and research propositions. *Journal of Current Issues and Research in Advertising*, *26*(2), 1–10.

Keller, K. L., & Lehmann, D. (2003). How do brands create value. *Marketing Management*, *12*(3), 27–31.

Kerin, R. G., & Sethuraman, R. (1998). Exploring the brand value-shareholder value nexus for consumer goods company. *Journal of the Academy of Marketing Science*, *26*(4), 260–273.

Kim, W. S., & Lyn, E. O. (1990). FDI theories and performance of foreign multinationals operating in the U.S.. *Journal of International Business Studies*, *21*(1), 41–54.

Kimbrough, M. D., & McAlister, L. (2009). Linking marketing actions to value creation and firm value: Insights from accounting research, commentaries and rejoinder to 'Marketing and firm value: Metrics, methods, findings, and future directions'. *Journal of Marketing Research*, *46*(3), 313–319.

Lehmann, D. R. (2004). Linking marketing to financial performance and firm value. *Journal of Marketing*, *68*(4), 73–75.

Lev, B., & Sougiannis, T. (1996). The capitalization, amortization, and value-relevance of R&D. *Journal of Accounting and Economics*, *21*(1), 107–138.

Luo, X. (2008). When marketing strategy first meets Wall Street: Marketing spendings and firms' initial public offerings. *Journal of Marketing*, *72*(4), 98–109.

Luo, X., & Bhattacharya, C. B. (2006). Corporate social responsibility, customer satisfaction, and market value. *Journal of Marketing*, *70*(4), 1–18.

Luo, X., & de Jong, P. J. (forthcoming). Does advertising spending really work? The intermediate role of analysts in the impact of advertising on firm value. *Journal of the Academy of Marketing Science*.

Luo, X., & Donthu, N. (2001). Benchmarking advertising efficiency. *Journal of Advertising Research*, *41*(6), 7–18.

Luo, X., & Donthu, N. (2005). Assessing advertising media spending inefficiencies in generating sales. *Journal of Business Research*, *58*(1), 28–36.

Luo, X., & Donthu, N. (2006). Marketing's credibility: A longitudinal investigation of marketing communication productivity and shareholder value. *Journal of Marketing*, *70*(4), 70–91.

Luo, X., & Homburg, C. (2007). Neglected outcomes of customer satisfaction. *Journal of Marketing*, *71*(2), 133–149.

Madden, T., Fehle, F., & Fournier, S. (2006). Brands matter: An empirical demonstration of the creation of shareholder value through branding. *Journal of the Academy of Marketing Science*, *34*(2), 224–235.

Malmquist, S. (1953). Index numbers and indifference surfaces. *Trabajos de Estadistica*, *4*, 209–242.

McAlister, L., Srinivasan, R., & Kim, M. (2007). Advertising, research, development, and systematic risk of the firm. *Journal of Marketing*, *71*(1), 35–48.

McDonald, M., & Mouncey, P. (2009). *Marketing accountability: How to measure marketing effectiveness* (Retrieved from http://www.management.org.in/Ebooks/Marketing%20Accountability.pdf. Accessed on 27 October 2010). London; Philadelphia, PA: Kogan Page.

Mittal, V., Anderson, E. W., Sayrak, A., & Tadikamalla, P. (2005). Dual emphasis and the long-term financial impact of customer satisfaction. *Marketing Science*, *24*(4), 544–555.

Mizik, N., & Jacobson, R. (2003). Trading off between value creation and value appropriation: The financial implications of shifts in strategic emphasis. *Journal of Marketing*, *67*(1), 63–76.

Mizik, N., & Jacobson, R. (2008). The financial value impact of perceptual brand attributes. *Journal of Marketing Research*, *45*(1), 15–32.

Mizik, N., & Jacobson, R. (2009). Financial markets research in marketing, commentaries and rejoinder to 'Marketing and firm value: Metrics, methods, findings, and future directions'. *Journal of Marketing Research*, *46*(3), 320–324.

Montgomery, C. A., & Wernerfelt, B. (1988). Diversification, Ricardian rents, and Tobin's q. *RAND Journal of Economics*, *19*(4), 623–632.

Morgan, N. A., & Rego, L. L. (2009). Brand portfolio strategy and firm performance. *Journal of Marketing*, *73*(1), 59–74.

Neff, J. (2005). ROI: The marketer's obsession. *Advertising Age*, *76*(25), S1–S2.

Ng, C. (2005). An empirical study on the relationship between ownership and performance in a family-based corporate environment. *Journal of Accounting, Auditing and Finance*, *20*(2), 121–146.

Pergelova, A., Prior, D., & Rialp, J. (2010). Assessing advertising efficiency: Does the internet play a role? *Journal of Advertising*, *39*(3), 39–54.

Raithel, S., Sarstedt, M., Scharf, S., & Schwaiger, M. (forthcoming). On the value relevance of customer satisfaction. Multiple drivers and multiple markets. *Journal of the Academy of Marketing Science*.

Rao, V., Agarwal, M. K., & Dahlhoff, D. (2004). How is manifest branding strategy related to the intangible value of a corporation? *Journal of Marketing*, *68*(4), 126–141.

Rego, L. L., Billet, M. T., & Morgan, N. A. (2009). Consumer-based brand equity and firm risk. *Journal of Marketing*, *74*(6), 47–60.

Rust, R. T., Ambler, T., Carpenter, G. S., Kumar, V., & Srivastava, R. K. (2004). Measuring marketing productivity: Current knowledge and future directions. *Journal of Marketing*, *68*(4), 76–89.

Scheel, H. (2004). Efficiency measurement system – Software. Retrieved from http://www.wiso.tu-dortmund.de/wiso/de/fakultaet/personen/institut/or/EXT-HOSC.html. Accessed on 27 October 2010.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium and conditions of risk. *The Journal of Finance*, *19*(3), 425–442.

Simpson, A. V. (2008). Voluntary disclosure of advertising expenditures. *Journal of Accounting, Auditing and Finance*, *23*(3), 403–436.

Sirmon, D. G., Hitt, M. A., & Ireland, R. D. (2007). Managing firm resources in dynamic environments to create value: Looking inside the black box. *Academy of Management Review*, *32*(1), 273–292.

Sorescu, A., Shankar, V., & Kushwaha, T. (2007). New product preannouncements and shareholder value: Don't make promises you can't keep. *Journal of Marketing Research*, *34*(3), 468–489.

Srinivasan, S., & Hanssens, D. M. (2009). Marketing and firm value: Metrics, methods, findings, and future directions. *Journal of Marketing Research*, *46*(3), 293–312.

Srinivasan, S., Pauwels, K., Silva-Risso, J., & Hanssens, D. M. (2009). Product innovations, advertising, and stock returns. *Journal of Marketing*, *73*(1), 24–43.

Srivastava, R. K., Shervani, T. A., & Fahey, L. (1998). Market-based assets and shareholder value: A framework for analysis. *Journal of Marketing*, *62*(1), 2–18.

Stewart, D. W. (Ed.) (2009). The role of method: Some parting thoughts from a departing editor*Journal of the Academy of Marketing Science*, *37*(4), 381–383.

Subrahmanyam, A. (2005). Distinguishing between rationales for short-horizon predictability in stock returns. *Financial Review*, *40*(1), 11–35.

Taylor, C. R. (Ed.) (2010). Editorial: Measuring return on investment from advertising: 'Holy grail' or necessary tool? *International Journal of Advertising*, *29*(3), 345–348.

Tone, K. (2004). Malmquist productivity index. In: W. W. Cooper, L. M. Seiford & J. Zhu (Eds.), *Handbook on data envelopment analysis* (pp. 203–227). New York, NY: Springer.

Vakratsas, D., & Ambler, T. (1999). How advertising works: What do we really know? *Journal of Marketing*, *63*(1), 26–43.

Verhoef, P. C., & Leeflang, P. S. H. (2009). Understanding the marketing department's influence within the firm. *Journal of Marketing*, *73*(2), 14–37.

Vorhies, D. W., Morgan, R. E., & Autry, C. W. (2009). Product-market strategy and the marketing capabilities of the firm: Impact on market effectiveness and cash flow performance. *Strategic Management Journal*, *30*(12), 1310–1334.

Wang, F., Zhang, X. P., & Ouyang, M. (2009). Does advertising create sustained firm value? The capitalization of brand intangible. *Journal of the Academy of Marketing Science*, *37*, 130–143.

APPENDIX

Table A1. Herfindahl Indices 2003–2007.

Sector	2003	2004	2005	2006	2007
Aerospace & defense	0.0488	0.0475	0.0455	0.0449	0.0422
Automobiles & parts	0.0326	0.0322	0.0316	0.0298	0.0278
Banks	0.0098	0.0102	0.0110	0.0108	0.0105
Beverages	0.0269	0.0257	0.0246	0.0231	0.0227
Financial services	0.0300	0.0246	0.0192	0.0231	0.0192
Food producers	0.0186	0.0168	0.0165	0.0156	0.0151
General industrials	0.0479	0.0437	0.0397	0.0412	0.0371
General retailers	0.0318	0.0315	0.0331	0.0312	0.0296
Industrial engineering	0.0110	0.0100	0.0097	0.0092	0.0085
Industrial transportation	0.0234	0.0220	0.0203	0.0234	0.0224
Leisure goods	0.0664	0.0607	0.0611	0.0618	0.0605
Life insurance	0.0362	0.0326	0.0370	0.0377	0.0422
Media	0.0189	0.0165	0.0163	0.0145	0.0137
Oil & gas Producers	0.0189	0.0165	0.0163	0.0145	0.0137
Personal goods	0.0124	0.0115	0.0116	0.0114	0.0114
Pharmaceuticals & biotechnology	0.0294	0.0290	0.0281	0.0268	0.0251
Software & computer services	0.0581	0.0554	0.0479	0.0386	0.0335
Support services	0.0506	0.0408	0.0128	0.0128	0.0133
Technology, hardware & equipment	0.0154	0.0142	0.0141	0.0161	0.0132
Travel & leisure	0.0092	0.0094	0.0086	0.0080	0.0077

Table A2. Overview of considered firms.

Companies/Brands			
Accenture	ebay	Kellogg's	PepsiCo
Adidas	Ford	Kraft	Pfizer
American Express	GAP	LG Electronics	Phillips
Audi	General Electric	McDonalds	SAP
BMW	Harley-Davidson	Merill Lynch	Shell
Boeing	Heinz	Microsoft	Siemens
BP	Hermes	Morgan Stanley	Starbucks
Canon	Hewlett-Packard	Motorola	Thomson Reuters
Cisco	Honda	Nestle	UBS
Citi Group	HSBC	Nike	UPS
Coca Cola	Hyundai	Nintendo	VW
Colgate	ING	Novartis	Walt Disney
Mercedes	IBM	Oracle	Xerox
Danone	Johnson & Johnson	Panasonic	Yahoo!

THE STATE OF METHODOLOGICAL PRACTICE IN INTERNATIONAL MARKETING RESEARCH

Charles R. Taylor, C. Luke Bowen and Hae-Kyong Bang

ABSTRACT

Purpose – *A considerable body of literature has evolved on the topic of appropriate research methodology for cross-national data collection. Additionally, prior commentaries on cross-national research in the marketing have cited significant deficiencies in this body of research in terms of the theoretical foundations, methods, and analytical techniques used. The purpose of this chapter is to summarize guidelines for conducting cross-national research in marketing and assess the degree to which these rules are being followed.*

Design/methodology/approach – *The literature on cross-national research methods in marketing studies is first reviewed to identify key issues and methodological guidelines. A content analysis of cross-national studies appearing in 10 major journals in the marketing and advertising field for the period from 2005 to 2010 is conducted to assess whether the guidelines for researchers are being followed. The chapter also explores whether recent research is addressing key deficiencies identified by prior commentaries on this body of research.*

Measurement and Research Methods in International Marketing
Advances in International Marketing, Volume 22, 143–167

ISSN: 1474-7979/doi:10.1108/S1474-7979(2011)0000022010

Findings –*Results are indicative of some promising trends. A wider range of theory bases, methodological techniques, and analytical techniques are being used in cross-national marketing studies. Additionally, methodological guidelines for conceptualizing studies, including following appropriate procedures to ensure equivalence and verifying the existence of cultural differences, are being followed at a higher rate than in the past. Still, some studies do not follow accepted guidelines, and there is a need for a wider range of theory bases and methods to be used.*

Research limitations/implications – *The study examines only cross-national studies published in 10 journals over a recent six years (2005–2010). As a result, no direct comparison to earlier periods is made.*

Originality/value of paper – *This chapter outlines key guidelines for conducting cross-national studies in marketing. It also calls attention to the need to follow these guidelines based on the trend toward a majority of studies complying with them. Finally, the chapter calls attention to the need for certain theory bases and methods to be used more frequently.*

Keywords: Methodology; culture; equivalence; translation; theory development

INTRODUCTION

It is widely acknowledged that conducting cross-cultural research is more complex than collecting data in a single country. As noted by Cateora, Gilly, and Graham (2009), communication issues and environmental differences create additional problems when obtaining data cross-nationally. It is inevitable that these predicaments introduce more noise into data sets than would be present when information is collected in a single market. However, in terms of conducting academic research, recent years have seen advances in available procedural guidelines and methodological tools for effectively conducting marketing research.

While there has been an increase in the number of cross-cultural marketing studies published in major academic journals, it has been widely observed for some time that issues can arise with the research methodologies used in these studies (e.g., Craig & Douglas, 2005; Taylor, 2002; Zinkhan, 1994). The body of research as a whole has been criticized on the grounds of unsophisticated theoretical bases, overabundance of content analysis studies (and an associated lack of surveys and experiments), insufficient assurance of data equivalence, and a lack of application of advanced analytical techniques (e.g., Moriarty & Duncan, 1991; Taylor, 2002).

At the same time, it is apparent that more studies published in the past decade are following long recommended guidelines for conducting research. Thus, the focus of this study is on assessing the degree of compliance with appropriate procedures and accepted principles for conducting cross-cultural research in studies published in major marketing, advertising, and international business journals. The purpose of this chapter is to examine the degree to which published cross-cultural marketing studies are adhering to guidelines and/or recommendations that have been established in the academic literature. The issues examined include:

- Geographic scope of studies and whether a rationale for the choice of countries is provided
- Whether cross-functional and cross-national research teams are used
- The degree to which various data collection techniques are used
- The types of theoretical bases used for the research
- Whether cultural dimensions are measured when used as a theory base
- The primary analytical technique used
- Whether a translation/backtranslation technique is used when needed
- Whether post-hoc equivalence tests are conducted for configural, scale, and metric equivalence
- The types of subjects used in the research

In addition to assessing whether recommended procedures are followed, the study will also examine the degree to which these techniques are used in three types of journals: general marketing journals (e.g., *Journal of Marketing*), advertising journals (e.g., *Journal of Advertising*), and internationally oriented marketing or business journals (e.g., *Journal of International Marketing*).

The chapter's scope is limited to cross-cultural marketing studies, appearing in major journals during the past six years. To be included in the analysis, data from at least two countries must be collected. Single-country studies therefore were not included. It should be noted that the goal of the study is to assess the degree to which guidelines are being followed as opposed to analyzing and critiquing individual studies. We also do not mean to suggest that all guidelines need to be followed in every case in order for a paper to make a contribution to the literature. However, assessing the extent to which researchers are following state-of-the-art practices is useful in evaluating the degree to which the international marketing literature is progressing from a methodological standpoint.

The remainder of the chapter begins by outlining the major recommendations for researchers based on prior literature. Research questions are posed, followed by a description of the study's methodology. Results are then discussed and conclusions are drawn.

GUIDELINES FOR RESEARCHERS

Geographic Scope of Studies

In the past, it has been observed that a large proportion of studies conducted on cross-cultural marketing practices have focused on the triad countries, most often comparing either the United States vs. Japan, or the United States vs. major EU markets (e.g., Taylor, 2005; Zinkhan, 1994). Fam and Grohs (2007) further noted that studies of values expressed in advertising normally took the form of comparing the U.S. and a foreign country, or else using a cluster of countries. As has been observed by Sawyer and Howard (1991), it is important to study marketing practices in diverse cultural contexts.

While it is not surprising that much of the research literature would focus on the largest and most economically developed economies of the world, the growth of the BRIC (Brazil, Russia, India, and China) and other nations suggests that the past six years would have been particularly ripe for additional countries to be included in cross-cultural marketing studies. Zinkhan (1994) and Taylor (2002 and 2005) have argued that this would be a worthwhile development, as would more studies investigating clusters of countries. Thus, it is worthwhile to assess the geographic scope of the studies that are being conducted.

> RQ1: What countries/regions have been most frequently compared in cross-national advertising studies?

Compelling Rationale for Selection of Countries

When a cross-cultural marketing study is designed, ideally, researchers would choose countries in a way that is consistent with the conceptual or theoretical perspective they are trying to test. For example, if the researcher is interested in whether the level of collectivism in a society is associated with how often groups are depicted in the advertising; it makes sense to choose countries that are good exemplars of individualistic and collectivistic cultures. However, it has often been supposed that convenience has driven the choice of countries, rather than the existence of a compelling theoretical rationale.

Researchers have noted that there are several appropriate bases for developing hypotheses in cross-national studies, including cultural differences (Milner, 2005), level of economic development (e.g., Cayla & Eckhardt, 2007; Pollay & Mittal, 1993; Sheth & Parvatiyar, 2001), and social factors (Davila & Rojas-Mendez, 2001; Yang, 2000). These and other factors can be used as a basis for choosing countries to include in a study.

It is important that researchers justify the reasons why specific counties are included in a study. Unless data are collected from a very large number of countries, researchers should be concerned about the degree to which their results apply (or don't apply) to additional countries. Thus, the countries chosen should match the theory base provided in the study. In some cases this may be as simple as testing a general consumer behavior theory, such as the Theory of Planned Action, to see if it applies across countries with different characteristics. In others, it may involve examining the impact of culture on an advertising phenomenon. Thus:

> RQ2: To what extent do cross-national marketing studies provide a compelling rationale for selecting the countries in which data was collected?

Use of Cross-National and Cross-Functional Research Teams

Writing in 1984, Gordon Miracle observed that too few of the published studies on cross-cultural marketing were conducted by cross-national teams. Cross-national teams can play the important role of combining emic and etic perspectives (Malhotra, Agarwal, & Peterson, 2006). Emic perspectives are more practically focused and action oriented (and hence, culture specific), while etic perspectives are value laden, yet theoretically strong. Cross-national research teams can allow for both of these important perspectives to be present. At an intuitive level, it also makes sense that more robust theories with practical implications can be developed if both emic and etic views are available.

The use of cross-functional research teams can also lead to important contributions based on combining perspectives. As observed by Rotfeld and Taylor (2009), the use of cross-functional teams can lend deeper insight into specific issues, especially issues related to managerial relevance and regulation of marketing practices. It simply makes sense that researchers from different disciplines can sometimes provide broader perspectives and that integration of theory can lead to new insights.

RQ3: How commonly are cross-national research teams used in cross-cultural marketing research?

RQ4: How commonly are cross-functional research teams in cross-cultural marketing research?

Variety of Data Collection Techniques Used

Previous analyses of the cross-cultural advertising literature have pointedly suggested that the field would be well served by more survey and experimental research (Taylor, 2002). Both Zinkhan (1994) and Taylor (2005) conducted analyses that indicated that a high proportion of cross-cultural studies were either conceptual in nature or were content analyses. A significant problem cited by these authors is that much of the literature is descriptive in nature. Okazaki and Mueller (2007) agreed and specifically highlighted the shortage of experimental studies, calling for more such research. Thus:

RQ5: Have a wide variety of data collection techniques been used in cross-cultural advertising research?

RQ6: Are surveys commonly used in cross-cultural advertising research?

RQ7: Are experiments commonly used in cross-cultural advertising research?

Theoretical Bases

It has been argued by some (e.g., de Mooij & Hofstede, 2010; Hall, 1976; House, Quigley, & Sully de Luque, 2010; Schudson, 1984) that culture and marketing practices are inextricably linked. However, others view the relationship between culture and marketing to be more variable. For example, Han and Shavitt (1994) found cultural differences to have a strong impact only when products are used and purchased collectively as opposed to individually.

There is little doubt that culture can still play a significant role in understanding differences in marketing practices in many instances. As a result, it is not surprising that several cultural frameworks, including, but not limited to, those proposed by Hofstede (1980), House, Hanges, Mansour, Dorfman, and Gupta (2004), House et al. (2010), Triandis (1994), Trompenaars and Hampden-Turner (1997), and S. H. Schwartz (1992) and S. Schwartz and Sagiv (1995) are still worth considering in conducting cross-national marketing research. However, in an environment where more global strategies are being used than ever before, it is worthwhile to test other theoretical perspectives that have been evolving (Ford, Mueller, & Taylor, 2011).

Indeed, literature on cross-cultural marketing before 2005 was largely dominated by the use of cultural perspectives as a theory base (Taylor, 2005). As a result, there was only a limited application of other perspectives from the consumer behavior, psychology, and sociology fields. However, since 2005, the time would seem particularly ripe for the application of theories from these fields, as well as relatively new perspectives drawn from the management and international business literatures. Recently, perspectives such as resource advantage theory, global consumer culture theory (and the global/local/foreign consumer culture positioning framework, see Alden, Steenkamp, & Batra, 1999), and global marketing strategy (GMS) theory (Zou & Cavusgil, 2002) have been applied to marketing issues in a cross-cultural context. These theories will be briefly described before stating the associated research question.

Resource advantage theory is derived from the resource based view of the firm from management. This perspective asserts that firms build competitive advantages based on the bundle of resources (e.g., core competencies or capabilities) they possess. In the context of marketing, the firm's brands, marketing employees, degree of marketing skill, and how these resources are deployed, affects success (Griffith & Yalcinkaya, 2010).

Zou and Cavusgil's (2002) global marketing strategy theory is also a promising framework that can be applied to cross-cultural studies. GMS outlines eight dimensions of a global marketing strategy; product standardization, promotion standardization, distribution standardization, and pricing standardization, along with four factors related to coordinating value-adding activities; integration of competitive moves, global market participation, coordination of marketing activities, and concentration of marketing activities. These dimensions can be applied to assess how the use of certain marketing strategies performance of firms across markets at both a financial and strategic level (e.g., Okazaki, Taylor, & Zou, 2006).

Building on global consumer culture theory (Arnould & Thompson, 2005), Alden et al.(1999) proposed the construct of global consumer culture positioning. This has become influential in international marketing studies (see Taylor, 2010). The global consumer culture positioning (GCCP) construct suggests a strategy via which brands are associated, widely recognized, and commonly interpreted symbols that are consistent with global consumer culture. Local consumer culture positioning (LCCP), marketers link the brand with local consumer culture and foreign consumer culture positioning, in which the brand is intentionally associated with a foreign culture are two alternative options. Studies can examine the degree to which employment of strategies and/or tactics consistent with these types of positioning are influential across markets.

Another recent advance is the development of the concept of perceived brand globalness, which refers to the degree to which consumer view the brand as being global (Steenkamp, Batra, & Alden, 2003). Evidence has been found in support of perceived brand globalness (PBG) having a positive impact on consumer perceptions of brand equity. This is an important finding that has led to PBG being employed in the international marketing literature (e.g., Becker-Olsen, Taylor, Hill, & Yalcinkaya, 2011).

With the recent advancement and application of international business frameworks, it is particularly useful to look at the extent to which various theoretical bases are being applied. Thus:

> RQ8: Which theoretical perspectives have been applied most often to cross-national advertising research?

Analytical Technique Used

Both Zinkhan (1994) and Taylor (2005) reported that many of the cross-cultural studies that had been conducted in the advertising field were either conceptual in nature or content analyses. While content analyses can provide descriptive data that is sometimes appropriate for a research purpose, it is limited in its ability to provide information on cause and effect or the underlying explanations behind what makes leads to a marketing strategy or tactic effective. Basic analyses such as χ^2 tests that are often used in conjunction with content analyses also have confines.

As observed by Joreskog and Sorbom, 1993, structural equations models (SEM) have proven to be useful to researchers when the aim is to make causal attributions about the impact of one variable on another in a non-experimental situation (e.g., as with data collected in a survey). Gerbing and Anderson (1993) and Jaccard and Wan (1996) make similar observations about χ^2 tests, noting that they are highly sensitive to multivariate normality and can produce unstable results when assumptions are violated (Gerbing & Anderson 1993; Jaccard & Wan, 1996).

While it is always important to choose the analytical technique most closely matched to the research purpose, advanced analytical techniques, such as SEM, and multivariate techniques, such as regression, factor analysis, and cluster analysis, have the potential to provide a level of insight on certain types of research questions that simple tests cannot. It also would make sense that as research has evolved over time, we would see more application of such advanced techniques. Some researchers have also called for the application of more qualitative techniques as well (e.g., Thompson, 1997) have called for more cross-cultural studies using qualitative techniques. Thus, it is worthwhile to examine whether various analytical techniques are being used in cross-national marketing studies:

> RQ9: Are a wide variety of different analytical techniques being applied in cross-national marketing studies?

Backtranslations and Equivalence Tests

Brislin (1970) commented on the need to translate and backtranslate survey instruments to ensure equivalence in cross-cultural studies. Subsequently, it has become a widely accepted principle that a procedure involving translation and backtranslation is needed when collecting cross-cultural data (e.g., Brislin, 1986; Craig et al., 2005; Kumar, 2000; Miracle, 1988; Van de Vijver & Leung, 1997). Over the years, additional literature on how to effectively conduct translations has evolved (e.g., Harkness, Fons, Van de Vijver, & Mohler, 2003) and the use of cross-national teams is sometime recommended depending on the researcher's budget and the specific task at hand. However, the need for backtranslation is now so well established that it can now be viewed as an error when a study does not follow this procedure.

In addition to taking measures to ensure equivalence before data is collected, Taylor (2005) observed that there is a need for advertising studies to use available measures for assessing equivalence after data has been collected. There are multiple possibilities for measuring various types of equivalence including Ewing, Salzberger, and Sinkovics' (2005) proposed Raasch-based technique. The technique that has become most standard, however, is Steenkamp and Baumgartner's (1998) approach, which uses confirmatory factor analysis, appears to have become the standard for cross-cultural research in marketing.

The focus of the Steenkamp and Baumgartner approach is on establishing three primary types of equivalence. They are configural (scalar) equivalence, metric equivalence, and functional equivalence. Configural invariance refers to the scalar item showing the same configuration of salient as opposed to nonsalient factor loadings across the different country groupings. Metric invariance refers to ensuring that respondents in different cultural settings reply to the various items in the same way. For functional equivalence, the researcher needs assurance that the underlying constructs are the same across the different cultural settings. Steenkamp and Baumgartner recommend measuring factor covariance invariance to establish functional equivalence.

RQ10: Are appropriate translation/backtranslation procedures being followed in cross-national studies?

RQ11: In what proportion of cross-national studies are appropriate post-hoc equivalence tests to measure configural, metric, and functional equivalence conducted?

Verifying the Existence of Cultural Differences

As discussed earlier, it is not uncommon for cross-national studies to use cross-cultural differences as the basis for hypotheses. Too often, however, when cultural variables are used to form hypotheses, the cultural variable in question is not measured; rather it is assumed that groups in two or more countries differ based on secondary data. As observed by Taylor (2005), this can be problematic, especially if a sub-group, such as students, is used as the sample. Given the availability of scales used by Hofstede (1980), House et al. (2004), and others, these dimensions can and should be measured. If this is done, the hypothesized difference can be verified (Taylor, 2005). Moreover, measuring such cultural variables can permit individual difference tests on cultural variables.

RQ12: What proportion of cross-national marketing studies that use cultural variables as a basis for hypotheses collect data from the sample pertaining to the cultural dimensions being studied.

Type of Subjects

It has been observed that a disproportionate number of the cross-national studies have used student subjects as opposed to a sample from the broader population. Use of student subjects has sometimes been justified based on ease of data collection and/or greater within group homogeneity (e.g., Calder, Phillips, & Tybout, 1981; Chan, Lyann, Diehl, & Terlutter, 2007; Soley & Reid, 1983). However, samples of student subjects are often criticized as having less external validity.

The degree to which a student sample is appropriate is clearly related to research goals. However, in an environment in which it is less difficult to collect data internationally in at least some settings, one might expect to see an increase in the proportion of studies that use non-student samples. This would likely be viewed as a positive development as it has been observed (e.g., Moriarty & Duncan, 1991; Zinkhan, 1994) that the general research stream would benefit from more studies using non-student samples.

RQ13: To what extent are non-student subjects vs. student subject used in cross-national advertising research?

METHODOLOGY

Articles in 10 SSCI listed journals from the marketing, advertising, and international business fields during the six year period from January 2005 through December 2010 were content analyzed. In addition to being listed in SSCI, a selection of the top general marketing, general advertising, and international marketing journals were intentionally chosen. The resulting list of journals included four general marketing journals: *Journal of Marketing*, *Journal of Marketing Research*, *Journal of Consumer Research*, *Marketing Science*; three advertising journals: *Journal of Advertising*, *International Journal of Advertising*, and *Journal of Advertising Research*; and three journals with an international marketing or business focus: *Journal of International Business Studies*, Journal *of International Marketing*, and *International Marketing Review*. In addition to addressing the listed research questions, the degree to which there were differences by journal category was also examined.

A coder with strong background in research methodology was trained extensively by the lead author. A data coding instrument was developed based on prior literature on international research methods. In addition to publication, the primary variables coded were: the number of countries studies; whether there was a stated rationale for the countries selected; primary rationale for selection of countries; use of cross-national and cross-disciplinary research teams; primary data collection technique; theory base; analytical technique; use of backtranslation; use of post hoc equivalence test; measurement of cultural factors; application of individual difference measures; and type of subject.

In the following results section, in addition to reporting frequencies for the variables of interest cross-tabulations for some of the variables by journal type will be reported. In instances where there is an interesting finding pertaining to a difference by journal type a separate table will be presented and discussed.

RESULTS AND DISCUSSION

While reliability was not measured for the entire sample as it primarily involves simple classification, the lead author coded 20 studies independently to assure that there were no systematic differences in coding of some of the items and found no significant discrepancies. The literature search found 80 studies that collected cross-national data on an advertising topic.

Journal Coverage

Table 1 shows the breakdown of articles by journal. As can be seen, five journals, the *International Marketing Review* (31%), *International Journal of*

Table 1. Journal.

	No.	%
Advertising journals		
Journal of Advertising	11	14
Journal of Advertising Research	10	13
International Journal of Advertising	12	15
International marketing/business		
International Marketing Review	25	31
Journal of International Marketing	11	14
Journal of International Business Studies	2	3
General marketing		
Journal of Marketing	3	4
Journal of Marketing Research	3	4
Journal of Consumer Research	1	1
Marketing Science	2	3

Advertising (15%), *Journal of International Marketing* (14%), *Journal of Advertising* (14%), and *Journal of Advertising Research* (13%), accounted for 87% of the papers published on cross-national advertising in this sample. It is notable that the general marketing category, consisting of four journals, accounted for just eight, or 14% of the cross-cultural articles. Additionally, *Journal of International Business Studies* published only two cross-national marketing studies over the six-year period. While this result may not be surprising given that the advertising and international marketing journals are more specialized, it nonetheless appears to be the case that relatively few cross-national marketing studies get published in the top four general marketing journals, as less than one article per year fell into this category. Meanwhile, it is clear that advertising journals and international marketing journals publish a substantial number of cross-cultural studies.

RQ1: Countries Studied

As Table 2 shows, the single most common geographic scope in the 80 studies reviewed was ''three or more continents,'' with 22 studies (28%) falling in this category. This was followed by North America vs. Asia with 19 (24%) and North America vs. Europe comparisons at 14 studies (18%). It is notable that only two of the two country studies analyzed here focused on Latin America, which appears to be an understudied region. On one hand, the results shown in Table 2 suggest a broader geographic scope of studies having been conducted than was reported by Taylor (2005). On the other hand, it is clear that a very substantial proportion of studies, well over

Table 2. Countries Studied.

	No.	%
North America vs. Asia	19	24
Europe vs. Asia	5	6
North America vs. Europe	14	18
Latin America vs. Asia	1	1
Latin America vs. Europe	1	1
Latin America vs. North America	0	0
Three or more continents	22	28
Same continent	15	19
Other	3	4

half, (when three country studies and same continent studies are included), still include a North American country, which is usually the United States. There were not notable differences across journal type in terms of geographic scope of the studies.

RQ2: Compelling Rationale for Countries

Table 3 shows that the overwhelming majority (85%) of the studies provided a compelling rationale for the selection of countries in the study. This finding suggests that this rule is being largely followed by academic researchers conducting international marketing research. While on the surface, this may seem to be an obvious point, it is nonetheless encouraging given that as recently as the 1980s and 1990s it was being observed that samples were being chosen based on convenience and without sufficient justification. Largely because only a small number of studies did not provide a compelling rationale, there were not significant differences by journal type.

RQs 3–4: Use of Cross-National and Cross-Functional Research Teams

As shown in Table 4, exactly half of the studies were conducted by cross-national research teams. Across journal type, the differences in relative frequency by magazine type were not statistically significant for this

Table 3. Compelling Rationale for Country Selection.

	No.	%
Yes	68	85
No	12	15

Table 4. Authors: Cross-National Teams.

	No.	%
Yes	40	50
No	40	50

variable. The use of cross-functional research teams was a bit less common (Table 5), with 31% of the studies including researchers from different disciplines. There were not significant differences based on journal type.

The figures for both cross-national and cross-functional teams are encouraging in that the proportion of studies in which a range of backgrounds perspectives is brought by teams of co-authors appears to be increasing. Indeed, it is shown to be a common practice. Teams of researchers from different countries and different disciplines working together on cross-national marketing research can likely provide multiple perspectives and perhaps deeper insights into research problems.

RQs 5–7: Data Collection Techniques

The next set of research questions examined whether a wide variety of data collection techniques are being used, and specifically, whether surveys and experiments are being used. As shown in Table 6(a), it does appear that a range of techniques are being used, thereby answering RQ5 affirmatively. As surveys are the single most used technique (44% of the studies), the answer to RQ6 is also, yes. Experiments were used in 14, or 18% of the studies, which is considerably less and more open to interpretation. Thus, it appears

Table 5. Authors: Cross-Functional Teams.

	No.	%
Yes	25	31
No	55	69

Table 6a. Data Collection Technique.

	No.	%
Content analysis	19	24
Experiment	14	18
Survey	35	44
Conceptual	0	0
Qualitative	3	4
Secondary data analysis	5	6
Other	4	5

Table 6b. Primary Data Collection Technique.

	Advertising		International Marketing/ Business		General Marketing	
	No.	%	No.	%	No.	%
Content analysis	15	45	3	8	1	11
Experiment	5	15	7	18	2	22
Survey	10	30	23	61	2	22
Conceptual	0	0	0	0	0	0
Qualitative	0	0	1	3	3	33
Secondary data analysis	3	9	1	3	1	11
Other	0	0	3	8	0	0

Notes: $\chi^2 = 57.4$; $p < 0.001$.

that the answer to RQ7 that asked whether experiments are commonly used is a qualified yes. Overall, surveys were followed by content analysis and experiments, as the three most commonly used techniques, accounting for just over three-quarters of all studies. There were no conceptual studies in this study's sample and just three qualitative (4%) and four studies (5%) based on secondary data analysis.

Table 6(b) shows additional insight into the results as there is a significant difference in the relative frequency of these techniques when analyzed by journal type ($\chi^2 = 57.4$; $p < 0.001$). The most notable issue here is clearly the higher frequency of content analyses in advertising journals (45%) compared to the international marketing/business journals (8%) and general marketing journals (11%). It is also notable that fully 61% of the articles in international marketing journals were surveys and that one-third of the cross-cultural marketing articles in the elite general marketing journals were qualitative. On the whole, however, it does appear that a range of techniques in cross-national marketing are being used across all the journal categories, with the exception of qualitative studies, none of which appeared in advertising or international business journals.

RQ8: Theory Bases

Research question 8 examined the extent to which different types of theory bases are used in cross-national marketing studies. As shown in Table 7(a), cultural dimensions still place first, with 26 (33%) of studies falling in this

Table 7a. Theoretical Basis.

	No.	%
Cultural	26	33
Social science	24	30
Grounded theory	7	9
No theory base	9	11
International business/management	14	18

Table 7b. Theory Base.

	Advertising		International Marketing/ Business		General Marketing	
	No.	%	No.	%	No.	%
Cultural	11	33	12	32	3	33
Social science	12	36	9	24	3	33
Grounded theory	2	6	2	5	3	33
No theory base	5	15	3	8	0	0
International business/management	2	6	12	32	0	0

Notes: $\chi^2 = 18.6$ $p < 0.001$.

category. However, 24 (30%) of the studies tested social science-based theories from consumer behavior or related fields, and 14 (18%) used relatively new perspectives from management and international business. These results do show various theory bases being used and are encouraging when the overall literature is assessed.

However, one major flaw that was found was that 9% or 11% of the studies did not offer any significant theory base. As shown in Table 7(b), the only interesting difference is that the international marketing/international business journals at 32% have been more prone to use the newer international business/management theories than have the advertising (6%) or general marketing journals (0%). This is perhaps not surprising given that one would expect the internationally oriented journals to take the lead in this regard. As shown in Table 7(b), the overall difference in relative frequency of journal type by theory base was statistically significant ($\chi^2 = 18.6$; $p < 0.001$).

RQ9: Primary Analytical Techniques

RQ9 examines the degree to which advanced analytical techniques are used in cross-national marketing studies. It should be noted that the coding was focused on the primary analytical technique used in the study. As a result, some techniques, such as factor analysis, that are frequently used secondarily, may be under-represented relative to their actual overall use. Table 8(a) shows that regression was the most commonly employed technique (24%), followed by *t*-tests (21%), analysis of variance (15%), structural equations modeling (14%), and cluster analysis (10%). Qualitative techniques and factor analysis were also used as the primary technique in some studies. As a result, it is clear that quite a wide variety of analytical techniques are being used.

Table 8(b) shows that there are some significant differences in the primary analytical technique used ($\chi^2 = 13.1$; $p = 0.041$). Most notably, the advertising journals are more prone to use regression, while general marketing journals

Table 8a. Analytical Technique.

	No.	%
t-test	17	21
ANOVA/MANOVA	12	15
Regression	19	24
Structural equations modeling	11	14
Factor analysis	7	9
Cluster analysis	8	10
Qualitative analysis	6	8

Table 8b. Primary Analytical Technique.

	Advertising		International Marketing/ Business		General Marketing	
	No.	%	No.	%	No.	%
t-test	7	21%	9	24%	1	11%
ANOVA/MANOVA	5	15%	3	8%	4	44%
Regression	10	30%	8	21%	1	11%
Structural equations modeling	3	9%	6	16%	2	22%
Factor analysis	1	3%	6	16%	0	0%
Cluster analysis	5	15%	2	5%	0	0%
Qualitative analysis	2	6%	4	11%	1	11%

Notes: $\chi^2 = 13.1$ $p = 0.041$.

used ANOVA/MANOVA disproportionately. However, it is apparent that a range of techniques are used for all three journal types.

RQ 10 and RQ 11: Backtranslation and Equivalence Tests

As Table 9(a) shows, of the 52 cross-national studies administered in countries where the native language is different, 34 employed back-translation, with 10 collecting data in another language without a backtranslation, and another 8 collecting data in English from non-native speakers. Thus, slightly under two-thirds (65%) of studies where data were collected in different languages employed a back-translation, indicative of its status as accepted and normally being necessary practice. However, it is somewhat concerning that the remaining 35% of studies did not use a back-translation. Moreover, the fact that 8 (10% in the overall sample) studies collected data in a non-native language raises is indicative of some studies not following ideal practices.

Ironically, Table 9(b) reveals that eight of the studies in international marketing/business did not use a backtranslation and another four

Table 9a. Backtranslation Used.

	No.	%
Yes	34	43
No	10	13
Administered in non-native language	8	10
Same native language/no need	28	35

Table 9b. Translation/Backtranslation.

	Advertising		International Marketing/ Business		General Marketing	
	No.	%	No.	%	No.	%
Yes	13	39	19	50	2	22
No	3	9	8	21	0	0
Same native language/no need	16	48	7	18	5	56
Administered in non-native language	1	3	4	11	2	22

Notes: $\chi^2 = 4.9$ $p = 0.027$.

administered a survey in a non-native language, meaning 32% did not choose this technique (overall χ^2 by journal type = 4.9; $p = 0.27$). While this figure is not that different from the overall sample, it is somewhat surprising given that on the whole this study has found that the international marketing/business journals appear to follow the guidelines for cross-national studies more closely than the other two categories.

In terms of post-hoc equivalence test, Table 10(a) shows that most studies that can conduct post-hoc equivalence tests are doing so. Of the 26 studies where such post-hoc tests could be conducted, 18 measured configural equivalence (69%), 18 measured scale equivalence (24%), and 23 measured metric equivalence (31%). Overall it is encouraging that such a substantial proportion of studies that have data that lends itself to post-hoc equivalence tests run such tests as this has not always been the case in the past. It is likely the case that in addition to heightened author sensitivity toward this issue reviewers and journal editors are mandating that these post-hoc tests be run.

Table 10(b) shows that there are not large differences by journal type in terms of the number of studies employing these tests with the exception of only 50% of the advertising studies that could have employed a test for scalar equivalence doing so. Chi-square tests were not run for these variables due to the small cell size (3) for general marketing journals.

Table 10a. Post-Hoc* Equivalence Tests.

	No.	%
Configural	18	69
Scale	18	69
Metric	23	86

*Numbers are out of 26 studies for which post-hoc tests would have been appropriate.

Table 10b. Post-Hoc Equivalence Tests by Journal Type.

	Advertising		International Marketing/Business		General Marketing	
	No.	%	No.	%	No.	%
Configural	5	63	11	73	2	67
Scale	4	50	12	80	2	67
Metric	7	88	13	87	3	100

RQ12: Verifying the Existence of Cultural Differences

RQ 12 examines the question of whether those studies that use a cultural dimension(s) collect data that verifies that the individuals in the countries sampled actually vary on the cultural dimension of interest. As indicated in Tables 11(a) and (b), 22 of the 26 studies that used cultural dimensions measured whether the subjects differed by country on this dimension. Because of the small cell size for general marketing journals (3), a χ^2 test on this variable was not conducted. However, it is clear that a strong majority of studies across journal types followed these guidelines.

RQ13: Type of Subjects

As shown in Table 12, of those studies in which subjects were used, 35 used a non-student sample, while two used a combination of students and non-students, and 21 used a student sample. Thus, 64% of the studies use non-student subjects in some capacity. These results suggest a large increase in the use of non-student subjects in cross-national marketing studies. This may be a result of editor and reviewer preferences for more generalizable samples. This was true across journal types as there were no statistically significant difference.

Table 11a. Measurement of Cultural Dimensions.

	No.	%
Yes	22	85
No	4	15

Table 11b. Measurement of Cultural Factors Test.

	Advertising		International Marketing/Business		General Marketing	
	No.	%	No.	%	No.	%
Yes	8	73	11	92	3	100
No	3	27	1	8	0	0

Table 12. Type of Subjects.

	No.	%
Students	21	26
Non-students	35	44
Students and non-students	2	3
No subjects	22	28

CONCLUSION

The results of this study are encouraging in that they suggest that substantial strides have been made in cross-national marketing research. The content analysis suggests that many guidelines or suggestions for collectively improving this general area of research are being followed. While it is true that few cross-national studies are being published in elite general marketing journals, it is also clear that advertising journals and international marketing/business journals are clearly receptive to this type of research. While there remains some room for improvement, research is being published across a wider geographic scope than has been the case in the past. Moreover, a significant proportion of the research is being conducted by cross-national teams of researchers.

It is also clear that researchers are doing a better job (or are being forced to) of providing a compelling rationale for the choice of countries included in a study. Notably, a broader array of theory bases are being used, and we are seeing a trend toward more of the newly developed management/international business theories being employed. It seems likely that the future will see more of these theories being implemented. Furthermore, there may be a trend toward simultaneously testing general theories and testing whether cultural variables appear to explain results.

The results of this content analysis also indicate that wider ranges of analytical techniques are being applied in cross-national marketing studies. The use of cross-national surveys appears to have risen considerably in recent years. It does appear that there is still additional room for the growth of cross-national studies employing experiments or qualitative research techniques. A wide range of analytical techniques are also being applied and, as a result, it is clear those conceptual papers and content analyses are no longer "king" when it comes to cross-national marketing research.

It is particularly encouraging that most studies conducted in these leading journals employ backtranslations when needed. A majority of studies that use methods that are conducive to post hoc equivalence tests are conducting such tests. While it can be argued that all studies should be employing these measures to ensure equivalence, at least, there is a clear trend toward a larger percentage of the studies doing so. It is also encouraging that a high proportion of those studies that use cultural dimensions to develop hypotheses actually measure whether the subjects in the countries studies adhere to the cultural assumptions made.

It is likely that the future will see a continued acceleration of cross-national research in marketing. Hopefully, the practices outlined here will be followed even more and, with additional refinement and improvement, will allow knowledge to advance even further.

REFERENCES

Alden, D. L., Steenkamp, J. B. E. M., & Batra, R. (1999). Brand positioning through advertising in Asia, North America, and Europe: The role of global consumer culture. *Journal of Marketing*, *63*(1), 75–87.

Arnould, E. J., & Thompson, C. J. (2005). Consumer culture theory (CCT): Twenty years of research. *Journal of Consumer Research*, *31*(4), 868–892.

Becker-Olsen, K. L., Taylor, C. R., Hill, R. P., & Yalcinkaya, G. (2011). A cross cultural look at corporate social responsibility marketing communications in Mexico and the United States: Strategies for global brands. *Journal of International Marketing, 19*(2), 30–44.

Brislin, R. (1970). Back translation for cross-cultural research. *Journal of Applied Psychology*, *1*, 185–216.

Brislin, R. (1986). The wording and translation of research instruments. In: W. Lonner & J. Berry (Eds.), *Field methods in cross-cultural research* (pp. 137–164). Newbery Park, CA: Sage.

Calder, B. J., Phillips, L. W., & Tybout, A. M. (1981). Designing research for applications. *Journal of Consumer Research*, *8*, 197–207.

Cateora, P. R., Gilly, M., & Graham, J. L. (2009). *International marketing* (14th ed.). Homewood, IL: Richard D. Irwin.

Cayla, J., & Eckhardt, G. M. (2007). Asian brands without borders: Regional opportunities and challenges. *International Marketing Review*, *24*(4), 44–56.

Chan, K., Lyann, L., Diehl, S., & Terlutter, R. (2007). Consumers' response to offensive advertising: A cross cultural study. *International Marketing Review*, *24*(5), 606–628.

Craig, C., Douglas, S., & Douglas, S. P. (2005). *International marketing research* (3rd ed). New York, NY: Wiley.

Davila, V., & Rojas-Mendez, J. (2001). Attitude toward advertising: Does the 7-factor model work in Chile? *International Journal of Organizational Theory and Behavior*, *4*(1&2), 3–19.

de Mooij, M., & Hofstede, G. (2010). The Hofstede model: Applications to global branding and advertising strategy and research. *International Journal of Advertising*, *29*(1), 85–110.

Ewing, M. T., Salzberger, T., & Sinkovics, R. R. (2005). An alternative approach to assessing cross-cultural measurement equivalence in advertising research. *Journal of Advertising*, *34*(1), 17–36.

Fam, K. S., & Grohs, R. (2007). Cultural values and effective executional techniques in advertising. *International Marketing Review*, *24*(5), 519–538.

Ford, J. B., Mueller, B., & Taylor, C. R. (2011). The tension between strategy and execution: Challenges for international advertising research. *Journal of Advertising Research*, *51*(1), 27–41.

Gerbing, D. W., & Anderson, J. C. (1993). Monte Carlo evaluations of goodness-of-fit indices for structural equation models. In: K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 40–65). Newbury Park, CA: Sage.

Griffith, D. A., & Yalcinkaya, G. (2010). Resource-advantage theory: A foundation for insights into global advertising research. *International Journal of Advertising*, *29*(1), 15–36.

Hall, E. T. (1976). *Beyond culture*. Garden City, NY: Anchor Press/Doubleday.

Han, S., & Shavitt, S. (1994). Persuasion and culture: Advertising appeals in individualistic and collectivistic societies. *Journal of Experimental Social Psychology*, *30*(4), 326–350.

Harkness, J. A., Fons, J. R., Van de Vijver, J. R., & Mohler, P. H. (2003). *Cross-cultural survey methods*. Hoboken, NJ: Wiley.

Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. Newbury Park, CA: Sage.

House, R. J., Hanges, P. J., Mansour, J., Dorfman, P. W., & Gupta, V. (2004). *Culture, leadership and organizations: The globe study of 62 societies*. Thousand Oaks, CA: Sage Publications.

House, R. J., Quigley, N. R., & Sully de Luque, M. (2010). Insights from project globe: Extending global advertising research through a contemporary framework. *International Journal of Advertising*, *29*(1), 119–139.

Jaccard, J., & Wan, C. K. (1996). *Lisrel approaches to interaction effects in multiple regression*. Thousand Oaks, CA: Sage.

Joreskog, K., & Sorbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Chicago, IL: Scientific Software International.

Kumar, V. (2000). *International marketing research*. Upper Saddle River, NJ: Prentice Hall.

Malhotra, N. K., Agarwal, J., & Peterson, M. (1996). Methodological issues in cross-cultural marketing research: A state-of-the-art review. *International Marketing Review*, *13*(5), 7–43.

Milner, L. (2005). Sex-role portrayal in African television advertising: A preliminary examination with implications for use of Hofstede's research. *Journal of International Consumer Research*, *17*(2/3), 73–91.

Miracle, G. E. (1984). An assessment of progress in research in international advertising. *Current Issues and Research in Advertising, 2*, 135–166.

Miracle, G. E. (1988). An empirical assessment of the usefulness of the back-translation technique for international advertising messages in print media. In: J. D. Leckenby (Ed.), *Proceedings of the 1988 Conference of the American Academy of Advertising*. RC51–RC54. Austin, TX: University of Texas.

Moriarty, S. E., & Duncan, T. R. (1991). Global advertising: Issues and practices. *Current Issues and Research in Advertising*, *13*(1&2), 313–341.

Okazaki, S., & Mueller, B. (2007). Cross-cultural advertising research: Where we have been and where we need to go. *International Marketing Review*, *24*(5), 499–518.

Okazaki, S., Taylor, C. R., & Zou, S. (2006). Advertising standardization's positive impact on the bottom line: A model of when and how standardization improves financial and strategic performance. *Journal of Advertising*, *35*(4), 17–33.

Pollay, R., & Mittal, B. (1993). Here's the beef: Factors, determinants, and segments in consumer criticism of advertising. *Journal of Marketing*, *57*(3), 99–114.

Rotfeld, H. J., & Taylor, C. R. (2009). The advertising regulation and self-regulation issues ripped from the headlines with (sometimes missed) opportunities for disciplined multi-disciplinary research. *Journal of Advertising*, *38*(4), 5–14.

Sawyer, A. G., & Howard, D. J. (1991). Effects of omitting conclusions in advertisements to involved and uninvolved audience. *Journal of Marketing Research*, *28*(11), 467–474.

Schudson, M. (1984). *Advertising, the uneasy persuasion: Its dubious impact on American society*. New York, NY: Basic Books.

Schwartz, S., & Sagiv, L. (1995). Identifying culture-specifics in the content and structure of values. *Journal of Cross-Cultural Psychology*, *26*(1), 92–116.

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. *Advances in Experimental Psychology*, *25*, 1–65.

Sheth, J. N., & Parvatiyar, A. (2001). The antecedents and consequences of integrated global marketing. *International Marketing Review*, *18*(1), 16–29.

Soley, L. C., & Reid, L. N. (1983). On the validity of students as subjects in advertising experiments. *Journal of Advertising Research*, *23*(4), 57–59.

Steenkamp, J. B. E. M., & Baumgartner, H. (1998). Assessing measurement equivalence in cross-national consumer research. *Journal of Consumer Research*, *25*(2), 78–90.

Steenkamp, J., Batra, R., & Alden, D. L. (2003). How perceived brand globalness creates brand value. *Journal of International Business Studies*, *34*(1), 53–65.

Taylor, C. R. (2002). What is wrong with international advertising research? *Journal of Advertising Research*, *42*(6), 48–54.

Taylor, C. R. (2005). Moving international advertising research forward: A new research agenda. *Journal of Advertising*, *34*(1), 7–16.

Triandis, H. C. (1994). *Culture and social behavior*. New York, NY: McGraw-Hill.

Trompenaars, F., & Hampden-Turner, C. (1997). *Riding the waves of culture: Understanding cultural diversity in business* (2nd ed.). London: Nicholas Brealey.

Thompson, C. J. (1997). Interpreting consumers: A hermeneutical framework for deriving marketing insights from the texts of consumers' consumption stories. *Journal of Marketing Research*, *34*(4), 438–455.

Van de Vijver, F., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Sage.

Yang, C. (2000). Taiwanese students' attitudes towards and beliefs about advertising. *Journal of Marketing Communications*, *6*(3), 171–183.

Zinkhan, G. M. (1994). International advertising: A research agenda. *Journal of Advertising*, *23*(1), 11–15.

Zou, S., & Cavusgil, S. T. (2002). The GMS: A broad conceptualization of global marketing strategy and its effect on firm performance. *Journal of Marketing*, *66*(4), 40–56.

ASSESSING HETEROGENEITY IN CUSTOMER SATISFACTION STUDIES: ACROSS INDUSTRY SIMILARITIES AND WITHIN INDUSTRY DIFFERENCES

Edward E. Rigdon, Christian M. Ringle, Marko Sarstedt and Siegfried P. Gudergan

ABSTRACT

Purpose – *Revisiting Fornell et al.'s (1996) seminal study, this chapter looks at the evidence for observed and unobserved heterogeneity within data underlying the American customer satisfaction index (ACSI) model. Examining data for two specific industries (utilities and hotels) reveals only modest differences. However, we suppose that unobserved heterogeneity critically affects the results. These insights provide the basis for shaping further differentiated ACSI model analyses and more precise interpretations.*

Methodology/approach – *This study applies the partial least squares (PLS) path modeling method and uses empirical data to estimate and compare the ACSI model results on the aggregate and industry-specific data levels. In addition, the finite mixture PLS path modeling (FIMIX-PLS)*

Measurement and Research Methods in International Marketing
Advances in International Marketing, Volume 22, 169–194

ISSN: 1474-7979/doi:10.1108/S1474-7979(2011)0000022011

method is employed to further examine across industry similarities and within industry differences.

Findings – *This research uncovers unobserved heterogeneity that guides forming three segments of customers within each industry. The major segment in each industry represents customers that are fairly loyal (i.e., neither disloyal nor extremely loyal) while the other two smaller segments are not as similar across the two industries. Our study identifies substantial differences across these segments within each industry. An importance-performance map analysis illustrates these differences and provides the basis for managerial implications.*

Originality/value of the chapter – *The unobserved heterogeneity revealed within industries in a given country (i.e., the US) underlines the need to be open to differences within populations, beyond the observed heterogeneity across distinct groups or cultures, and the need to reconsider reporting requirements in academic research.*

Keywords: American customer satisfaction index (ACSI) model; partial least squares (PLS) path modeling; unobserved heterogeneity; multigroup analysis; important-performance matrix

INTRODUCTION

Customer satisfaction is a fundamental concept in marketing (Anderson, Fornell, & Lehmann, 1994; Anderson & Sullivan, 1993) and represents a key to business success, evidenced through established effects on customer retention and profitability. With the development of satisfaction indices in Sweden (Fornell, 1992), Norway (Andreassen & Lindestad, 1998), and the US (Fornell, Johnson, Anderson, Cha, & Bryant, 1996), customer satisfaction has gained national and international significance. Today, the American customer satisfaction index (ACSI) ranks among the most salient models in studying customer satisfaction not only in the US but also manifold international contexts. Specifically, the ACSI has provided a basis for similar indices such as the European customer satisfaction index (ECSI), which is, in essence, a variation of the ACSI model (Eklöf & Westlund, 2002).

Since 1994, ACSI data have been used widely and research has been published in various academic journals. Specifically, ACSI data have recently been used to examine the link between marketing activity and shareholder value (Aksoy, Cooil, Groening, Keiningham, & Yalçin, 2008; Anderson, Fornell, & Mazvancheryl, 2004; Anderson & Mansi, 2009; Gruca & Rego, 2005; Lou & Homburg, 2008; Mittal, Anderson, Sayrak, & Tadikamalla, 2005; Raithel, Sarstedt, Scharf, & Schwaiger, 2011) and advertising and promotion efficiency (Lou & Homburg, 2007). Commonly, these studies draw directly – or indirectly as it is in international contexts – on firm-level data that have been tested in the original PLS path modeling analysis by Fornell et al. (1996) for reliability and validity. Although their results provide a basis for comparing the effects of antecedent constructs on overall satisfaction and loyalty, as is demonstrated in the seminal ACSI model, there are variations to this model that have been applied to and that are based on indices used in international contexts such as the slightly different ECSI model. The issue that arises from such model adaptations points toward the universal validity of the underlying model structure within and across contexts.

In their analyses, Fornell et al. (1996) assume that the data stem from a homogenous population – a single model represents all observations. However, this assumption of homogeneity is not necessarily justified as customers are likely to be different in their perceptions and evaluations of as well as familiarity with firms' offerings (Jedidi, Jagpal, & DeSarbo, 1997; Rigdon, Ringle, & Sarstedt, 2010). In this context, one needs to distinguish between observed and unobserved heterogeneity (Lubke & Muthén, 2005). Observed heterogeneity addresses forming groups of data based on theoretical assumptions and prior knowledge. Unobserved heterogeneity on the contrary describes circumstances when groups of data are unknown. This kind of situation may exist when theory is not well developed (i.e., well-established knowledge about the existence of groups and their specific characteristics and differences does not exist).

The question that arises for those interested in understanding and measuring customer satisfaction is how to identify a customer satisfaction model structure that may help illuminate the reasons behind adaptations across industries and international contexts (i.e., across countries). As a first step, we argue, it is important to examine heterogeneity within national contexts, which is, in particular, the case for the ACSI as it has provided a basis for adaptations in international contexts. Hence, in this study on the ACSI, we aim at examining the effect of heterogeneity across industries with a focus on a specific country.

However, even if a priori theory is capable of accounting to some extent for observed heterogeneity (e.g., analyses for industries and countries;

Johnson, Herrmann, & Gustafsson, 2002), a large amount of heterogeneity is frequently unobservable and its true sources are unknown. Studies report the existence of substantial consumer heterogeneity within a given product, service, or industry class (Wu & Desarbo, 2005). Hence, observable characteristics are often inadequate in capturing the apparent heterogeneity in the data (e.g., Wedel & Kamakura, 2000). Ignoring such heterogeneity can easily lead to biased parameter estimates and, consequently, potentially flawed conclusions as illustrated, for instance, by Sarstedt, Schwaiger, and Ringle (2009). Recent advances in PLS path modeling have incorporated several approaches – such as the finite mixture PLS method (FIMIX-PLS; Hahn, Johnson, Herrmann, & Huber, 2002; Sarstedt & Ringle, 2010) – that permit uncovering unobserved heterogeneity. The availability of the FIMIX-PLS method (e.g., in the SmartPLS software application; Ringle, Wende, & Will, 2005) allows us to revisit Fornell et al.'s (1996) seminal ACSI study toward the goal of clarifying the role of heterogeneity.

In this chapter, we extend prior research by examining observed and unobserved moderating factors in the ACSI model. By contrasting results from a priori partitioning of the observations into industries with the results of a latent class analysis per industry, this study considers different types of heterogeneity at various levels of analysis. Moreover, by applying importance-performance map analyses, this research provides further differentiated results and, thereby, establishes the necessary grounds for deriving segment-related managerial implications (Höck, Ringle, & Sarstedt, 2010; Völckner, Sattler, Hennig-Thurau, & Ringle, 2010). The analyses suggest similarities across industry lines and reveal unobserved heterogeneity in within-industry data. This points to a much larger implication for international marketing researchers whose datasets feature obvious bases for observed heterogeneity, but may well conceal substantial unobserved heterogeneity.

THE AMERICAN CUSTOMER SATISFACTION INDEX MODEL AND DATA

The ACSI conducted by the University of Michigan's National Quality Research Center is a national system of customer satisfaction measurement that was established in 1994. It measures customer satisfaction with respect to more than 200 corporate and government organizations from a broad cross-section of industries representing 43% of the US GDP. The raw data for the ACSI are collected on an annual basis by random telephone surveys with about 250 customers of each organization's goods or services. For a given year, the database contains more than 200,000 customer responses. The data collection process is carried out in a manner so that the final data are comparable across industries (Fornell et al., 1996). Respondents are asked to rate each firm on the basis of a set of 15 items that are then used to define six constructs, centered around overall customer satisfaction. Fornell et al. (1996) identified three antecedent constructs of overall customer satisfaction: perceived quality, customer expectations of quality, and perceived value. The overall customer satisfaction construct itself directly predicts both customer loyalty and customer complaints.

This study draws on ACSI data from the first quarter of 1999 including 17,265 observations for all industries. In terms of observed heterogeneity, we focus on data for the utilities sector ($N = 7{,}053$) and the hotel industry ($N = 2{,}879$). To ensure the validity of our analysis, we adjusted the dataset by carrying out a missing value analysis. In standard PLS path model estimations, researchers frequently revert to mean replacement algorithms. However, when replacing relatively high numbers of missing values per variable and case by mean values, latent class segmentation methods such as FIMIX-PLS (Hahn et al., 2002; Sarstedt & Ringle, 2010) will most likely form these observations into their own segment. As we aim at uncovering unobserved heterogeneity by applying FIMIX-PLS in this chapter, incomplete observations were excluded (i.e., casewise deletion). As this procedure would have led to the exclusion of a vast number of observations, we decided to reduce the original ACSI model as presented by Fornell et al.'s (1996). Consequently, we excluded two items from the customer loyalty construct and the construct customer complaints, measured by a binary single item, as these items had a high number of missing values. Thus, this study retained "likely to repurchase" as the sole indicator of customer loyalty, but otherwise included all of the items used by Fornell et al. (1996) to measure the constructs customer expectations of quality, perceived quality, perceived value, overall customer satisfaction, and customer loyalty. The final sample hence comprises 10,417 observations for all industries and, more specifically, $N = 4{,}015$ observations for the utilities sector and $N = 1{,}383$ observations for the hotel sector. Fig. 1 illustrates the revised ACSI model that we employ for this study.

All latent variables (e.g., perceived value and overall customer satisfaction) use a "mode A" specification for their items (i.e., manifest or observed variables) in their measurement models, which is associated with reflective measurement (Hair, Ringle, & Sarstedt, 2011). Alternatively, "mode B" that

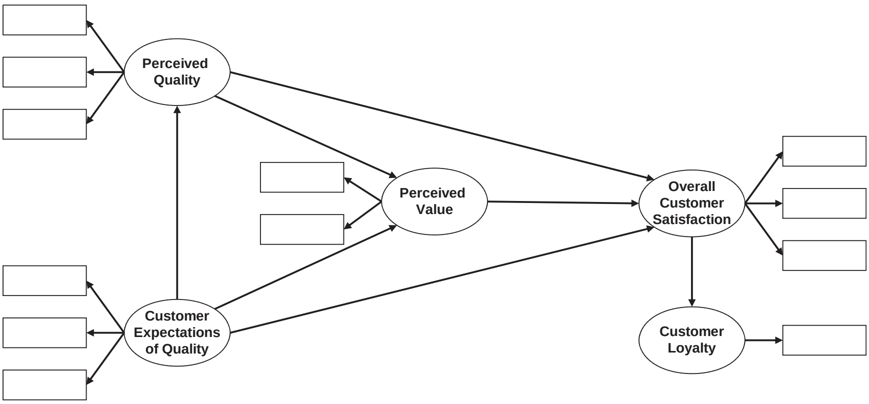

Fig. 1. Revised ACSI Model.

is associated with formative measurement models and relationships from the items to the latent variable is feasible (Diamantopoulos, Riefler, & Roth, 2008; Diamantopoulos & Winklhofer, 2001). A theoretical evaluation of the measurement models' "mode A" by qualitative criteria (Jarvis, MacKenzie, & Podsakoff, 2003) *and* an empirical assessment by using the confirmatory tetrad analysis for PLS path modeling (CTA-PLS; Gudergan, Ringle, Wende, & Will, 2008) do not give evidence for the appropriateness of the alternative "mode B" measurement model specification. Hence, in this study, we draw on the ACSI path model with reflective measurement models as depicted in Fig. 1.

MODEL ESTIMATION

For the estimation of our modified ACSI model with empirical data, we use the PLS path modeling method (Hair, Sarstedt, Ringle, & Mena, 2012; Lohmöller, 1989; Wold, 1982) and the SmartPLS 2.0 software application (Ringle et al., 2005). Table 1 presents the results on the aggregate data level (i.e., for all industries in the ACSI data). To analyze and evaluate the PLS path modeling results, we follow recommendations by Henseler, Ringle, and Sinkovics (2009) and Hair et al. (2012).

Measurement model parameter estimates and diagnostics provide evidence for the reliability and validity of the reflective construct measures.

Table 1. Overall and Industry-Specific PLS Path Modeling Results for the ACSI Model.

		Overall	Utilities	Hotels	$\lvert\Delta\rvert$
Number of observations		10,417	4,015	1,383	2,632
Path coefficients	Customer expectations of quality → perceived quality	.556***	.569***	.495***	.074**
	Customer expectations of quality → perceived value	.072***	.132***	−.018	.150***
	Customer expectations of quality → overall customer satisfaction	.021***	.022***	.047***	.026
	Perceived quality → overall customer satisfaction	.558***	.496***	.568***	.072**
	Perceived quality → perceived value	.619***	.547***	.694***	.147***
	Perceived value → overall customer satisfaction	.394***	.450***	.360***	.090***
	Overall customer satisfaction → customer loyalty	.686***	.734***	.695***	.039*
R^2	Perceived quality	.309	.324	.245	.079
	Perceived value	.439	.399	.470	.071
	Customer loyalty	.471	.538	.483	.055
	Overall customer satisfaction	.777	.747	.772	.025
ρ_c	Customer expectations of quality	.822	.802	.798	.004
	Perceived quality	.894	.882	.875	.007
	Perceived value	.940	.925	.947	.022
	Overall customer satisfaction	.927	.914	.915	.001
AVE	Customer expectations of quality	.610	.581	.566	.015
	Perceived quality	.739	.716	.706	.010
	Perceived value	.887	.860	.900	.040
	Overall customer satisfaction	.808	.779	.781	.002

Note: $\lvert\Delta\rvert$ = Absolute difference of utilities' data and hotels' data results.
***Significant at .01 (reported for path coefficients only).
**Significant at .05 (reported for path coefficients only).
*Significant at .10 (reported for path coefficients only).

All multi-item scales exhibit average variance extracted (AVE) values and composite reliability (ρ_c) values well above the commonly suggested thresholds of .50 for the AVE and .70 for ρ_c. Computations of the Fornell and Larcker (1981) criterion provide evidence for the constructs' discriminant validity (i.e., the squared root of each latent variable's AVE is higher than its correlations with other latent variables).

Evaluation of the prediction-oriented PLS path modeling method's results for the structural model centers on the R^2 values. The key target construct, overall customer satisfaction, exhibits a relatively high R^2 value of above .70 (i.e., the ACSI model explains overall customer satisfaction by more than 70%), whereas all other constructs show moderate levels of resulting R^2 values. The standardized path coefficients provide the basis for assessing the relative importance of relationships in the ACSI model. To test whether path coefficients differ significantly from zero, we calculated t-values using a bootstrapping routine[1] (Henseler et al., 2009). The analysis substantiates that all relationships in the structural model have statistically significant estimates. With a path coefficient of .558, perceived quality is the most important construct to explain overall customer satisfaction. In contrast, customer expectation of quality has the weakest effect (path coefficient of .021) on overall customer satisfaction. With the exception of the path relationship from customer expectations of quality to perceived value (.072), all additional path coefficients in the structural model have relatively high values of at least .394. The important link between overall customer satisfaction and customer loyalty has the highest coefficient (.686) for the PLS path model estimation on the aggregate data level.

The results presented in this replication study are consistent with the results and findings on the overall set of data in the original ACSI study by Fornell et al. (1996). However, we assume that these results on the aggregate data level are affected by observed heterogeneity (Haenlein and Kaplan 2011) and unobserved heterogeneity (Rigdon et al., 2010).

OBSERVED HETEROGENEITY BY INDUSTRY

Conceptual assumptions and theoretical considerations guide forming groups of data by an explanatory variable (e.g., age, income, country, and industry). This kind of information allows conducting moderator (Henseler & Fassott, 2010) and multigroup (Chin & Dibbern, 2010; see also the chapter by Sarstedt, Henseler, & Ringle, in this volume) analyses in PLS path modeling. Significant differences in the group specific results represent observed heterogeneity. In this study, we assume – in accordance with, for instance, Johnson et al. (2002) – that the PLS path modeling results for the ACSI model differ across industries and compare the ACSI model results for two presumably heterogeneous industries with a relatively large sample size (i.e., $N = 4{,}015$ observations for utilities and $N = 1{,}383$ observations for hotels).

Table 1 presents the group-specific PLS path modeling results for the utilities and hotels industries. The previous findings for the evaluation of reflective measurement models also hold for the industry-specific subsets. In the structural model, the R^2 values of latent variables in the industry-specific ACSI model estimation also show the same rank order compared with the results for the R^2 values on the aggregate data level. The bootstrapping results[2] show that – with a single exception (i.e., the weak relationship from customer expectations of quality to perceived value for the hotels industry) – all paths in each industry are significant. In comparison with the aggregate level results, the industry-specific path coefficients show differences (in absolute values) in the range from .001 to .090. Nevertheless, structural model path coefficients principally exhibit the same order of relative relevance.

The industry-specific ACSI model estimations, however, entail stronger differences that range (in absolute values) between .026 and .150. The PLS multigroup analysis (PLS-MGA), conducted by means of Chin and Dibbern's (2010) permutation test routine, substantiates that all structural model relationships – with one exception (i.e., the relationship from customer expectations of quality to overall customer satisfaction) – significantly differ across industries. For example, in comparison with the hotel group, the utilities group exhibits a considerably *stronger* relationship ($|\Delta| = .150$) between customer expectations of quality and perceived value, and a *weaker* relationship ($|\Delta| = .147$) between perceived quality and perceived value. Differences across these two industries certainly make sense. Customers experience utilities every day and have well-defined quality expectations. By contrast, customers have less experience with hotels and encounter more variation both in quality and in strategies to deliver value, challenging their expectations.

To better illustrate the consequence of observed heterogeneity, we provide an importance-performance map of the group-specific PLS path modeling results for the hotels and the utilities industry (Höck et al., 2010; Slack, 1994; Völckner et al., 2010). For a particular endogenous latent variable (i.e., customer loyalty), the assessment builds on the PLS estimates for the preceding path model relationships (i.e., total effects) and adds index values (i.e., average latent variable scores of the preceding constructs) and, thereby, an additional performance dimension to the analysis. However, the computation of index values is carried out by means of rescaling the latent variable scores to a range of 0 and 100 (Anderson & Fornell, 2000).[3] Fig. 2 presents the importance-performance map for customer loyalty that includes the index values of the latent variables in the structural model – except customer loyalty, which is the target construct – and the total effects of preceding latent variables on customer loyalty.

The constructs' index values, one of the ACSI project's key deliverables, point to generally small differences across industries. The largest difference across industries is in the index value of customer loyalty, with a value of 66.41 for hotels and 71.67 for utilities. Considering that customer loyalty is defined solely as repurchase intent, the directionality of this slight difference certainly makes sense. Customers have much greater choice when it comes to hotels, as opposed to utilities, in most US markets. All in all, this across-industry analysis suggests more similarities than differences, perhaps leading one to believe that heterogeneity is not an issue, even across industries.

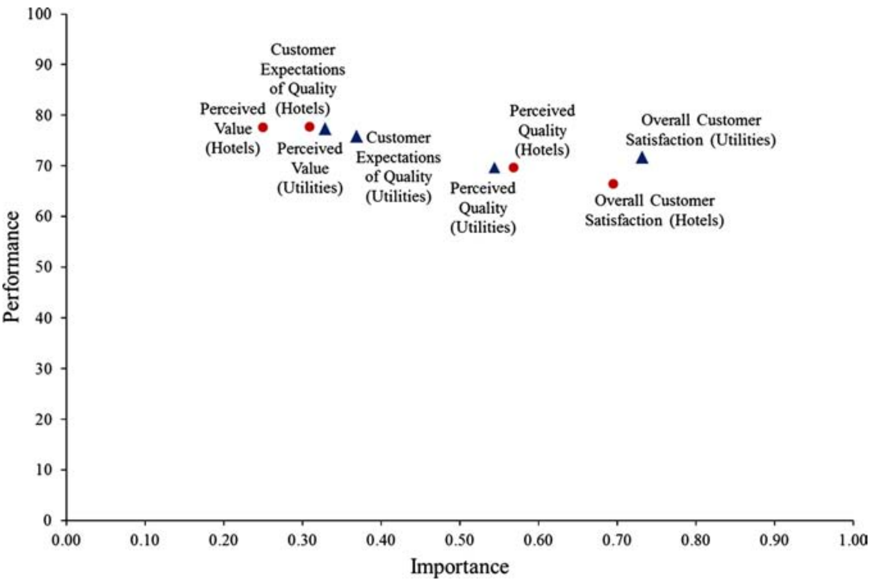

Fig. 2. Importance-Performance Map for Customer Loyalty in the Utilities and Hotel Industries.

UNCOVERING UNOBSERVED HETEROGENEITY WITHIN INDUSTRIES

In research that spans multiple nations or cultures, it is natural to address heterogeneity in terms of variation across borders. In PLS path modeling, this kind of examination may be conducted by moderator (Henseler & Fassott, 2010) and multigroup (Chin & Dibbern, 2010) analyses. However, modeling segments based on a priori information suffers from serious limitations. In many instances, substantive theory on the variables causing heterogeneity is unavailable or incomplete. Furthermore, observable characteristics such as nationality often do not capture even the majority of heterogeneity present in the data (Wedel & Kamakura, 2000). Even if a priori theory and associated observed variables are able to account to some extent for heterogeneity, a large amount of heterogeneity is frequently unobservable and its true sources are unknown. Studies report the existence of substantial consumer heterogeneity within a given product, service, or industry class (Wu & Desarbo, 2005). Hence, it is important to also uncover unobserved heterogeneity within segments defined by variables such as culture or nationality.

Among other approaches, the FIMIX-PLS method (Hahn et al., 2002; Sarstedt & Ringle, 2010) represents one of the best-known approaches for uncovering unobserved heterogeneity in PLS path modeling (Sarstedt, 2008). The FIMIX-PLS approach describes a dataset as representing the combined influence of a specified number of subpopulations (segments), each with its own distinct set of parameter values (e.g., McLachlan & Peel, 2000; Frühwirth-Schnatter, 2006). Assuming that each endogenous construct is distributed as a mixture of conditional multivariate normal densities, FIMIX-PLS uses a maximum likelihood approach to uncover latent classes. More specifically, the FIMIX-PLS approach uses scores for the constructs in the structural model to identify heterogeneity in the relationships between constructs. Simultaneously, the method also calculates the probability of each observation belonging to each subpopulation (segment) (Sarstedt, Becker, Ringle, & Schwaiger, 2011). Prior applications demonstrate this method's effectiveness for uncovering and explaining unobserved heterogeneity in PLS path modeling leading to further differentiated and more effective findings, conclusions, and recommendations (e.g., Navarro, Acedo, Losada, & Ruzo, 2011; Ringle, Wende, & Will, 2010a; Sarstedt et al., 2009). Still, it is important to remember that a mixture model solution involving multiple subpopulations does not necessarily mean that actual distinct subpopulations exist. The mixture model solution may be just one way to represent the data (Cudeck & Henley, 2003; Lubke & Spies, 2008), but this representation may be nevertheless a valuable alternative perspective.

This research uses the FIMIX-PLS module of SmartPLS (Ringle et al., 2005) to examine differences across observations based on the estimated group-specific construct scores for the ACSI model. Starting with the one-segment solution, the number of segments is increased sequentially until the relative segment size of additional latent classes are so small (e.g., about 5% of the sample) that extra segments become uninterpretable and not managerially relevant (i.e., the $K = 7$ segment solution in this ACSI application; Table 3) (Sarstedt & Ringle, 2010). As with all mixture models, local optima are always a concern, and therefore, this analysis involved 30 random start replications of the algorithm for each industry and number of segments, selecting the best solution (Ringle, Sarstedt, & Mooi, 2010b).

Information and classification criteria values (Hahn et al., 2002; Sarstedt et al., 2011) indicate that unobserved heterogeneity within industries represents a critical issue (Table 2). For the FIMIX-PLS results evaluation and the decision on the best fitting number of segments, we rely on information criteria. The criteria's minimum value indicates the most suitable number of segments. The simulation study by Sarstedt et al. (2011) however reveals serious under- and overfitting issues for the different segment retention criteria in FIMIX-PLS, consistent with other research findings (Andrews & Currim, 2003; Hawkins, Allen, & Stromberg, 2001). Among the best performing criteria, BIC and CAIC show a strong underfitting tendency, AIC_3 is subject to strong overfitting, and HQ and AIC_4 exhibit both over- and

Table 2. FIMIX-PLS Results for Segment Retention Criteria (*K* = *Number of Prespecified Segments*).

	Utilities						
	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$	$K=6$	$K=7$
CAIC	33,523.88	31,656.88	30,977.59	31,336.87	31,034.17	31,337.95	31,538.29
AIC_3	33,454.58	31,511.98	30,719.28	31,040.75	30,851.46	30,890.63	31,317.78
EN	n/a	.52	.65	.52	.44	.44	.46
	Hotels						
CAIC	11,565.73	11,052.55	10,966.78	11,173.98	11,220.33	11,298.16	11,200.05
AIC_3	11,508.18	10,869.43	10,720.87	10,865.30	10,848.86	10,863.91	10,803.01
EN	n/a	.47	.59	.52	.46	.53	.53

underfitting tendencies (for a description and formal presentation of several information and classification criteria, see Sarstedt et al., 2011). In accordance with Sarstedt et al. (2011), the decision of how many segments to retain from the data is based on a joint consideration of AIC_3 and CAIC. In addition, this study draws on the entropy normed (EN) criterion to ensure that the segments are sufficiently distinct (Ringle et al., 2010a).

In this ACSI study for the utilities and hotels industries, the one-segment solution is clearly inferior as segment retention criteria values are considerably better (smaller) for solutions with two or more segments. Table 2 summarizes the results of the relevant segment retention criteria (i.e., CAIC, AIC_3, and EN). For both the utilities and the hotel industry data, information criteria suggest that a three-segment solution may be appropriate. In both industries, we find the minimum values for CAIC and AIC_3 for $K=3$, which provides strong support for the adequacy of this segment solution (Sarstedt et al., 2011).

Even though Table 2 notes some improvements when switching from the $K=4$ to $K=5$ segments solution, the results for the criteria remain above the minimum outcome for $K=3$. Moreover, solutions for higher numbers of

Table 3. FIMIX-PLS Results for the Relative Segment Sizes (*K* = *Number of Prespecified Segments*).

	1	2	3	4	5	6	7	Sum
				Utilities				
$K=1$	1.000							1.000
$K=2$	.656	.344						1.000
$K=3$	.571	.264	.165					1.000
$K=4$	.627	.228	.103	.042				1.000
$K=5$	.310	.304	.168	.144	.073			1.000
$K=6$	.495	.161	.131	.091	.088	.034		1.000
$K=7$	.470	.173	.103	.074	.067	.057	.055	1.000
				Hotels				
$K=1$	1.000							1.000
$K=2$	.648	.352						1.000
$K=3$	.526	.338	.136					1.000
$K=4$	.465	.312	.121	.102				1.000
$K=5$	.468	.260	.125	.085	.061			1.000
$K=6$	.419	.203	.126	.116	.075	.060		1.000
$K=7$	.422	.244	.151	.060	.058	.045	.019	1.000

segments are not favorable when taking the development of relative segment sizes into account (Table 3). For the utilities group, only solutions with three or fewer segments avoid relative segment sizes of .05 or less. For the hotel group, only solutions with four or fewer segments meet this criterion for extracting relevant and interpretable segments.

Another important indicator for model selection is the normed entropy statistic (EN), proposed by Ramaswamy, DeSarbo, Reibstein, and Robinson (1993), which reveals whether a solution provides well-separated segments. Ranging between 0 and 1, a higher EN value indicates a greater probability of assigning each observation to one particular segment, whereas lower values suggest that observation could be assigned to multiple classes with similar likelihood. In the present analysis, for both industries, EN is maximized for a three-segment solution (Table 2), which supports our findings for the information criteria.

To summarize, in the light of the information criteria results, we opt for a three-segment solution for both industries. In addition, the high three-segment solution's EN for both industries is well above the critical value of .50 (Ringle et al., 2010a) and, thus, provides well separated groups of data for the ex post analysis.

SEGMENT-SPECIFIC FIMIX-PLS RESULTS

Following Ringle et al.'s (2010b) recommendations for carrying out FIMIX-PLS analyses, in the next step, we partitioned the industry datasets for utilities and hotels by assigning each observation to the segment with the maximum assignment probability. This way of splitting the FIMIX-PLS segments (i.e., that are based on an observation's probability of membership) into separated (i.e., disjunctive) groups of data provides the basis for estimating the ACSI model for each uncovered segments within an industry. Tables 4 and 5 summarize the results for the segment-specific PLS path modeling results. Their assessment confirms that all relevant criteria for the measurement and structural model evaluation (Hair et al., 2011; Henseler et al., 2009) have been satisfactorily met. For instance, all relevant criteria for reliability and validity reveal satisfactory results for the measures. For the structural model, the R^2 values for overall customer satisfaction are high, whereas R^2 values for loyalty are more moderate. The bootstrapping results reveal that almost all structural model path coefficients are statistically significant.[4] Additional analyses show that all constructs exhibit discriminant validity.

Table 4. FIMIX-PLS Results for the Three-Segment Solutions of Utilities.

		Segment 1	Segment 2	Segment 3	$\lvert\Delta_{12}\rvert$	$\lvert\Delta_{13}\rvert$	$\lvert\Delta_{23}\rvert$
Number of observations		2,924	687	404			
Relative segment size		.73	.17	.10			
Path coefficients	Customer expectations of quality → perceived quality	.71***	.23***	.98***	.48***	.27***	.75***
	Customer expectations of quality → perceived value	.19***	−.12***	.94***	.31***	.76***	1.06***
	Customer expectations of quality → overall customer satisfaction	.08***	−.09***	.31***	.17***	.23***	.39***
	Perceived quality → overall customer satisfaction	.41***	.33***	.29***	.08***	.12***	.04***
	Perceived quality → perceived value	.54***	.42***	−.05	.12***	.59***	.47***
	Perceived value → overall customer satisfaction	.49***	.59***	.41***	.10***	.08**	.19***
	Overall customer satisfaction → customer loyalty	.79***	.51***	.96***	.29***	.17***	.46***
R^2	Perceived quality	.50	.05	.95	.45	−.45	−.90
	Perceived value	.47	.17	.80	.30	−.33	−.63
	Overall customer satisfaction	.78	.60	.95	.17	−.17	−.35
	Customer loyalty	.63	.26	.93	.37	−.30	−.68
ρ_c	Customer expectations of quality	.80	.74	.89	.06	−.09	−.15
	Perceived quality	.86	.89	.91	−.03	−.05	−.02
	Perceived value	.92	.90	.93	.01	−.01	−.03
	Overall customer satisfaction	.90	.90	.95	.00	−.05	−.06
AVE	Customer expectations of quality	.58	.50	.74	.09	−.15	−.24
	Perceived quality	.68	.73	.77	−.05	−.09	−.04
	Perceived value	.85	.83	.87	.03	−.02	−.05
	Overall customer satisfaction	.75	.75	.87	.00	−.12	−.13

Notes: $\lvert\Delta_{ij}\rvert$, absolute differences between path coefficients between groups *i* and *j*; permutation-based multigroup comparison test by Chin and Dibbern (2010) for the *path coefficients* only.
***Significant at .01 (reported for path coefficients only).
**Significant at .05 (reported for path coefficients only).
*Significant at .10 (reported for path coefficients only).

Table 5. FIMIX-PLS Results for the Three-Segment Solutions of Hotels.

		Segment 1	Segment 2	Segment 3	$\|\Delta_{12}\|$	$\|\Delta_{13}\|$	$\|\Delta_{23}\|$
Number of observations		845	424	114			
Relative segment size		.61	.31	.08			
Path coefficients	Customer expectations of quality → perceived quality	.79***	.40***	−.09***	.39***	.88***	.49***
	Customer expectations of quality → perceived value	.16***	−.14	−.27***	.30***	.43***	.13
	Customer expectations of quality → overall customer satisfaction	.14***	−.06	.08	.20***	.06	.14*
	Perceived quality → overall customer satisfaction	.45***	.57***	.39***	.13***	.05	.18**
	Perceived quality → perceived value	.66***	.28***	.74***	.38***	.08	.46***
	Perceived value → overall customer satisfaction	.42***	.25***	.59***	.17***	.17***	.35***
	Overall customer satisfaction → customer loyalty	.83***	.18***	.95***	.65***	.12***	.77***
R^2	Perceived quality	.63	.16	.01	.47	.62	.16
	Perceived value	.63	.07	.65	.57	−.02	−.58
	Overall customer satisfaction	.87	.43	.83	.44	.04	−.40
	Customer loyalty	.68	.03	.89	.65	−.21	−.86
ρ_c	Customer expectations of quality	.81	.75	.71	.06	.10	.04
	Perceived quality	.87	.75	.90	.12	−.03	−.15
	Perceived value	.95	.91	.95	.04	.00	−.04
	Overall customer satisfaction	.92	.80	.96	.12	−.05	−.16
AVE	Customer expectations of quality	.60	.53	.47	.07	.13	.06
	Perceived quality	.69	.53	.75	.16	−.06	−.22
	Perceived value	.91	.83	.90	.08	.01	−.07
	Overall customer satisfaction	.79	.58	.90	.21	−.11	−.32

Notes: $|\Delta_{ij}|$, absolute differences between path coefficients between groups *i* and *j*; permutation-based multigroup comparison test by Chin and Dibbern (2010) for the *path coefficients* only.
***Significant at .01 (reported for path coefficients only).
**Significant at .05 (reported for path coefficients only).
*Significant at .10 (reported for path coefficients only).

For both industries, the FIMIX-PLS analysis uncovers one large segment 1 and two rather small segments 2 and 3 (Tables 4 and 5). The segment-specific results for each industry's large segment are similar (i.e., the mean absolute deviation of structural model path coefficients has a value of .06). The smaller segments however differ across industries. For this reason, we now analyze the segment-specific result for each industry separately.

Utilities

For utilities, we find one large segment with a relative segment size of 73%. This segment 1 may be characterized as incorporating *average* customers reflecting those who seem to be *somewhat loyal* (i.e., the index value for customer loyalty is 74.09 of 100 points). Segment 2 (size of 17%) includes those respondents who are *less loyal* (i.e., the index value for customer loyalty is 52.65 of 100 points), and segment 3 (size of 10%) captures those who are *loyal* (i.e., the index value for customer loyalty is 86.69 of 100 points). In segment 1, the strongest relationships in the structural model exist between overall customer satisfaction and customer loyalty (.79) and customer expectations of quality and perceived quality (.71). The customer expectations of quality to overall customer satisfaction relationship (.08), which may be subject to strong mediation, and the customer expectations of quality to perceived value relationship (.19) have particularly low values. All other path coefficients in the structural model have medium-high to high values that range between .41 and .54 (Table 4). Interesting differences appear when comparing the three segments. Notably, the effect of customer expectations of quality on perceived quality is .23 in segment 2 and .98 in segment 3; the one of overall customer satisfaction on customer loyalty .51 in segment 2 and .96 in segment 3; and the one of customer expectations of quality on perceived value is −.12 in segment 2 and .94 in segment 3. Overall, then, segment 3 appears to include highly loyal customers who have strong (and positive) perceptions of their current provider's quality of service. Segment 2, by contrast, appears to include customers who are willing and able to change providers.

We utilize the importance-performance map representation of PLS path modeling results (Höck et al., 2010; Völckner et al., 2010) to summarize and interpret the findings for the group-specific ACSI model estimation for each subsample. This kind of analysis uses the total effects of the PLS estimates (importance) and the construct index values (performance) as the axes of a grid. The importance-performance map for the target construct customer loyalty focuses not only on its direct antecedent construct, overall customer satisfaction, which always has the highest impact, but also on the importance of the three indirect (i.e., via overall customer satisfaction in the structural model) driver constructs customer expectations of quality, perceived quality, and perceived value (Fig. 3). Managerial actions can address those levers that have not only an impact on customer loyalty but also have a relatively high importance. Moreover, managers may achieve greater efficiency if they focus improvement efforts in those areas where performance is currently low – that is, where there appears to be room for improvement.

Looking at segment 1, for example, the performance level of overall customer satisfaction is relatively high (71.56 of 100 points), but it still offers sufficient potential for future improvements, which in turn can lead to higher customer loyalty. Customer expectations of quality and perceived quality appear to be the key concepts for improving customer loyalty (via

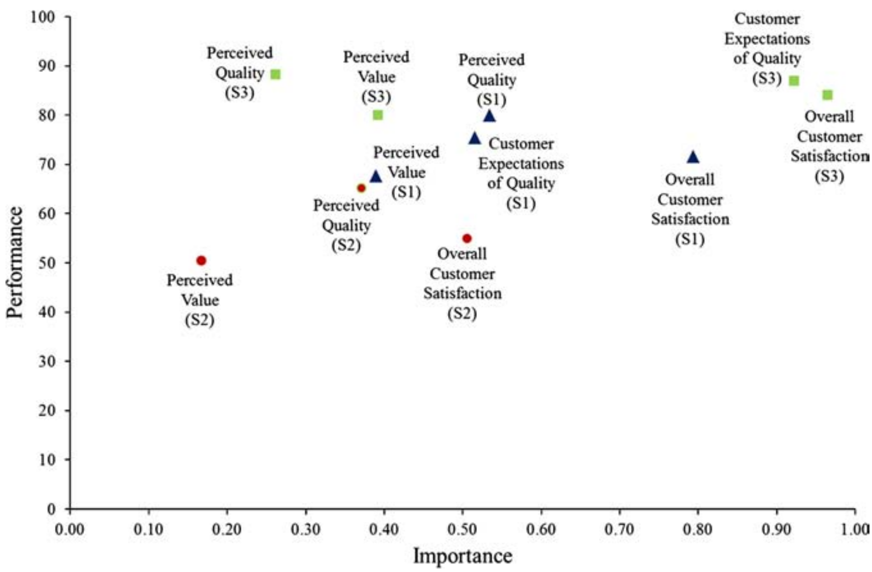

Fig. 3. Importance-Performance Map Analysis for Utilities (Target Construct: Customer Loyalty). *Notes*: S1 = segment 1; S2 = segment 2; S3 = segment 3. The total effect of customer expectations of quality on customer loyalty in segment 2 is not significant, and thus, the customer expectations of quality (S2) construct is not included in the figure.

overall customer satisfaction). Both concepts have a relatively high impact on customer loyalty (total effects: .52 and .53). However, while the performance of perceived quality (79.87 of 100 points) already is at a relatively high level in this segment, customer expectations of quality (75.38 of 100 points) offer somewhat more headroom for future improvements. Perceived value (67.00 of 100 points) offers the highest potential for future improvements, but the importance of this construct in influencing loyalty (total effect: .39) is rather low and, hence, may not justify specific managerial attention.

When analyzing the smaller segments 2 (relative segment size of 17%) and 3 (relative segment size of 10%), we find that segment 2 has low performance levels of the constructs (e.g., overall customer satisfaction and customer loyalty with index values of 54.97 and 52.65, respectively, of 100 points) and that segment 3 has very high performance levels (e.g., overall customer satisfaction and customer loyalty with index values of 84.11 and 86.69, respectively, of 100 points). The reduced mean level of overall customer satisfaction in segment 2 is associated with substantially lower construct index values for both perceived quality and perceived value. These differences between the large segment 1 (*average customers*) and the smaller segment 2 (*less loyal and less satisfied customers*), which range upward from a minimum of about 4.01 points (customer expectations of quality) to a maximum of over 21.44 points (customer loyalty), outline the need to examine heterogeneity in such data. More so, the effects observed in segment 3 (*loyal and satisfied customers*) show that both overall customer satisfaction and customer expectations of quality have a high importance (i.e., total effects of .96 and .92, respectively) in managing customer loyalty.

Common across the three segments within the utilities industry is that managing overall customer satisfaction plays an important role in enhancing customer loyalty. There are, however, differences in how to improve customer loyalty by means of influencing antecedent constructs. Notably, the role of customer expectations of quality differs.

Hotels

For the hotel industry, we identify one large segment with a relative segment size of 61%. Similarly to the utilities industry analyses, this segment 1 may be characterized as incorporating *standard* customers reflecting those who seem to be *somewhat loyal* (i.e., the index value for customer loyalty is 72.58 of 100 points). Segment 2 (size of 31%) includes those respondents who are *less loyal* (i.e., the index value for customer loyalty is 62.37 of 100 points), and segment 3 (size of 8%) captures those who are *disloyal* (i.e., the index value for customer loyalty is 35.67 of 100 points). Again, it may be important to remember that it is typically easier for customers to switch hotels than to switch utilities provider.

Similar to the other industry, in segment 1, the strongest relationships in the structural model exist between overall customer satisfaction and customer loyalty (.83) and customer expectations of quality and perceived quality (.79). Also, the customer expectations of quality to overall customer satisfaction relationship (.14), which may be subject to strong mediation, and the customer expectations of quality to perceived value relationship (.16) have particularly low values. All other path coefficients in the structural model have medium-high to high values that range between .42 and .66 (Table 5). However, remarkable differences are evident when comparing the three segments. For instance, the effect of customer expectations of quality on perceived value is .16 in segment 1, −.14 in segment 2, and −.27 in segment 3; the one of overall customer satisfaction on customer loyalty .18 in segment 2 and .95 in segment 3; and the one of perceived quality on perceived value is .28 in segment 2 and .74 in segment 3. Unlike in the utilities industry, in the hotel industry, overall customer satisfaction is not the construct with highest impact on customer loyalty across all three segments. One possible interpretation is that segment 2 consists of price shoppers who are indifferent to variations in quality (within reason) and will almost always choose the low-cost provider (i.e., hotels), whereas segment 3 customers are very sensitive to quality but are displeased with their provider and thus are inclined to switch providers precisely because their current provider has failed to meet their standards.

We again draw on the importance-performance map representation of PLS path modeling results to interpret the findings for the group-specific ACSI model estimation for each subsample within the hotel industry (Fig. 4). For the large segment 1 of hotels, we find a strong impact of overall customer satisfaction on customer loyalty (total effect: .83). Although the performance level of overall customer satisfaction is relatively high (74.29 of 100 points), it represents an adequate opportunity for investment into improvements, which in turn can lead to higher customer loyalty (granted, we do not know the cost of these improvements). Again, customer expectations of quality and perceived quality appear to be the key concepts for improving customer loyalty (via overall customer satisfaction). Both concepts have a relatively high impact on customer loyalty (total effects: .65

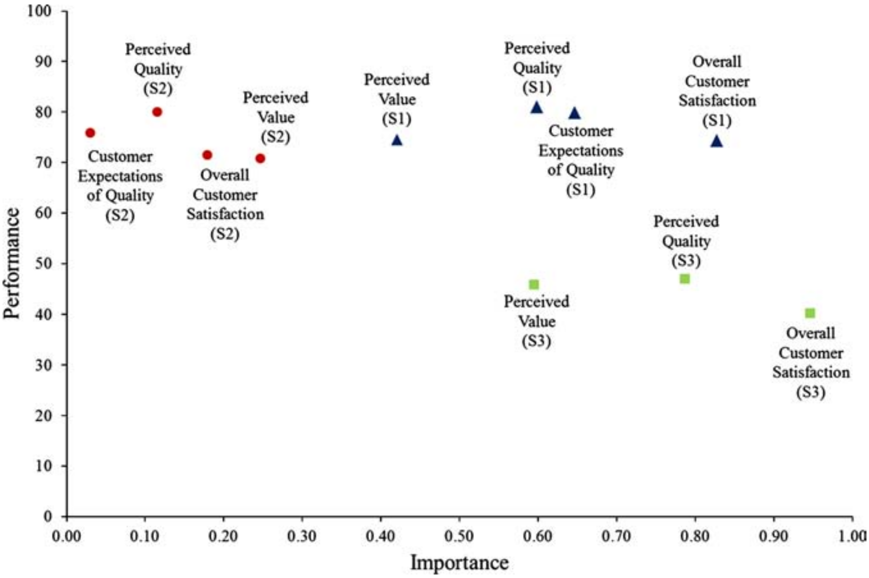

Fig. 4. Importance-Performance Map Analysis for Hotels (Target Construct: Customer Loyalty). *Notes*: S1 = segment 1; S2 = segment 2; S3 = segment 3. The total effect of customer expectations of quality on customer loyalty in segment 3 is not significant, and thus, the customer expectations of quality (S3) construct is not included in the figure.

and .60). Both are already performing at high levels (i.e., 79.79 and 80.98 of 100, respectively) and, thus, provide similar opportunity for improvement. Compared with the other constructs, the importance of perceived value (total effect: .42) is too low to warrant managerial attention.

There are differences when examining the two smaller segments. For the *less loyal customers* in segment 2, the importance of overall customer satisfaction for driving customer loyalty is very low (total effect: .18) with a moderate current performance (71.49 of 100 points), whereas for segment 3 with the *disloyal customers*, the impact is very high (total effect: .95) with a performance level of 40.22 of 100 points, which is sharply lower than the performance levels (above 80) reported by the other two segments. Similarly, in segment 3, the performance levels of perceived value (45.79 of 100 points) and of perceived quality (46.91 of 100 points), respectively, are quite low and are accompanied by sufficiently strong importance ratings (i.e., .59 and .79 respectively), whereas in segment 2, performance levels are at 70.76 and 80.08 of 100 points, respectively, and indicate very low importance (i.e., total effects of .25 and .12, respectively). These results tend to affirm a characterization of segment 3 as *dissatisfied switchers* and segment 2 as *indifferent price shoppers*.

SUMMARY AND CONCLUSION

Customer satisfaction has become a fundamental and well-documented construct in marketing that is critical in respect of demand and for any business's success given its importance and established relation with customer retention and corporate profitability (Anderson et al., 1994; Mittal et al., 2005; Morgan, Anderson, & Mittal, 2005). Although it is often acknowledged that there are no truly homogeneous segments of consumers, studies usually do not address this critical area of concern. Some studies however uncover substantial unobserved customer heterogeneity in the ACSI model (Ringle et al., 2010b) and others even within a given product, service, or industry class (Wu & Desarbo, 2005). Dealing with this unobserved heterogeneity is critical for forming groups of consumers that are homogeneous in terms of the benefits that they seek or their response to marketing programs (e.g., product offering and price discounts). Segmentation is therefore a key element for marketers in developing and improving their targeted marketing strategies.

In this study, the PLS path modeling estimations for the ACSI model on the aggregate data level and across the utilities and hotels industries differ little. Thus, modeling observed heterogeneity suggests little variance in the model across customers. By contrast, FIMIX-PLS analysis suggests that this ACSI dataset could be meaningfully conceptualized as reflecting the influence of three subpopulations or segments within both the utilities and the hotel industries. The analysis points to segments, which differ markedly from the majority segment on actionable variables and which are large enough to be strategically valuable. More generally, these results suggest that unobserved heterogeneity, defined by latent classes, may indeed be more important and more dramatic than observed heterogeneity defined by observed variables such as industry sector (Jedidi et al., 1997). Uncovering unobserved heterogeneity and assessing this phenomenon through segment-specific importance-performance map analyses may inspire further insights. This, in turn, can lead to more specific and effective managerial response.

For international marketing practice and research, obvious differences between different national or cultural groups may cause researchers to overlook potentially greater levels of heterogeneity within such groups. International marketing researchers should be alert to the possibility that supposedly homogeneous single-nation samples may actually include substantial diversity. Moreover, along with hidden differences within one sample, researchers should also be alert for similarities across national boundaries in terms of the different segments that, in fact, may underlie overall results. For example, many researchers have pointed to the rise of a global middle class, a worldwide group of consumers with strong similarities in their wants and tastes (Banerjee & Duflo, 2008; Das, 2009). Researchers should bear in mind the possibility that within data from different nations lie different mixtures of the same component classes or groups, with substantial implications for both academics and practitioners.

NOTES

1. We used the following setting when conducting the bootstrapping routine for PLS path modeling using SmartPLS: Number of cases per subsample: $N = 10{,}417$ for the overall set of data; random data generation (with replacement) for 5,000 subsamples per bootstrapping analysis; no sign change, which is the most conservative option when conducting a bootstrapping analysis for PLS path modeling.
2. We used the following setting when conducting the bootstrapping routine for PLS path modeling using SmartPLS: Number of cases per subsample: $N = 4{,}015$ for analyses on the utilities and $N = 1{,}383$ for the hotels; random data generation (with replacement) for 5,000 subsamples per bootstrapping analysis; no sign change.
3. PLS path modeling estimates the latent variable scores; the coefficients for the measurements models; and the path coefficients to obtain the direct, indirect, and total effects in the structural model so as to maximize the explained variance of the endogenous constructs. The resulting construct scores are transformed to a 0 to 100-point scale. The construct values are computed by aggregating firm-level results, weighted by firm sales (for industry-level), industry sales (for sector-level), and each sector's contribution to the US GDP (for economy-level). Notwithstanding this approach to aggregation, heterogeneity may play a role but is commonly not taken into account.
4. We used the following setting when conducting the bootstrapping routine for PLS path modeling using SmartPLS: Number of cases per subsample: *see the number of observations for each segment per industry in Tables 4 and 5*; random data generation (with replacement) for 5,000 subsamples per bootstrapping analysis; no sign change.

REFERENCES

Aksoy, L., Cooil, B., Groening, C., Keiningham, T. L., & Yalçin, A. (2008). The long-term stock market valuation of customer satisfaction. *Journal of Marketing*, *72*(4), 105–122.

Anderson, E., & Sullivan, M. (1993). The antecedents and consequences of customer satisfaction for firms. *Marketing Science*, *12*(2), 125–143.

Anderson, E. W., & Fornell, C. (2000). Foundations of the American customer satisfaction index. *Total Quality Management*, *11*(7), 869–882.

Anderson, E. W., Fornell, C. G., & Lehmann, D. R. (1994). Customer satisfaction, market share, and profitability: Findings from Sweden. *Journal of Marketing*, *58*(3), 53–66.

Anderson, E. W., Fornell, C. G., & Mazvancheryl, S. K. (2004). Customer satisfaction and shareholder value. *Journal of Marketing*, *68*(4), 172–185.

Anderson, E. W., & Mansi, S. A. (2009). Does customer satisfaction matter to investors? Findings from the bond market. *Journal of Marketing Research*, *46*(5), 703–714.

Andreassen, T. W., & Lindestad, B. (1998). The effects of corporate image in the formation of customer loyalty. *Journal of Services Marketing*, *1*(1), 82–92.

Andrews, R. L., & Currim, I. S. (2003). Retention of latent segments in regression-based marketing models. *International Journal of Research in Marketing*, *20*(4), 315–321.

Banerjee, A. V., & Duflo, E. (2008). What is middle class about the middle classes around the world? *Journal of Economic Perspectives*, *22*(2), 3–28.

Chin, W. W., & Dibbern, J. (2010). A permutation based procedure for multi-group PLS analysis: Results of tests of differences on simulated data and a cross cultural analysis of the sourcing of information system services between Germany and the USA. In: V. Esposito Vinzi, W.W. Chin, J. Henseler, & H. Wang (Eds), *Handbook of partial least squares: Concepts, methods and applications in marketing and related fields* (pp. 171–193). Berlin: Springer.

Cudeck, R., & Henley, S. J. (2003). A realistic perspective on pattern representation in growth data: Comment on Curran and Bauer. *Psychological Methods*, *8*(3), 378–383.

Das, D. K. (2009). Globalisation and an emerging global middle class. *Economic Affairs*, *29*(3), 89–92.

Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research*, *61*(12), 1203–1218.

Diamantopoulos, A., & Winklhofer, H. M. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research*, *38*(2), 269–277.

Eklöf, J. A., & Westlund, A. H. (2002). The pan-European customer satisfaction index programme – Current work and the way ahead. *Total Quality Management & Business Excellence*, *13*(8), 1099–1106.

Fornell, C. G. (1992). A national customer satisfaction barometer: The Swedish experience. *Journal of Marketing*, *56*(1), 6–21.

Fornell, C. G., Johnson, M. D., Anderson, E. W., Cha, J., & Bryant, B. E. (1996). The American customer satisfaction index: Nature, purpose, and findings. *Journal of Marketing*, *60*(4), 7–18.

Fornell, C. G., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39–50.

Frühwirth-Schnatter, S. (2006). *Finite mixture and markov switching models*. New York, NY: Springer.

Gruca, T. S., & Rego, L. L. (2005). Customer satisfaction, cash flow, and shareholder value. *Journal of Marketing*, *69*(3), 115–130.

Gudergan, S. P., Ringle, C. M., Wende, S., & Will, A. (2008). Confirmatory tetrad analysis in PLS path modeling. *Journal of Business Research*, *61*(12), 1238–1249.

Haenlein, M., & Kaplan, A. M. (2011). The influence of observed heterogeneity on path coefficient significance: Technology acceptance within the marketing discipline. *Journal of Marketing Theory and Practice*, *19*(2), 153–168.

Hahn, C., Johnson, M. D., Herrmann, A., & Huber, F. (2002). Capturing customer heterogeneity using a finite mixture PLS approach. *Schmalenbach Business Review*, *54*(3), 243–269.

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory and Practice*, *19*(2), 139–151.

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*, forthcoming (online available).

Hawkins, D. S., Allen, D. M., & Stromberg, A. J. (2001). Determining the number of components in mixtures of linear models. *Computational Statistics & Data Analysis*, *38*(1), 15–48.

Henseler, J., & Fassott, G. (2010). Testing moderating effects in PLS path models: An illustration of available procedures. In: V. Esposito Vinzi, W. W. Chin, J. Henseler & H. Wang (Eds.), *Handbook of partial least squares: Concepts, methods and applications in marketing and related fields* (pp. 713–735). Berlin: Springer.

Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing. *Advances in International Marketing*, *20*, 277–320.

Höck, C., Ringle, C. M., & Sarstedt, M. (2010). Management of multi-purpose stadiums: Importance and performance measurement of service interfaces. *International Journal of Services Technology and Management*, *14*(2/3), 188–207.

Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research*, *30*(2), 199–218.

Jedidi, K., Jagpal, H. S., & DeSarbo, W. S. (1997). Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science*, *16*(1), 39–59.

Johnson, M. D., Herrmann, A., & Gustafsson, A. (2002). Comparing customer satisfaction across industries and countries. *Journal of Economic Psychology*, *23*(6), 749–769.

Lohmöller, J.-B. (1989). *Latent variable path modeling with partial least squares*. Heidelberg: Physica.

Lou, X., & Homburg, C. (2007). Neglected outcomes of customer satisfaction. *Journal of Marketing*, *71*(3), 133–149.

Lou, X., & Homburg, C. (2008). Satisfaction, complaint, and the stock value gap. *Journal of Marketing*, *72*(4), 29–43.

Lubke, G. H., & Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods*, *10*(1), 21–39.

Lubke, G. H., & Spies, J. R. (2008). Choosing a 'correct' factor mixture model: Power, limitations, and graphical data exploration. In: G. R. Hancock & K. M. Samuelson (Eds.), *Latent variable mixture models* (pp. 343–361). Charlotte, NC: Information Age Publishing.

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York, NY: Wiley.

Mittal, V., Anderson, E. W., Sayrak, A., & Tadikamalla, P. (2005). Dual emphasis and the long-term financial impact of customer satisfaction. *Marketing Science*, *24*(4), 531–543.

Morgan, N., Anderson, E. W., & Mittal, V. (2005). Understanding firms' customer satisfaction information usage. *Journal of Marketing*, *69*(3), 121–135.

Navarro, A., Acedo, F. J., Losada, F., & Ruzo, E. (2011). Integrated model of export activity: Analysis of heterogeneity in managers' orientations and perceptions on strategic marketing management in foreign markets. *Journal of Marketing Theory and Practice*, *19*(2), 187–204.

Raithel, S., Sarstedt, M., Scharf, S., & Schwaiger, M. (2011). On the value relevance of customer satisfaction. Multiple drivers and multiple markets. *Journal of the Academy of Marketing Science*.

Ramaswamy, V., DeSarbo, W. S., Reibstein, D. J., & Robinson, W. T. (1993). An empirical pooling approach for estimating marketing mix elasticities with PIMS data. *Marketing Science*, *12*(1), 103–124.

Rigdon, E. E., Ringle, C. M., & Sarstedt, M. (2010). Structural modeling of heterogeneous data with partial least squares. In: N. K. Malhotra (Ed.), *Review of marketing research* (Vol. 7, pp. 255–296). Armonk, NY: Sharpe.

Ringle, C. M., Wende, S., & Will, A. (2005). *SmartPLS 2.0 (Beta)* (Retrieved from www.smartpls.de). Hamburg: SmartPLS.

Ringle, C. M., Wende, S., & Will, A. (2010a). Finite mixture partial least squares analysis: Methodology and numerical examples. In: V. Esposito Vinzi, W. W. Chin, J. Henseler & H. Wang (Eds.), *Handbook of partial least squares: Concepts, methods and applications in marketing and related fields* (pp. 195–218). Berlin: Springer.

Ringle, C. M., Sarstedt, M., & Mooi, E. A. (2010b). Response-based segmentation using finite mixture partial least squares: Theoretical foundations and an application to American customer satisfaction index data. *Annals of Information Systems*, *8*, 19–49.

Sarstedt, M. (2008). A review of recent approaches for capturing heterogeneity in partial least squares path modelling. *Journal of Modelling in Management*, *3*(2), 140–161.

Sarstedt, M., Becker, J.-M., Ringle, C. M., & Schwaiger, M. (2011). Uncovering and treating unobserved heterogeneity with FIMIX-PLS: Which model selection criterion provides an appropriate number of segments? *Schmalenbach Business Review*, *63*(1), 34–62.

Sarstedt, M., & Ringle, C. M. (2010). Treating unobserved heterogeneity in PLS path modelling: A comparison of FIMIX-PLS with different data analysis strategies. *Journal of Applied Statistics*, *37*(8), 1299–1318.

Sarstedt, M., Schwaiger, M., & Ringle, C. M. (2009). Do we fully understand the critical success factors of customer satisfaction with industrial goods? – Extending Festge and Schwaiger's model to account for unobserved heterogeneity. *Journal of Business Market Management*, *3*(3), 185–206.

Slack, N. (1994). The importance-performance matrix as a determinant of improvement priority. *International Journal of Operations and Production Management*, *44*(5), 59–75.

Völckner, F., Sattler, H., Hennig-Thurau, T., & Ringle, C. M. (2010). The role of parent brand quality for service brand extension success. *Journal of Service Research*, *13*(4), 359–361.

Wedel, M., & Kamakura, W. A. (2000). *Market segmentation: Conceptual and methodological foundations* (2nd ed.). Boston, MA: Kluwer.

Wold, H. (1982). Soft modeling: The basic design and some extensions. In: K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observations: Part II* (pp. 1–54). Amsterdam: North-Holland.

Wu, J., & Desarbo, W. S. (2005). Market segmentation for customer satisfaction studies via a new latent structure multidimensional scaling model. *Applied Stochastic Models in Business and Industry*, *21*(4/5), 303–309.

MULTIGROUP ANALYSIS IN PARTIAL LEAST SQUARES (PLS) PATH MODELING: ALTERNATIVE METHODS AND EMPIRICAL RESULTS

Marko Sarstedt, Jörg Henseler and Christian M. Ringle

ABSTRACT

Purpose – *Partial least squares (PLS) path modeling has become a pivotal empirical research method in international marketing. Owing to group comparisons' important role in research on international marketing, we provide researchers with recommendations on how to conduct multigroup analyses in PLS path modeling.*

Methodology/approach – *We review available multigroup analysis methods in PLS path modeling and introduce a novel confidence set approach. A characterization of each method's strengths and limitations and a comparison of their outcomes by means of an empirical example extend the existing knowledge of multigroup analysis methods. Moreover, we provide an omnibus test of group differences (OTG), which allows testing the differences across more than two groups.*

Measurement and Research Methods in International Marketing
Advances in International Marketing, Volume 22, 195–218

ISSN: 1474-7979/doi:10.1108/S1474-7979(2011)0000022012

Findings – *The empirical comparison results suggest that Keil et al.'s (2000) parametric approach can generally be considered more liberal in terms of rendering a certain difference significant. Conversely, the novel confidence set approach and Henseler's (2007) approach are more conservative.*

Originality/value of paper – *This study is the first to deliver an in-depth analysis and a comparison of the available procedures with which to statistically assess differences between group-specific parameters in PLS path modeling. Moreover, we offer two important methodological extensions of existing research (i.e., the confidence set approach and OTG). This contribution is particularly valuable for international marketing researchers, as it offers recommendations regarding empirical applications and paves the way for future research studies aimed at comparing the approaches' properties on the basis of simulated data.*

INTRODUCTION

Studies on international marketing have frequently made use of partial least squares (PLS) path modeling (Hair, Ringle, & Sarstedt, 2011; Hair, Sarstedt, Ringle, & Mena, 2012; Lohmöller, 1989; Wold, 1975, 1982) to empirically test theoretical models (for an overview, see Henseler, Ringle, & Sinkovics, 2009). As part of international marketing researchers' toolbox, PLS path modeling has become a pivotal instrument for estimating and analyzing complex path relationships between latent variables. This method belongs to a family of alternating least squares algorithms that extend principal component analysis and canonical correlation analysis to estimate (mainly linear) relationships between latent variables (Lohmöller, 1989).

As with any other statistical method, PLS path modeling applications are usually based on the assumption that the analyzed data stem from a single population (i.e., a unique global model represents all the observations well). However, in many real-world applications, such as in international marketing, this assumption of homogeneity is unrealistic, because individuals are likely to be heterogeneous in their perceptions and evaluations of latent constructs (e.g., Jedidi, Jagpal, & DeSarbo, 1997; Sarstedt & Ringle, 2010). This notion holds specifically for research on international marketing, which often analyzes differences in parameters in respect of different subpopulations such as countries and cultures (Brettel, Engelen, Heinemann, & Vadhanasindhu, 2008; Graham, Mintu, & Rodgers, 1994; Grewal, Chakravarty, Ding, & Liechty, 2008; Rodríguez & Wilson, 2002). Although several studies explicitly broach the issue of group-specific effects in their research questions, ignoring population heterogeneity – when performing PLS path modeling on an aggregate data level – can seriously bias the results and, thereby, yield inaccurate management conclusions (Sarstedt, Schwaiger, & Ringle, 2009).

Although cross-national or cross-cultural differences are related to observed heterogeneity, there can also be unobserved heterogeneity that cannot be attributed to one (or more) pre-specified variable(s). Similar to ignoring observed heterogeneity, unobserved heterogeneity is a serious problem in respect of interpreting PLS path modeling results if it is not considered in the analysis. Various response-based segmentation approaches have recently been developed to deal with unobserved heterogeneity. These segmentation approaches generalize, for example, genetic algorithm (Ringle, Sarstedt, & Schlittgen, 2010), and typological regression approaches (Esposito Vinzi, Ringle, Squillacciotti, & Trinchera, 2007; Esposito Vinzi, Trinchera, Squillacciotti, & Tenenhaus, 2008) to PLS path modeling. Finite mixture PLS (FIMIX-PLS; Sarstedt & Ringle, 2010; Hahn, Johnson, Herrmann, & Huber, 2002; Sarstedt, Becker, Ringle, & Schwaiger, 2011) is currently regarded the primary approach of all these segmentation techniques, and has become mandatory for evaluating PLS path modeling results (Sarstedt, 2008; Hair et al., 2012). Hair et al. (2011, p. 147), for example, point out that "using this technique, researchers can either confirm that their results are not distorted by unobserved heterogeneity or they can identify thus far neglected variables that describe the uncovered data segments." Although these response-based segmentation approaches rely on different statistical concepts, they all share the same final analysis step: A comparison of the PLS parameter estimates across the identified latent segments (e.g., Rigdon, Ringle, & Sarstedt, 2010; Ringle, Sarstedt, & Mooi, 2010). Therefore, no matter whether heterogeneity is observed or unobserved, there is a need for PLS-based approaches to multigroup analysis.

Despite its obvious importance for the international marketing discipline, research on multigroup analysis is a rather new field. Only a small number of methodologically oriented articles have to date been dedicated to the discussion of available approaches (e.g., Chin & Dibbern, 2010; Rigdon et al., 2010). Researchers' discussions, for example, on internet forums like http://www.smartpls.de, show that there is a strong need to clarify how multigroup analysis can be carried out within a PLS path modeling framework. Given this background, the purpose of this chapter is to illustrate the use of multigroup analysis procedures in PLS path modeling. Specifically, we describe available multigroup analysis approaches, comment on their strengths and limitations, and illustrate their use by means of an empirical example. We also propose a novel nonparametric approach based on a comparison of bootstrap confidence intervals. This method has been designed as a more conservative approach to PLS multigroup analysis.

Prior approaches to PLS multigroup analysis are restricted in that they only allow testing the differences in two groups' parameters. However, researchers in international marketing and other cross-cultural research fields frequently encounter situations in which they would like to compare more than two groups. A naive approach would be to conduct all possible pairwise group comparisons, which would, however, quickly boost the familywise error rate beyond any prespecified acceptable Type-I error level (Mooi & Sarstedt, 2011). To overcome this problem, we introduce a permutation-based analysis of variance approach, which maintains the familywise error rate, does not rely on distributional assumptions, and exhibits an acceptable level of statistical power.

MULTIGROUP ANALYSIS IN PLS PATH MODELING

Conceptually, the comparison of group-specific effects entails the consideration of a categorical moderator variable which, in line with Baron and Kenny (1986, p. 1174), "affects the direction and/or strength of the relation between an independent or predictor variable and a dependent or criterion variable." Following this concept, group effects are nothing more than a variable's moderating effect whereby the categorical moderator variable expresses each observation's group membership (Henseler et al., 2009). As a consequence, multigroup analysis is generally regarded as a special case of modeling continuous moderating effects (Henseler & Chin, 2010; Henseler & Fassott, 2010). Fig. 1 illustrates the categorical moderator variable concept graphically. Here, x_1 to x_3 represent (reflective) indicator variables of an exogenous latent variable ξ, y_1 to y_3 represent (reflective) indicator variables of an endogenous latent variable η, and θ is the parameter of the relationship between ξ and η. Lastly, m represents a categorical moderating variable, which potentially exerts an influence on all

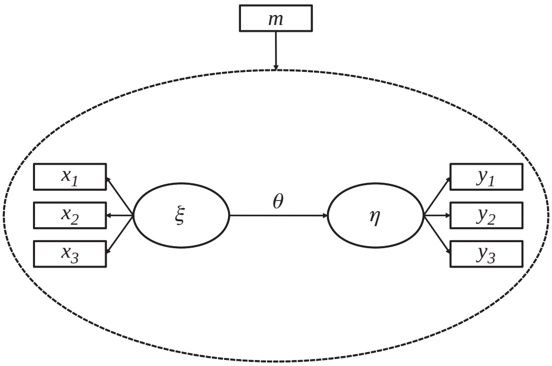

Fig. 1. Moderator Modeling Framework.

model relations. Researchers are usually interested in analyzing group effects related to structural model relations. More precisely, a population parameter θ is hypothesized as different across two subpopulations (i.e., $\theta^{(1)}$ and $\theta^{(2)}$), which are expressed by different modalities in m.

A primary concern when comparing model estimates across groups is ensuring that the construct measures are invariant across the groups. Amongst other criteria, as described by Steenkamp and Baumgartner (1998), this entails, for example, that the estimates satisfy the requirement of measurement invariance. With reference to Fig. 1, this requirement implies that the moderator variable's effect is restricted to the parameter θ and does not entail group-related differences in the item loadings.

Three approaches to multigroup analysis have been proposed within a PLS path modeling framework thus far. The first approach, introduced by Keil et al. (2000), involves estimating model parameters for each group separately, and using the standard errors obtained from bootstrapping as the input for a parametric test. This method is generally labeled the parametric approach (Henseler, 2007). Chin (2003b) proposed and further described a distribution-free data permutation test (Chin & Dibbern, 2010; Dibbern & Chin, 2005), because the parametric approach's distributional assumptions do not fit PLS path modeling's distribution-free character. This test seeks to scale the observed differences between groups by comparing these differences to those between groups randomly assembled from the data. Henseler (2007) proposed and described another nonparametric procedure, which directly compares group-specific bootstrap estimates from each bootstrap sample (see also Henseler et al., 2009).

Parametric Approach

The parametric approach was initially applied by Keil et al. (2000) (see also Chin, 2000) and depicts a modified version of the two independent samples t-test. As such, this approach requires the data (i.e., the PLS estimations of a certain path coefficient across all bootstrapping subsamples) to be normally distributed, which runs contrary to PLS path modeling's distribution-free character. Consequently, researchers should run a Kolmogorov–Smirnov test with Lilliefors correction – or, in the case of small sample sizes below 50, the Shapiro–Wilk test – to assess whether the data follow a normal distribution (Mooi & Sarstedt, 2011). In addition to carrying out these tests, researchers should also visually inspect the theoretical and empirical probability distributions by means of q–q plots (Chambers, Cleveland, Kleiner, & Tukey, 1983).

Executing the parametric test requires researchers to first run the standard PLS path modeling algorithm for each group, followed by the bootstrapping procedure (e.g., Hair et al., 2011; Henseler et al., 2009) to obtain the standard errors of the group-specific parameter estimates (Keil et al., 2000). The choice of test statistic depends on whether the parameter estimates' standard deviations differ significantly across the groups, which can be assessed by means of Levene's test. If the parameter estimates' standard deviations are equal, the test statistic is computed as follows (Keil et al., 2000; the equation provided by these authors has a flaw that we corrected):

$$t = \frac{\tilde{\theta}^{(1)} - \tilde{\theta}^{(2)}}{\sqrt{((n^{(1)}-1)^2/(n^{(1)}+n^{(2)}-2)) \cdot se^2_{\theta^{(1)}} + ((n^{(2)}-1)^2/(n^{(1)}+n^{(2)}-2)) \cdot se^2_{\theta^{(2)}}} \cdot \sqrt{(1/n^{(1)}) + (1/n^{(2)})}} \tag{1}$$

Here, $\tilde{\theta}^{(1)}$ ($\tilde{\theta}^{(2)}$) denote the original parameter estimate for a path relationship in group one (two), $n^{(1)}$ ($n^{(2)}$) the number of observations in group one (two), and $se_{\theta^{(1)}}$ ($se_{\theta^{(2)}}$) the path coefficient's standard error in group one (two) obtained from the bootstrapping procedure. Moreover, t represents the empirical t-value that must be larger than the critical value from a t-distribution with $n^{(1)} + n^{(2)} - 2$ degrees of freedom.[1] In cases where Levene's test indicates that the standard errors are unequal, the test statistic takes the following form (Chin, 2000):

$$t = \frac{\tilde{\theta}^{(1)} - \tilde{\theta}^{(2)}}{\sqrt{((n^{(1)}-1)/n^{(1)})se^2_{\theta^{(1)}} + ((n^{(2)}-1)/n^{(2)})se^2_{\theta^{(2)}}}} \tag{2}$$

This test statistic is asymptotically t-distributed and the degrees of freedom (df) are determined by means of the Welch–Satterthwaite equation. The equation below was derived by Nitzl (2010) for use in combination with bootstrapping (note that the first draft by Chin (2000) is not entirely correct):

$$df = \left\| \frac{\left((n^{(1)}-1)/n^{(1)} \cdot se^2_{\theta^{(1)}} + (n^{(2)}-1)/n^{(2)} \cdot se^2_{\theta^{(2)}}\right)^2}{(n^{(1)}-1)/n^{(1)^2}) \cdot se^4_{\theta^{(1)}} + (n^{(2)}-1)/n^{(2)^2} \cdot se^4_{\theta^{(2)}}} - 2 \right\| \tag{3}$$

Permutation-Based Approach

The permutation-based approach was developed by Chin (2003b) and subsequently further described by Chin and Dibbern (2010), as well as Dibbern and Chin (2005). Analogous to Edgington and Onghena (2007), the permutation-based test procedure builds on the observations' random assignment to groups. The procedure is as follows:

1. Run the PLS path modeling algorithm separately for each group.
2. Randomly permute the data; that is, the observations are randomly exchanged between the two groups. More precisely, $n^{(1)}$ observations are drawn without replacement and assigned to the first group; all remaining observations are assigned to the second group. Thus, in each permutation run u ($u \in \{1, \ldots, U\}$), the group-specific sample size remains constant (i.e., $n_u^{(1)} = n^{(1)}$ and $n_u^{(2)} = n^{(2)}, \forall u$). In accordance with commonly suggested rules of thumb for bootstrapping sample sizes (Hair et al., 2012), the minimum number of permutation runs should be 5,000.
3. Run the PLS path modeling algorithm for each group per permutation run u to obtain the group-specific parameter estimates $\tilde{\theta}_u^{(1)}$ and $\tilde{\theta}_u^{(2)}$.
4. Compute the differences in the permutation run-specific parameter estimates $d_u = \tilde{\theta}_u^{(1)} - \tilde{\theta}_u^{(2)}$.
5. Test the null hypothesis that the population parameters are equal across the two groups ($H_0 : \theta^{(1)} = \theta^{(2)}$).

By not relying on distributional assumptions, the permutation-based approach overcomes a key disadvantage of the parametric approach and, thus, fits the PLS path modeling method's characteristics. However, the permutation-based approach requires group-specific sample sizes to be fairly similar (Chin & Dibbern, 2010), which is a central limitation.

Henseler's PLS Multigroup Analysis

From a procedural perspective, the approach proposed by Henseler (2007) closely resembles the parametric approach. Initially, the subsamples are exposed to separate bootstrap analyses, and the bootstrap outcomes serve as a basis for testing the potential group differences. However, Henseler's (2007) approach differs in the way the bootstrap estimates are used to assess the robustness of the group-specific parameter estimates. Instead of relying on distributional assumptions, the new approach evaluates the bootstrap outcomes' observed distribution. Given two subsamples with different parameter estimates $\tilde{\theta}^{(1)}$ and $\tilde{\theta}^{(2)}$, groups can be indexed – without any loss of generality – so that $\tilde{\theta}^{(1)} > \tilde{\theta}^{(2)}$. In order to assess the significance of a group effect, the conditional probability $p(\theta^{(1)} \leq \theta^{(2)} | \tilde{\theta}^{(1)}, \tilde{\theta}^{(2)}, CDF(\theta^{(1)}), CDF(\theta^{(2)}))$ has to be determined on the basis of the group-specific parameter estimates $\tilde{\theta}^{(g)} (g \in \{1, 2\})$ and the empirical cumulative distribution functions (CDFs).

In an initial step, the centered bootstrap estimates ($\tilde{\theta}_i^{(g)\bar{*}}$) have to be computed as follows:

$$\tilde{\theta}_i^{(g)\bar{*}} = \tilde{\theta}_i^{(g)*} - \frac{1}{B} \sum_{i=1}^{B} \tilde{\theta}_i^{(g)*} + \tilde{\theta}^{(g)} \tag{4}$$

where $\tilde{\theta}_i^{(g)*}$ represents the bootstrap estimate in group g ($g \in \{1, 2\}$) and bootstrap sample i ($i \in \{1, \ldots, B\}$). By using the Heaviside step function $H(x^*)$, as defined by

$$H(x^*) = \frac{1 + \mathrm{sgn}(x^*)}{2} \tag{5}$$

and the bootstrap estimates as discrete manifestations of the CDFs, the conditional probability is computed as follows:

$$p(\theta^{(1)} \leq \theta^{(2)} | \tilde{\theta}^{(1)}, \tilde{\theta}^{(2)}, CDF(\theta^{(1)}), CDF(\theta^{(2)})) = \frac{1}{B^2} \sum_{i=1}^{B} \sum_{j=1}^{B} H\left(\tilde{\theta}_j^{(2)\bar{*}} - \tilde{\theta}_i^{(1)\bar{*}}\right) \tag{6}$$

The idea behind Henseler's (2007) approach is simple. Each centered bootstrap estimate of the second group is compared with each centered bootstrap of the first group across all the bootstrap samples. The number of positive differences divided by the total number of comparisons (i.e., B^2) indicates the probability that the second group's population parameter will be greater than that of the first group.

Henseler's (2007) approach does not build on any distributional assumptions and is simple to apply by using the bootstrap outputs generated by established PLS path modeling software packages such as SmartPLS (Ringle, Wende, & Will, 2005) and PLS-graph (Chin, 2003a). Researchers can easily make the final calculations with available spreadsheet software applications. However, Henseler's (2007) approach only allows testing the one-sided hypotheses. As the bootstrap-based distribution is not necessarily symmetric, it cannot be used to test two-sided hypotheses.

Nonparametric Confidence Set Approach

As an answer to prior methods' deficiencies, we propose the confidence set approach, which builds conceptually on Keil et al.'s (2000) parametric test. Keil et al.'s (2000) approach is a modified version of the two independent samples *t*-test, which accounts for the fact that the standard deviation is obtained through bootstrapping. As such, the test indirectly compares two bootstrap confidence intervals, assuming that the data are normally distributed.

In accordance with this test, researchers can directly compare the group-specific bootstrap confidence intervals, regardless of whether the data are normally distributed or not. The procedure is as follows:

1. Run the PLS path modeling algorithm separately for each group.
2. Construct the bias-corrected α%-bootstrap confidence intervals (preferably 95% in order to avoid Type-II error inflation) for groups one and two, $(\tilde{\theta}^{(1)}_{low}, \tilde{\theta}^{(1)}_{up})$, and $(\tilde{\theta}^{(2)}_{low}, \tilde{\theta}^{(2)}_{up})$.
3. If the parameter estimate for a path relationship of group one $\tilde{\theta}^{(1)}$ sided falls within the corresponding confidence interval of group two $(\tilde{\theta}^{(2)}_{low}, \tilde{\theta}^{(2)}_{up})$, *or* if the parameter estimate of group two $\tilde{\theta}^{(2)}$ falls within the corresponding confidence interval of group one $(\tilde{\theta}^{(1)}_{low}, \tilde{\theta}^{(1)}_{up})$, it can be assumed that there are *no* significant differences between the group-specific path coefficients with regard to a significance level α. Conversely, if there is no overlap, one can assume that group-specific path coefficients are significantly different.

An important element of the confidence set approach is the bootstrap confidence interval. Several methods for constructing bootstrap confidence intervals have been proposed in the literature (e.g., Davison & Hinkley, 1997; Efron & Tibshirani, 1993). An obvious way to construct a confidence interval for a parameter based on bootstrap estimates is to use a set of B bootstrap samples $x^*_i (i \in \{1, \ldots, B\})$ and calculate the bootstrap-specific parameters $\tilde{\theta}_i^*$. Similar to random subsampling, it is presumed that an interval containing 90% of the $\tilde{\theta}_i^*$ is a 90% confidence interval for θ if the estimates are sorted in ascending sequence. Although this so-called percentile method (Efron, 1981) is appealing due to its easy implementation, prior research has shown that – in the case of small samples (especially regarding asymmetric distributions) – the percentile method does not work well (Chernick, 2008). In addition, this method has a clear tendency to underestimate the upper confidence limit, leading to severe under-coverage (Shi, 1992).

The double bootstrap is an alternative approach which generally provides more accurate bootstrap confidence intervals (i.e., bootstrap the bootstrap; McCullough & Vinod, 1998). Articles on double bootstrap methods appear regularly in the statistical literature (e.g., Davidson & MacKinnon, 2007; McKnight, McKean, & Huitema, 2000), but this technique has not yet found its way into methodological research on PLS path modeling. The double bootstrap's basic principle is to take resamples from each bootstrap resample; that is, for each element of $x_i^* = (x_1^*, x_2^*, \ldots, x_B^*)$ (i.e., the first-level bootstrap), further resamples $x_{ij}^{**} = (x_{11}^{**}, \ldots, x_{1M}^{**}, \ldots, x_{B1}^{**}, \ldots, x_{BM}^{**})(j \in \{1, \ldots, M\})$ are drawn from the second level. Both types of bootstrap samples are used to estimate path coefficients on the two levels; that is, $\tilde{\theta}_i^*$ (first level) and $\tilde{\theta}_{ij}^{**}$ (second level). Fig. 2 illustrates the general concept.

However, this approach is computationally demanding. Specifically, the second-level bootstrap generates M bootstrap samples for each first-level bootstrap, leading to an overall number of $B \cdot M + B$ bootstrap samples. For example, following Hair et al.'s (2011) recommendation to use at least 5,000 bootstrap samples would require drawing more than 25×10^6 bootstrap samples.

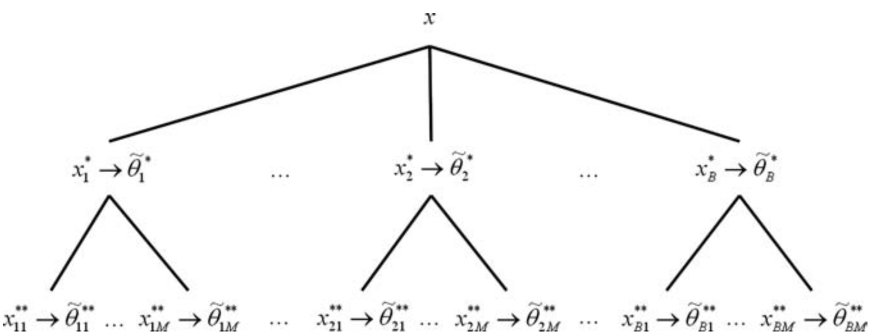

Fig. 2. The Double Bootstrap Method.

Based on this principle, Shi (1992) proposed an accurate and efficient double bootstrap method to estimate bootstrap confidence intervals. In this method, the bootstrap distribution is estimated using

$$Q_i^* = \frac{1}{M}\sum_{j=1}^{M} H(\tilde{\theta}_{ij}^{**} - \tilde{\theta}) \tag{7}$$

where $Q_i^* \in \{0; 1\}$ is random under the empirical distribution $\tilde{F}_i$. Its values are sorted in ascending sequence ($Q_{(1)}^* \leq Q_{(2)}^* \leq \cdots \leq Q_{(B)}^*$) and are used to determine the lower and upper confidence limits:

$$(\tilde{\theta}_{low}^{(g)}, \tilde{\theta}_{up}^{(g)}) = (\tilde{\theta}_{[l]}^{(g)}, \tilde{\theta}_{[u]}^{(g)}) \tag{8}$$

where $[\cdot]$ is a nearest integer function with the arguments given by

$$l = (B+1) \cdot Q_{(\alpha/2)}, \text{ and} \tag{9}$$

$$u = (B+1) \cdot Q_{(1-\alpha/2)} \tag{10}$$

Since estimating the bootstrap confidence interval (Efron & Tibshirani, 1993) entails potential systematic errors, Davison and Hinkley (1997) proposed a bias correction, which should be considered when constructing the interval. The use of bias-corrected confidence intervals was introduced to PLS path modeling in the context of the confirmatory tetrad analysis (Gudergan, Ringle, Wende, & Will, 2008) and bootstrapping-based significance testing (Henseler et al., 2009). The bias correction is as follows:

$$\begin{aligned} bias &= \frac{1}{B}\sum_{i=1}^{B} \tilde{\theta}_i^* - \tilde{\theta} - \left(\frac{1}{BM}\sum_{i=1}^{B}\sum_{j=1}^{M} \tilde{\theta}_{ij}^{**} - \frac{2}{B}\sum_{i=1}^{B} \tilde{\theta}_i^* + \tilde{\theta}\right) \\ &= \frac{3}{B}\sum_{i=1}^{B} \tilde{\theta}_i^* - \frac{1}{BM}\sum_{i=1}^{B}\sum_{j=1}^{M} \tilde{\theta}_{ij}^{**} - 2\tilde{\theta} \end{aligned} \tag{11}$$

This bias correction is used to estimate the confidence interval's lower and upper limits:

$$(\tilde{\theta}_{low}^{(g)}, \tilde{\theta}_{up}^{(g)}) = (\tilde{\theta}_{[l]}^{(g)} - bias, \tilde{\theta}_{[u]}^{(g)} - bias) \tag{12}$$

Although Shi's (1992) method for estimating double bootstrap-based confidence intervals has proven to be accurate in various data constellations, the improvement in accuracy comes at the expense of computational demand.

MULTIGROUP ANALYSIS WITH MORE THAN TWO GROUPS

All previously presented approaches to group comparison in PLS path modeling have in common that they test the difference in the parameters between two groups. As previously mentioned, researchers in international marketing and other cross-cultural research fields frequently encounter situations in which they would like to compare more than two groups. As soon as there are more than two groups, two questions arise: Does a parameter differ between groups? And, if so, between which groups does it differ? Although the second question can be answered by means of pairwise group comparisons, the first question demands more attention. Again, as mentioned, a naive approach would be to conduct all possible pairwise group comparisons, which would lead to the well-known multiple testing problem; that is, the familywise error rate quickly exceeds any prespecified acceptable Type-I error level.

There are, however, several ways of controlling the familywise error rate. A standard remedy is the Bonferroni correction, which aims at retaining the familywise error rate by dividing each comparison's error-rate by the overall number of comparisons. The Bonferroni correction tends to be conservative; that is, it sacrifices statistical power for the sake of a predefined level of Type-I error. An alternative would be to conduct an ANOVA (i.e., an overall *F*-test), comparing the different groups' bootstrap outputs. However, using an ANOVA would mean relying on distributional assumptions (e.g., Hair, Black, Babin, & Anderson, 2010; Mooi & Sarstedt, 2011), which Chin and Dibbern (2010) criticize. An optimal test for the differences between multiple groups in a PLS path modeling framework should (1) maintain the familywise error rate, (2) deliver an acceptable level of statistical power, and (3) not rely on distributional assumptions. Another desirable feature is that such a test should be available in PLS path modeling software packages. In this section, we propose such an omnibus test of group differences (OTG).

Our OTG approach uses bootstrapping, permutation, and random selection's asymptotic properties. The underlying idea of this nonparametric OTG dates back to Pitman (1938) – although the concrete implementation is inspired by Bortz, Lienert, and Boehnke (2003), who proposed a "randomized ANOVA" method. Their method tests the hypothesis that *G* samples are drawn from populations with identical means. Applied to PLS path modeling, the OTG approach consists of the following steps:

1. The first step encompasses groupwise bootstrapping. Per group, a large number of bootstrap samples are drawn and estimated in order to obtain

an empirical distribution of the group-specific model parameters. The number of bootstrap samples should be equal across the groups. The presentation of the bootstrap estimates may be structured as shown in Table 1.

2. The bootstrap results of the previous step facilitate the variance ratio's computation. Analogous to a one-way ANOVA (e.g., Mooi & Sarstedt, 2011), the variance explained by the grouping variable is evaluated relatively to the overall variance:

$$F_R = \frac{s^2_{\text{between}}}{s^2_{\text{within}}} = \frac{G \cdot B \cdot (1/(G-1)) \cdot \sum_{g=1}^{G} (\bar{A}_g - \bar{A})^2}{1/(B-1) \cdot \sum_{g=1}^{G} \sum_{i=1}^{B} (\tilde{\theta}_i^{(g)*} - \bar{A}_g)^2} \tag{13}$$

In this equation, $\tilde{\theta}_i^{(g)*}$ is the parameter estimate from the i^{th} bootstrap sample ($i = 1,\ldots, B$) of group g ($g = 1,\ldots, G$), $\bar{A}_g$ the average over the bootstrap parameter estimates of group g, and $\bar{A}$ the grand mean of all the bootstrap values.

3. This permutation step uses the previously generated bootstrap estimates (e.g., as displayed in Table 1). The elements of the first row – the outcomes of the first bootstrap estimation in each group – can be permuted in $G!$ different ways, whereby each permutation has the same likelihood of occurrence. If this idea is extended to all B rows, this results in $(G!)^B$ permutations. Since the test outcomes are independent of the group index, there are only $(G!)^{B-1}$ different permutations.

   For many bootstrap samples, the associated number of permutations becomes extremely high (e.g., in the case of $B = 5{,}000$ bootstraps and $G = 3$ groups, $(3!)^{4{,}999} = 9.508 \times 10^{3{,}889}$ permutations are required). Such extensive computations are not feasible within a reasonable time.

Table 1. Arranging the Groupwise Bootstrap Estimates of a Specific Model Parameter.

Bootstrap Estimation	Groups			
	1	2	…	G
1	$\tilde{\theta}_1^{(1)}$	$\tilde{\theta}_1^{(2)}$	…	$\tilde{\theta}_1^{(G)}$
2	$\tilde{\theta}_2^{(1)}$	$\tilde{\theta}_2^{(2)}$	…	$\tilde{\theta}_2^{(G)}$
…	…	…	…	…
B	$\tilde{\theta}_B^{(1)}$	$\tilde{\theta}_B^{(2)}$	…	$\tilde{\theta}_B^{(G)}$

Consequently, we draw on the random selection (i.e., Monte Carlo) concept. A reasonably high number of permutations (e.g., 5,000) are sufficient to obtain an outcome that approximates the results for $(G!)^{B-1}$ different permutations. Subsequently, the variance ratio F_R can be computed for each randomly selected permutation (e.g., one obtains 5,000 F_R values).

4. The error probability is computed in the final step. As is usual with regard to randomization tests, one has to examine whether the empirical F_R value from Step 1 is among the α% largest values of the empirical F_R value distribution obtained from the previous step. The error probability p can be determined as follows:

$$p = \frac{1}{U}\sum_{u=1}^{U} H(F_R - F_{R_u}) \tag{14}$$

In this equation, $H(\cdot)$ is again the Heaviside step function, U denotes the number of permutations, and F_{R_u} the empirical F_R-value obtained in permutation run u.

The proposed OTG approach offers a possibility to control the familywise error rate. This approach does not rely on distributional assumptions, nor is it as conservative as the Bonferroni correction. The OTG approach can be applied to the regular bootstrap output of standard PLS path modeling software implementations, such as SmartPLS (Ringle et al., 2005).[2]

EMPIRICAL EXAMPLE

Overview

In this section, we use a well-established PLS path model and empirical data to illustrate and compare the different multigroup analysis approaches. The selected PLS path model draws on prior studies by Homburg and Rudolph (1997), as well as by Festge and Schwaiger (2007), and examines the effects of customer satisfaction drivers on customer loyalty in industrial markets.[3] Since the focus of this section is not centered on the substantive model as such, but on an illustration of the multigroup analysis approaches, we only provide a brief description of the data and model set-up.

Measures and Data

The data originate from a survey – by means of standardized mail questionnaires – of a major industrial firm's customers in three countries (Germany, $n = 65$; the United Kingdom, $n = 115$; and France, $n = 170$). All the respondents rated their satisfaction with the different performance features related to the firm's products and services. Our model includes the following three performance features, which have been shown to significantly affect customer satisfaction in industrial markets (e.g., Festge & Schwaiger, 2007; Sarstedt et al., 2009): (1) satisfaction with products, (2) satisfaction with services, and (3) satisfaction with pricing. The corresponding construct measures were adapted from Homburg and Rudolph (1997), as well as from Festge and Schwaiger (2007), using a five-item scale to measure satisfaction with products, and two three-item scales to measure satisfaction with services and pricing. Loyalty was measured with three well-known items (intention to repurchase, word-of-mouth recommendation, and intention to remain a customer in the long run) from prior research (Zeithaml, Berry, & Parasuraman, 1996). We used reflective indicator variables measured on seven-point Likert-type scales.

Results

Principal components analysis supports the scales' unidimensionality. In addition, we computed coefficient β values which range from 0.59 (satisfaction with services; German subsample) to 0.83 (satisfaction with products; German subsample), and are thus above the commonly suggested threshold of 0.50 (Revelle, 1979). Subsequent PLS path model analyses reveal that all measures meet the commonly suggested criteria for measurement model assessment as described, for example, by Chin (1998), Henseler et al. (2009), and Hair et al. (2012). Specifically, the analyses per country show that all indicators exhibit loadings above 0.70, and that the constructs' average variance extracted (AVE) values are above 0.50. Likewise, all constructs achieve high composite reliability values of 0.80 and higher (Table 2).

We used two approaches to assess the constructs' discriminant validity. First, we examined the indicators' cross loading, which revealed that no indicator loads higher on an opposing construct (Hair et al., 2012). Second, we applied the Fornell and Larcker (1981) criterion and tested whether each construct's AVE is greater than its squared correlation with the remaining constructs. Both analyses clearly indicate that the constructs exhibit

Table 2. Country-Specific Results.

		Germany	United Kingdom	France
Latent variables				
Satisfaction with services	Composite reliability	0.829	0.860	0.848
	AVE	0.619	0.672	0.650
Satisfaction with products	Composite reliability	0.910	0.889	0.895
	AVE	0.669	0.616	0.630
Satisfaction with prices	Composite reliability	0.829	0.833	0.895
	AVE	0.619	0.625	0.623
Loyalty	Composite reliability	0.869	0.836	0.846
	AVE	0.689	0.631	0.646
n		65	115	170
Path relationships				
Satisfaction with services → Loyalty		0.040	0.238***	0.195***
Satisfaction with products → Loyalty		0.669***	0.130*	0.289***
Satisfaction with prices → Loyalty		0.163*	0.500***	0.398***
R^2		0.690	0.600	0.609

Notes: *Significance at 0.10, **significance at 0.05, ***significance at 0.01.

discriminant validity. Overall, these results provide clear support for the measures' reliability and convergent validity.

Table 2 shows the results of the structural model evaluation. The bootstrap analyses using 5,000 samples and a number of cases equal to the country-specific sample size (using the individual sign change option) show that all the satisfaction features – with the exception of satisfaction with services in the German subsample – have a significant ($p \leq 0.10$) effect on customer loyalty. A comparison of the country-specific path coefficients reveals several differences in the effects. For example, whereas satisfaction with products has the strongest effect on loyalty in the German subsample, it has a much weaker effect in the UK subsample. Instead, satisfaction with prices exerts the strongest influence on loyalty in the UK subsample. In respect of the French subsample, the effects are somewhat balanced across the three satisfaction types. However, the question emerges whether these numeric differences between country-specific path coefficients are statistically significant.

In a first step, we applied the OTG approach to assess if the path coefficients are equal across the three groups. The analysis reveals that in respect of all three structural model relations, the null hypothesis that the three path coefficients are equal across the three groups can be rejected.

Specifically, the analysis yields F_R values of 579.93 (Services → Loyalty), 3,393.36 (Products → Loyalty), and 1,504.48 (Prices → Loyalty), rendering all differences significant at $p \leq 0.01$. These results suggest that, in respect of all three relationships, at least one path coefficient differs from the remaining two across the three countries.

Table 3 shows the differences in three comparisons' path coefficient estimates (Germany vs. the United Kingdom, Germany vs. France, and the United Kingdom vs. France), and provides the results of multigroup comparisons based on the parametric approach, the permutation test, and Henseler's (2007) approach. The analysis shows that, generally, the multigroup comparison test results correspond very closely. However, differences emerge in respect of Keil et al.'s (2000) parametric tests, which, in most cases, yields higher t-values than the permutation test. For example, in the comparison of the German and the UK subsamples, Keil et al.'s (2000) test renders the relationship between satisfaction with services and loyalty significant ($p \leq 0.10$), whereas this result does not occur in the permutation test. Consequently, the parametric approach can generally be considered more liberal in terms of rendering a certain difference significant. Conversely, Henseler's (2007) approach appears to be rather conservative in this respect. Although the approach indicates several significant differences, one has to bear in mind that it only allows testing a one-sided hypothesis. Comparing its results with, for example, the critical t-values of a one-sided

Table 3. Multigroup Comparison Test Results.

Relationship	Comparison	$\lvert\text{diff}\rvert$	$t_{\text{Parametric}}$	$t_{\text{Permutation}}$	p_{Henseler}
Services → Loyalty	Germany vs. United Kingdom	0.198	1.930*	1.632	0.095
	Germany vs. France	0.155	1.530	1.351	0.130
	United Kingdom vs. France	0.043	0.410	0.441	0.363
Products → Loyalty	Germany vs. United Kingdom	0.539	4.285***	3.285***	0.005
	Germany vs. France	0.270	2.662***	2.614***	0.013
	United Kingdom vs. France	0.159	1.503	1.367	0.107
Prices → Loyalty	Germany vs. United Kingdom	0.338	2.156**	2.052**	0.021
	Germany vs. France	0.235	1.967**	1.802*	0.063
	United Kingdom vs. France	0.102	0.930	0.959	0.193

Notes: *Significant at 0.10, **significant at 0.05, ***significant at 0.01.
Results for Henseler (2007) eligible for a one-sided test.

parametric test (e.g., 1.28 for $\alpha = 0.10$), clearly shows that Henseler's (2007) approach reveals fewer significant effects.

Table 4 shows the bias-corrected 95% confidence intervals according to Shi's (1992) approach, as well as the corresponding multigroup analysis results. Again, if the parameter estimate for a path relationship of one group (Table 2) does not fall within the corresponding confidence interval of another group (Table 4) and vice versa, there exists no overlap and we can assume that the group-specific path coefficients are significantly different with regard to a significance level α.

Comparing the confidence set approach's results with those of prior tests shows that the former is more conservative than Keil et al.'s (2000) test. Whereas the parametric approach indicates a significant difference ($p \leq 0.05$) between the German and French subsamples in terms of the satisfaction with prices and loyalty relationship, this is not the case with the confidence set approach. Overall, in terms of significant differences, the approach closely resembles the permutation test's results.

Table 4. Bias-corrected 95% Confidence Intervals (Shi 1992) and Multigroup Comparison Results.

Relationship	Confidence Intervals			Comparison	Significance
	Germany	United Kingdom	France		
Services → Loyalty	[−0.206,0.250]	[0.035,0.380]	[0.065,0.325]	Germany vs. United Kingdom	Nsig.
				Germany vs. France	Nsig.
				United Kingdom vs. France	Nsig.
Products → Loyalty	[0.329,0.991]	[−0.021,0.275]	[0.115,0.469]	Germany vs. United Kingdom	Sig.
				Germany vs. France	Sig.
				United Kingdom vs. France	Nsig.
Prices → Loyalty	[−0.158,0.447]	[0.303,0.658]	[0.239,0.551]	Germany vs. United Kingdom	Sig.
				Germany vs. France	Nsig.
				United Kingdom vs. France	Nsig.

Notes: Sig. denotes a significant difference at 0.05; Nsig. denotes a nonsignificant difference at 0.05.

SUMMARY, CONCLUSIONS, AND FUTURE RESEARCH

PLS path modeling is a key multivariate analysis method for empirical research in international marketing (e.g., Henseler et al., 2009), and multigroup analyses are of primary interest in this field (e.g., Hoffmann, Mai, & Smirnova, 2011). This research contributes to the literature on PLS path modeling in several ways: First, we present and compare the procedures available for multigroup analysis in PLS path modeling. Second, we introduce the novel nonparametric confidence set approach based on the comparison of parameter estimates and bootstrap confidence intervals. Third, we address the issue of simultaneously comparing more than two groups by providing a permutation-based analysis of variance approach that maintains the familywise error rate, does not rely on distributional assumptions, and exhibits an acceptable level of statistical power.

The results of the empirical example suggest that Keil et al.'s (2000) parametric approach is the most liberal of the procedures as it, compared to the permutation test, generally yields higher *t*-values. Furthermore, Keil et al.'s (2000) approach renders more differences significant than the novel confidence set approach does. The confidence set approach, just like Henseler's (2007) procedure, appears to be very conservative, as they indicate fewer significant differences vis-à-vis alternative multigroup comparison tests.

International marketing research often deals with relatively small sample sizes and a relatively large number of groups (i.e., data from different cultures or countries). Our novel confidence set approach specifically provides researchers with certain advantageous functions in these kinds of situations. The confidence set approach is nonparametric, can handle relatively small sizes, and is more conservative than the other approaches and, thus, is less prone to Type-II errors. These aspects are particularly relevant when conducting multigroup analysis in international marketing research.

Overall, our findings suggest that if researchers need to compare more than two groups (e.g., countries or cultures), they should first conduct the OTG in order to test the hypothesis that a model parameter differs across groups. If this hypothesis is supported, or if there are only two groups, researchers should subsequently apply the novel confidence set approach to multigroup analysis with regard to comparing two groups of data.

Obviously, our empirical illustration using satisfaction data can only be a first step toward understanding the different multigroup analysis approaches' adequacy. For example, with regard to the confidence set approach's conservative performance, it is unclear if the corresponding path relationships are truly identical in the population, or if the approach's potential lack of statistical power biases the outcome. The approaches may perform differently, depending on the model set-up and sample at hand. It is therefore necessary to compare the approaches' point estimation accuracy, and their statistical power in systematically changed data constellations by conducting a Monte Carlo experiment. In a related context, Qureshi and Compeau (2009) evaluate the ability of variance and covariance-based approaches to structural equation modeling to detect between-group differences and to accurately estimate the moderating effects' strength. However, the authors only consider Keil et al.'s (2000) parametric approach, rather than comparing the performance of different approaches within a PLS path modeling framework.

Another important issue of, and avenue for future research on, multi-group comparisons is exploring ways to test for measurement invariance (e.g., Steenkamp & Baumgartner, 1998; Vandenberg & Lance, 2000) in a PLS path modeling context. If measurement invariance cannot be established, the differences in path coefficients cannot be fully attributed to true relationships, because respondents from different groups might have systematically interpreted a given measure in conceptually different ways. Although measurement invariance should be added to the well-established criteria reliability, homogeneity, and validity when performing multigroup analysis, prior research on PLS path modeling has largely neglected this issue. Haenlein and Kaplan (2011) proposed an approach to control for gamma change, which occurs when the construct's domain (i.e., its meaning) differs in each group. Specifically, the authors propose a combination of Box's M test and ordinary least squares regressions, which can help assess this bias's magnitude and, hence, support researchers when they have to decide whether parameter estimates can be trusted or not. However, Rigdon et al. (2010, p. 269) provide a different perspective on measurement invariance in PLS path modeling, stating that "an insistence on measurement invariance across groups carries its own assumption that the impact of group membership is limited to the structural parameters of the structural model. In many cases, this assumption is questionable or even implausible, and researchers should consider group membership effects on both structural and measurement parameters." Furthermore, the authors point out that PLS path modeling is a method based on approximation and designed for situations with a less firmly established theoretical base (Wold, 1982). Therefore, researchers should interpret the results from PLS path modeling involving multiple groups with the necessary caution.

NOTES

1. Sarstedt and Wilczynski (2009) describe a complementary approach for paired samples.
2. A code file for R (R-Development-Core-Team, 2011), which performs the approach, can be obtained from the second author upon request.
3. Sarstedt et al. (2009) applied the FIMIX-PLS method (Hahn et al., 2002; Rigdon et al., 2010; Ringle, Wende, & Will, 2010; Ringle, Sarstedt, & Mooi, 2010; Sarstedt et al., 2011; Sarstedt & Ringle, 2010) to the original study by Festge and Schwaiger (2007) to uncover unobserved heterogeneity.

REFERENCES

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182.

Bortz, J., Lienert, G. A., & Boehnke, K. (2003). *Verteilungsfreie Methoden in der Biostatistik* (3 ed.). Berlin: Springer.

Brettel, M., Engelen, A., Heinemann, F., & Vadhanasindhu, P. (2008). Antecedents of market orientation: A cross-cultural comparison. *Journal of International Marketing*, *16*(2), 84–119.

Chambers, J. M., Cleveland, W. S., Kleiner, B., & Tukey, P. A. (1983). *Graphical methods for data analysis*. Belmont: Wadsworth.

Chernick, M. R. (2008). *Bootstrap methods. A guide for practitioners and researchers* (Wiley Series in Probability and Statistics, 2: Wiley).

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: Erlbaum.

Chin, W. W. (2000). Multi-group analysis with PLS. Retrieved from http://disc-nt.cba.uh.edu/chin/plsfaq/multigroup.htm.

Chin, W. W. (2003a). *PLS graph 3.0*. Houston: Soft Modeling Inc.

Chin, W. W. (2003b). A permutation procedure for multi-group comparison of PLS models. In: M. Vilares, M. Tenenhaus, P. Coelho, V. Esposito Vinzi, A. Morineau (Eds.), *PLS and Related Methods: Proceedings of the International Symposium PLS'03* (pp. 33–43). Lisbon: Decisia.

Chin, W. W., & Dibbern, J. (2010). An introduction to a permutation based procedure for multi-group PLS analysis: Results of tests of differences on simulated data and a cross cultural analysis of the sourcing of information system services between Germany and the USA. In: V. Esposito Vinzi, W. W. Chin, J. Henseler & H. Wang, et al. (Eds.), *Handbook of partial least squares. Concepts, methods and applications* (pp. 171–193). Berlin, Heidelberg: Springer-Verlag.

Davidson, R., & MacKinnon, J. G. (2007). Improving the reliability of bootstrap tests with the fast double bootstrap. *Computational Statistics & Data Analysis*, *51*(7), 3259–3281.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge: Cambridge Univiversity Press.

Dibbern, J., & Chin, W. W. (2005). Multi-group comparison: Testing a PLS model on the sourcing of application software services across Germany and the USA using a permutation based algorithm. In: F. Bliemel, A. Eggert, G. Fassott & J. Henseler, et al. (Eds.), *Handbuch PLS-Pfadmodellierung. Methode, Anwendung, Praxisbeispiele* (pp. 135–160). Stuttgart: Schäffer-Poeschel.

Edgington, E., & Onghena, P. (2007). *Randomization tests* (4th ed.). London: Chapman & Hall.

Efron, B. (1981). Nonparametric standard erroros and confidence intervals. *Canadian Journal of Statistics*, *9*(2), 139–172.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.

Esposito Vinzi, V., Ringle, C. M., Squillacciotti, S., & Trinchera, L. (2007). Capturing and treating unobserved heterogeneity by response based segmentation in PLS path modeling: A comparison of alternative methods by computational experiments. Cergy Pontoise Cedex: ESSEC Research Center, Working Paper No. 07019.

Esposito Vinzi, V., Trinchera, L., Squillacciotti, S., & Tenenhaus, M. (2008). REBUS-PLS: A response-based procedure for detecting unit segments in PLS path modelling. *Applied Stochastic Models in Business and Industry*, *24*(5), 439–458.

Festge, F., & Schwaiger, M. (2007). The drivers of customer satisfaction with industrial goods: An international study. *Advances in International Marketing*, *18*, 179–207.

Fornell, C. G., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39–50.

Graham, J. L., Mintu, A. T., & Rodgers, W. (1994). Explorations of negotiation behaviors in ten foreign cultures using a model developed in the United States. *Management Science*, *40*(1), 72–95.

Grewal, R., Chakravarty, A., Ding, M., & Liechty, J. (2008). Counting chickens before the eggs hatch: Associating new product development portfolios with shareholder expectations in the pharmaceutical sector. *International Journal of Research in Marketing*, *25*(3), 261–272.

Gudergan, S. P., Ringle, C. M., Wende, S., & Will, A. (2008). Confirmatory tetrad analysis in PLS path modeling. *Journal of Business Research*, *61*(12), 1238–1249.

Haenlein, M., & Kaplan, A. M. (2011). The influence of observed heterogeneity on path coefficient significance: Technology acceptance within the marketing discipline. *Journal of Marketing Theory and Practice*, *19*(2), 153–168.

Hahn, C., Johnson, M. D., Herrmann, A., & Huber, F. (2002). Capturing customer heterogeneity using a finite mixture PLS Approach. *Schmalenbach Business Review*, *54*(3), 243–269.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7 ed.). Englewood Cliffs: Prentice Hall.

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory and Practice*, *19*(2), 139–151.

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*, forthcoming (online available).

Henseler, J. (2007). A new and simple approach to multi-group analysis in partial least squares path modeling. In: H. Martens & T. Næs (Eds.), *Causalities explored by indirect observation: Proceedings of the 5th international symposium on PLS and related methods* (PLS'07) (pp. 104–107). Oslo.

Henseler, J., & Chin, W. W. (2010). A comparison of approaches for the analysis of interaction effects between latent variables using partial least squares path modeling. *Structural Equation Modeling*, *17*(1), 82–109.

Henseler, J., & Fassott, G. (2010). Testing moderating effects in PLS path models: An illustration of available procedures. In: V. Esposito Vinzi, W. W. Chin, J. Henseler & H. Wang (Eds.), *Handbook of partial least squares. Concepts, methods and applications* (pp. 713–735). Berlin, Heidelberg: Springer-Verlag.

Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing. *Advances in International Marketing*, *20*, 277–320.

Hoffmann, S., Mai, R., & Smirnova, M. (2011). Development and validation of a cross-nationally stable scale of consumer animosity. *Journal of Marketing Theory and Practice*, *19*(2), 235–251.

Homburg, C., & Rudolph, B. (1997). Customer satisfaction in industrial markets: Dimensional and multiple role issues. *Journal of Business Research*, *52*(1), 15–33.

Jedidi, K., Jagpal, H. S., & DeSarbo, W. S. (1997). Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science*, *16*(1), 39–59.

Keil, M., Saarinen, T., Tan, B. C. Y., Tuunainen, V., Wassenaar, A., & Wei, K.-K. (2000). A cross-cultural study on escalation of commitment behavior in software projects. *Management Information Systems Quarterly*, *24*(2), 299–325.

Lohmöller, J.-B. (1989). *Latent variable path modeling with partial least squares*. Heidelberg: Physica.

McCullough, B. D., & Vinod, H. D. (1998). Implementing the double bootstrap. *Computational Economics*, *12*(1), 79–95.

McKnight, S. D., McKean, J. W., & Huitema, B. E. (2000). A double bootstrap method to analyze linear models with autoregressive error terms. *Psychological Methods*, *5*(1), 87–101.

Mooi, E. A., & Sarstedt, M. (2011). *A concise guide to market research: The process, data, and methods using IBM SPSS statistics*. Berlin: Springer.

Nitzl, C. (2010). Eine anwenderorientierte Einführung in die Partial Least Square (PLS)-Methode. Hamburg: Universität Hamburg: Industrielles Management, Arbeitspapier Nr. 21.

Pitman, E. J. G. (1938). Significance tests which may be applied to samples from any population. *Journal of the Royal Statistical Society Supplement*, *4*(1), 119–130.

Qureshi, I., & Compeau, D. R. (2009). Assessing between-group differences in information systems research: A comparision of covariance- and component-based SEM. *Management Information Systems Quarterly*, *33*(1), 197–214.

R-Development-Core-Team. (2011). R: A language and environment for statistical computing. Vienna.

Revelle, W. (1979). Hierarchical clustering and the internal structure of tests. *Multivariate Behavioral Research*, *14*(1), 57–74.

Rigdon, E. E., Ringle, C. M., & Sarstedt, M. (2010). Structural modeling of heterogeneous data with partial least squares. In: N. K. Malhotra (Ed.), *Review of marketing research* (Vol. 7, pp. 255–296). Armonk: Sharpe.

Ringle, C. M., Sarstedt, M., & Mooi, E. A. (2010). Response-based segmentation using finite mixture partial least squares: Theoretical foundations and an application to American Customer Satisfaction Index data. *Annals of Information Systems*, *8*, 19–49.

Ringle, C. M., Sarstedt, M., & Schlittgen, R. (2010). Finite mixture and genetic algorithm segmentation in partial least squares path modeling: Identification of multiple segments in a complex path model. In: A. Fink, B. Lausen, W. Seidel & A. Ultsch (Eds.), *Advances in data analysis, data handling and business intelligence* (pp. 167–176). Berlin: Springer.

Ringle, C. M., Wende, S., & Will, A. (2005). SmartPLS 2.0 (Beta). Hamburg: SmartPLS. Retrieved from www.smartpls.de.

Ringle, C. M., Wende, S., & Will, A. (2010). Finite mixture partial least squares analysis: Methodology and numerical examples. In: V. Esposito Vinzi, W. W. Chin, J. Henseler & H. Wang (Eds.), *Handbook of partial least squares. Concepts, methods and applications* (pp. 195–218). Berlin, Heidelberg: Springer-Verlag.

Rodríguez, C. M., & Wilson, D. T. (2002). Relationship bonding and trust as a foundation for commitment is U.S.-Mexican strategic alliances: A structural equation modeling approach. *Journal of International Marketing*, *10*(4), 53–76.

Sarstedt, M. (2008). A review of recent approaches for capturing heterogeneity in partial least squares path modelling. *Journal of Modelling in Management*, *3*(2), 140–161.

Sarstedt, M., Becker, J.-M., Ringle, C. M., & Schwaiger, M. (2011). Uncovering and treating unobserved heterogeneity with FIMIX-PLS: Which model selection criterion provides an appropriate number of segments? *Schmalenbach Business Review*, *63*(1), 34–62.

Sarstedt, M., & Ringle, C. M. (2010). Treating unobserved heterogeneity in PLS path modelling: A comparison of FIMIX-PLS with different data analysis strategies. *Journal of Applied Statistics*, *37*(8), 1299–1318.

Sarstedt, M., Schwaiger, M., & Ringle, C. M. (2009). Do we fully understand the critical success factors of customer satisfaction with industrial goods? – Extending Festge and Schwaiger's model to account for unobserved heterogeneity. *Journal of Business Market Management*, *3*(3), 185–206.

Sarstedt, M., & Wilczynski, P. (2009). More for less? A comparison of single-item and multi-item measures. *Die Betriebswirtschaft*, *69*(2), 211–227.

Shi, S. G. (1992). Accurate and efficient double-bootstrap confidence limit method. *Computational Statistics and Data Analaysis*, *13*(1), 21–32.

Steenkamp, J. B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross national consumer research. *Journal of Consumer Research*, *25*(1), 78–107.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4–70.

Wold, H. (1975). Path models with latent variables: The NIPALS approach. In: H. M. Blalock, A. Aganbegian, F. M. Borodkin, R. Boudon & V. Capecchi (Eds.), *Quantitative sociology: International perspectives on mathematical and statistical modeling* (pp. 307–357). New York: Academic Press.

Wold, H. (1982). Soft modeling: The basic design and some extensions. In: K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observations: Part II* (pp. 1–54). Amsterdam: North-Holland.

Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1996). The behavioral consequences of service quality. *Journal of Marketing*, *60*(2), 31–46.

INTRODUCTION TO SECTION II: REGULAR ARTICLES

Many thanks to Professors Marko Sarstedt, Manfred Schwaiger, and Charles R. Taylor, Volume 22 has assembled a set of outstanding articles addressing the methodological issues in international marketing research. Readers should find these articles informative and valuable. In addition to these articles on the special topic of international marketing research methods, a regular article is included in Volume 22. *Advances in International Marketing* encourages innovative research and "out-of-the-box" research ideas in international marketing. In future volumes, it will continue to promote special topic-based volumes, while also publishing "regular" papers that are reviewed outside of the themed volumes. The regular papers must show innovative research that addresses any significant issues in international marketing and should be submitted to the Series Editor.

Consumer complicity has been a taxing issue faced by many multinational corporations (MNCs). Complicity consumers are those who purchase counterfeit products. We often hear the talk of consumers in a particular country purchasing counterfeit products, such as in China and India recently. But little is known as to why consumers become complicit and whether Chinese and Indian consumers are particularly complicit. Academic research is limited on consumer complicity in foreign markets and almost absent in comparing consumer complicity across the emerging markets, which are often accused of having a major problem of complicit consumers. In the regular article in this volume, Ronald Paul Hill, Goksel Yalcinkaya, Peggy E. Chaudhry, and Stephen A. Stumpf have presented an intriguing study on consumer complicity in BRIC countries, the four major emerging markets of Brazil, Russia, India, and China. On the basis of a web survey of 1,600 consumers in the BRIC countries, these authors have found that the widely held belief that consumer complicity only happens in a particular country like China is false, because the country variable and demographic variables do not explain consumer complicity. Instead, the author found that consumer complicity happens in all four BRIC countries and that the best predictors of consumer complicity are perceived quality, price, and the hedonic shopping experience. The authors also found that the expected utility theory and the cognitive dissonance theory should be combined to examine the consumer complicity issue. In addition to their substantive contributions, the authors also used multiple-group confirmatory factor analysis to assess the cross-national measurement invariance of their measures from four BRIC countries, demonstrating how to implement the cross-national measurement invariance test that is one of the methodological issues in international marketing research addressed by the authors in this Volume 22.

I hope that readers will find this regular article interesting and that they will be inspired by the authors to adopt the state-of-the-art methodological tools in future cross-national research. I certainly hope that readers will submit their similarly well-done research to *Advances in International Marketing*.

Shaoming Zou
Editor

CONSUMER COMPLICITY ACROSS EMERGING MARKETS

Peggy E. Chaudhry, Ronald Paul Hill,
Stephen A. Stumpf and Goksel Yalcinkaya

ABSTRACT

The purpose of this investigation is to examine the explanatory powers of a consumer complicity framework that uses counterfeit products and five emerging country markets (Brazil, Russia, India, and China). A web survey was administered to 1,600 consumers in Brazil, Russia, India, and China to test whether demographics, national origin, perceived quality, price, and a hedonic shopping environment predicted consumers' complicity in these emerging markets. Overall, the results found little support for either demographics or national origin to predict this type of illicit consumption. The best predictive variables were perceived quality, price, and hedonic shopping experience. The study concludes with a model that incorporates these results and suggests that future research employ demarketing tactics using both cognitive dissonance and expected utility theories to obtain a more holistic view for curbing complicity that goes beyond product attributes and the shopping environment.

Keywords: Consumer complicity; counterfeit products; demarketing; emerging markets

Measurement and Research Methods in International Marketing
Advances in International Marketing, Volume 22, 223–239

ISSN: 1474-7979/doi:10.1108/S1474-7979(2011)0000022014

INTRODUCTION

The International Chamber of Commerce (ICC) claims that between 5% and 7% of world trade, an estimated $600 billion annually, will be lost this year to counterfeit trade. Furthermore, the Counterfeiting Intelligence Bureau of the ICC has investigated more than 600 incidents of piracy in 35 countries ranging from fake pharmaceuticals to alcoholic beverages (Counterfeiting Intelligence Bureau, 2011). Consequences of these marketplace distortions are severe. Damage to brand equity for licit products has caused significant drops in sales, leading to concomitant job losses, and Wilcox, Kim, and Sen (2009) lament these knockoffs are responsible, in part, for societal ills such as trafficking of narcotics and weapons.

The World Health Organization estimates that in over 50% of cases, pharmaceuticals obtained on the web from illegal sites that disguise physical addresses have been discovered to be counterfeit (*Medicine: Counterfeit Medicine*, 2010). The Center for Medicine in the Public Interest predicted that the global growth in counterfeit drugs at 13% annually (twice the growth rate for pharmaceuticals) – representing a $75 billion illicit pharmaceutical industry by 2010 (*Counterfeit Drugs and China*, 2011). *The Economist* warns that fake pharmaceuticals are encroaching on the developed world and reports that Pfizer discovered that 20% of European consumers in 14 countries admitted to obtaining their medicines by way of illegal channels (Poison Pills, 2010).

A general definition of consumer complicity is: *A complicit consumer is one who intentionally obtains, uses, and/or shares an illicit product*. The underlying consumer rationale for obtaining counterfeits is the ability to extend purchasing power by paying a fraction of what genuine products cost, while receiving full benefits of branded goods (Commuri, 2009). Scholars suggest that a number of variables underlie motivation to consume, from attitudes and beliefs (Cordell, Kieschnick, & Wongtada, 1996), to moral reasoning (Nill & Shultz, 1996), to corporate image and product characteristics (Penz & Stöttinger, 2008), to social goals (Wilcox et al., 2009). Consequently, this research stream has received criticism that these results provide an incomplete roadmap for understanding and influencing complicity. Most findings are based on studies conducted in North American or Asian countries (Eisend & Schuchert-Güler, 2006) – often with students (Wilcox et al., 2009). Such research focuses on illicit acquisition with a limited array of products, including high-status fashion items (Gucci bags, Rolex watches) or downloaded digital material (music, movies). Few use multiple counterfeits, leading scholars to suggest that complicity is driven by specific product attributes and shopping environments, reducing generalizability across nations (Eisend & Schuchert-Güler, 2006). Thus, researchers now advocate use of a range of consumer subpopulations (Wilcox et al., 2009) across national and regional boundaries that represent illegal markets for counterfeits with different benefits and costs (Commuri, 2009).

Another obstacle to the advancement of our understanding of consumer complicity is the lack of theory that adequately explains the consumer processes involved in market participation. After completing their review of available research, Eisend and Schuchert-Güler (2006) advance the reduction of cognitive dissonance as driving post-complicity justifications for consumers' past behaviors, leading to repeated future complicity. Based on the original work of Festinger (1957), Eisend and Schuchert-Güler suggest that cognitive dissonance occurs as a coping mechanism to bring present attitudes in line with past behaviors. Accordingly, complicit consumers develop rationales that reduce negative beliefs associated with past complicity to justify their ongoing counterfeit buying behaviors – actions they know harm genuine producers and consumers as well as society as a whole and are illegal in most countries. Cognitive dissonance rationales include minimizing importance of the complicit decision, reinterpreting inferior product attributes, or justifying their social conscience or guilt in ways that reduce this dissonance. To counter these justifications, marketers and governments are directed to give dissonance *increasing* information pre-, during, and post-purchase.

Nonetheless, the Von Neumann-Morgenstern (1944) utility theory may support a more complete understanding of consumer complicity (Peace, Galletta, & Thong, 2003). Utility theory focuses on expected personal net gain of an action relative to an alternative action in situations of perceived risk (Hauser & Urban, 1979) and/or uncertainty (Currim & Sarin, 1983). Consumers may be complicit when they expect to benefit more from complicity than from taking alternative actions (e.g., going without and obtaining the product legally). Introducing reference prices (Putler, 1992), altering product quality perceptions (Mehta, Chen, & Narasimhan, 2008), and stimulating equity desires (Ding, 2007; O'Shaughnessy, 2005) are ways that utility theory has been applied to predict consumer behavior. This use of utility theory has revealed a complex mix of wants and motivations that influence buyer decisions (including complicit consumers) related to branded items compared with counterfeits (Desarbo, Kim, Choi, & Spaulding, 2002).

The purpose of this research is to advance international marketing theory and practice using expected utility theory and cognitive dissonance theory to explain and influence consumer complicity across important emerging markets (Brazil, Russia, India, and China), and to consider possible efficacy of anti-counterfeiting tactics on complicit consumers in these markets. The next section provides the framework that guides the empirical research along with methodological considerations, followed by relevant findings and proposal of a framework that both explains consumer complicity and guides actions to reduce it. The close gives insights for marketing scholars as well as proactive strategies for practitioners to successfully combat consumer complicity with relevant products across national markets.

CONCEPTUAL FRAMEWORK

Based on seminal attitude-intention-behavior work by Fishbein and Ajzen (1975), marketers have investigated effects of consumer attitudes toward piracy on intentions and decisions to acquire illicit products (Chiou, Huang, & Lee, 2005; Kwong, Yau, Lee, Sin, & Tse, 2003; Penz & Stöttinger, 2004; Wang, Zhang, Zang, & Ouyang, 2005; Wee, Tan, & Cheok, 1995). In this study, we investigate demographics (e.g., age, gender, education, and income), product attributes (price and quality), national origin (Brazil, Russia, India, and China), and purchase situation/mood as antecedents to consumers' attitude toward counterfeits and decision/intention to obtain that shapes consumer complicity with counterfeits.

Linking Demographics and National Origin to Consumer Complicity

Demographic factors of age, gender, education, and income are advanced as predictors of complicity. For instance, research by Wee et al. (1995) with Singaporeans shows educational attainment and income predicted complicity. Also, Prendergast, Chuen, and Phau (2002) studied students as well as blue-collar workers in Hong Kong, and they profiled low- and high-spending complicit consumers based on age and income. Bian and Veloutsou (2007) found support for a gendered profile (male) for British consumers, but not for Chinese consumers. In a later study, Bian and Moutinho (2009) did not find that age, income, or educational attainment predicted complicity among the British. Such a mixture of findings from a variety of studies reveal the inability to replicate demographic profiles of complicit consumers and further illustrates the pervasiveness of this problem without providing a clear path to eradication of consumer complicity. As a consequence, we assume that such characteristics have no influence on complicity among consumers in emerging markets:

H1. Demographics will have no impact on consumer complicity.

In their study of the predictors of software piracy in 20 Latin American nations, Robertson, Gilley, Crettenden, and Crittenden (2008) note that there have been few studies that address piracy on a cross-national basis. Penz and Stöttinger (2008) in research on the role of corporate image and product attributes to propel consumer demand for counterfeits in Austria, Mexico, Sweden, Slovakia, Ukraine, and the United States assert that cross-country research is significantly underdeveloped. Penz and Stöttinger (2008) found that their conceptual model was relevant in all of the countries studied, but that "no consistent cross-cultural patterns predicting the intention to purchase counterfeit goods emerged" (p. 369). Thus, we predict that national origin of consumers in emerging markets also does not influence complicity:

H2. National origin of consumers will have no impact on consumer complicity.

Linking Product Attributes to Consumer Complicity

Various product attributes likewise may affect purchase of counterfeit goods. Cordell et al. (1996) found that both price and performance expectations toward illicit goods are important determinants of counterfeit purchase intentions. Penz and Stöttinger (2008) later established in a multinational study (the United States, Austria, Mexico, Sweden, Slovakia, and Ukraine) that consumers believed quality was similar between fake and genuine goods. Although consumers may face some confusion due to a quality continuum ranging from shoddy imitations to high-quality products pirated from production overruns, consumers may use product cues to differentiate genuine items from knock-offs and seek counterfeit goods intentionally to receive the desired price/quality trade-off (Chaudhry & Zimmerman, 2009). Thus, a consumer's readiness to buy counterfeit products *increases* if they discern high quality prior to purchase and *decreases* if they cannot make this judgment (Gentry, Putrevu, & Shultz, 2006). Thus, we predict that consumers in emerging markets will respond similarly:

H3. Perceived quality of counterfeit goods is positively related to consumer complicity.

Of course, support for the relationship between lower price and consumer demand has received much support in the literature. For example, the Anti-Counterfeiting Group in the United Kingdom interviewed nearly 1,000 consumers and found that about one-third of these respondents would knowingly purchase counterfeit products if offered at the right (i.e., lower) price relative to the authentic goods (ACG, 2004). An Intellectual Property Theft and Organized Crime study also determined that British consumers would purchase counterfeits if they were significantly cheaper, and this motivation to buy was particularly true for illicit DVDs, music, computer games, business software, and fashion apparel (KPMG, 2006). Thus, we predict that consumers in emerging markets will follow a similar pattern:

H4. Price of counterfeit goods is positively related to consumer complicity.

Linking Hedonic Shopping Experience to Consumer Complicity

Consumers who experience illicit goods purchases as uplifting or mood-enhancing have been observed to be more complicit, particularly when the counterfeit good was acquired in a stimulating shopping environment (Chaudhry & Stumpf, 2010; Leisen & Nill, 2001; Nia & Zaichkowsky, 2000). The original concept of hedonic consumption was advanced by Holbrook and Hirschman (1982), who suggest that a positive affective reaction to the process of buying may influence consumer attitudes toward brands as well as purchase intentions (Voss, Spangenberg, & Grohmann, 2003). Positive moods may lead to ‘‘increased arousal, perceived freedom, fantasy fulfillment, and escapism’’ (Babin, Darden, & Griffin, 1994, p. 646), and this effect may exist across cultures (Schwartz, 1999). If counterfeit purchase situations are perceived as ‘‘fun’’ for consumers, then shopping excursions will likely result in the acquisition of fake products. We concur with recent complicity research that mood is an antecedent variable that influences the decision to purchase and can moderate a consumer’s attitude toward pirated products (Eisend & Schuchert-Güler, 2006; Gentry et al., 2006; Penz & Stöttinger, 2008). Thus, we predict that the impact of hedonic shopping will operate in a similar way for consumers in emerging markets:

H5. Hedonic shopping experience is positively related to consumer complicity.

METHODOLOGY

A web survey was developed to examine the predictive value of the antecedents to consumer complicity suggested in the previous section. Qualtronics was employed to obtain 1,600 voluntary responses equally spread among prescreened persons living in Brazil, China, India, and Russia. The survey was translated into the local languages and refined through back-translation. A soft launch of the survey was performed as a pre-test to verify the consistency of items and scales across respondents and countries.

Through regression analysis, we empirically investigate the extent to which emerging market consumers in Brazil, Russia, India, and China are complicit. We selected these four countries because of large populations, distinct national cultures (Hofstede, 1980), and socio-economic conditions. Implicit in this selection is a desire to explore whether some countries provide a more supportive context for consumer complicity than others among emerging markets or whether they follow the same patterns as developed economies. To broaden the usefulness of our investigation, two dissimilar products were selected – movies and pharmaceuticals. Movies represent the category of digital counterfeits in both virtual and physical markets, are consumed across demographic profiles, and are used for entertainment purposes. Pharmaceuticals are also available in both virtual and physical markets, but consumption includes potentially severe downside risks to physical health and well-being.

ANALYSIS AND RESULTS

Measurement Validation

With the exception of product price, all measures were multiple-item scales based on constructs utilized in previously published research (Babin et al., 1994; Treise, Weigold, Conna, & Garrison, 1994; Wang et al., 2005). Because we collected data from four different countries, it was necessary to assess measurement invariance across samples to make valid cross-national comparisons. Based on multigroup measurement invariance comparison procedures proposed by Steenkamp and Baumgartner (1998), the measurement instruments for each construct were examined, and cross-cultural invariance was assessed using EQS 6.2 with maximum likelihood as the estimation method. By examining sequential models, we evaluated whether imposing additional equality parameter constraints across the countries resulted in substantially inferior models.

Table 1 reports the results of the measurement models. The baseline model (Model 1) assesses whether imposing the same factor structure across all countries yields acceptable results. The overall fit of the baseline model was adequate as indices (i.e., IFI, CFI, TLI, and RMSEA) passed the generally accepted cutoff value of 0.9 (χ^2 (587) = 1267.43, IFI = 0.94; CFI = 0.94; TLI = 0.92, and RMSEA = 0.06). Next, we estimated a model (Model 2) that constrained the factor loadings to be invariant across groups. The fit indices for this model were acceptable (χ^2 (771) = 1282.48, IFI = 0.94; CFI = 0.94; TLI = 0.92, and RMSEA = 0.06), and the difference in fit between this model and the baseline measurement model was not significant (χ^2diff = 15.74, d.f.diff = 16, $p > 0.10$). Thus, the loadings for all measurement items were invariant across nations. The third model (Model 3) imposes an additional constraint on Model 2 requiring that all (co)variances between constructs be equal across the nations. This model also results in acceptable fit indices (χ^2 (646) = 1297.36, IFI = 0.93; CFI = 0.94; TLI = 0.92, and RMSEA = 0.06), and the difference in fit between this model and the baseline measurement model was not significant (χ^2diff = 32.29, d.f.diff = 59, $p > 0.10$), suggesting correlational (covariance) relations among model constructs, as well as construct (factor) variances, were equal across nations.

Finally, we estimated a fully invariant measurement model (Model 4). This model imposes an additional constraint on Model 3 that the error variances be equal across all samples. Although the indices presented adequate fit (χ^2 (681) = 1418.94, IFI = 0.92; CFI = 0.93; TLI = 0.91, and RMSEA = 0.07), this model's fit differed from the baseline measurement model (χ^2diff = 178.52, d.f.diff = 89, $p < 0.01$). However, prior research suggested that "full measurement invariance is particularly unlikely" (Steenkamp & Baumgartner, 1998, p. 81) and partial measurement invariance can be regarded as sufficient. Therefore, we relaxed constraints

Table 1. Assessment of Measurement Invariance.

Model	Models Compared	χ^2 (df)	$\Delta\chi^2(\Delta df)$	RMSEA	IFI	CFI	TLI
Model 1		1267.43 (587)		0.06	0.94	0.94	0.92
Model 2	M2 vs. M1	1282.48 (607)	15.74 (16)	0.06	0.94	0.94	0.92
Model 3	M3 vs. M1	1297.36 (646)	32.29 (59)	0.06	0.93	0.94	0.92
Model 4	M4 vs. M1	1418.94 (681)	178.52 (89)	0.07	0.92	0.93	0.91

on parameters to test for partial measurement invariance. The results indicated that the partial measurement invariance was supported. As a result, we believe that our measures show a high level of cross-national equivalence and, therefore, are appropriate to contrast the relationships among these constructs across all nations.

Next, following the work of Anderson and Gerbing (1988), we first tested the validity of the measures using confirmatory factor analysis. We followed the procedure recommended by Bagozzi and Yi (1988) to evaluate the fit of the measurement model. Overall fit was satisfactory, and all relevant loadings were substantial and highly significant (Table 2). Findings indicate that the model converged well. Discriminant validity of measures was tested by performing, one at a time, χ^2 difference tests between a model in which a factor (construct) correlation is fixed at 1.0 and the original (unrestricted) model (Anderson & Gerbing, 1988). Because every restricted model exhibited significantly worse fit than the unrestricted model, we concluded that there is sufficient degree of discriminant validity between factors. Moreover, construct reliability values exceeded the recommended threshold of 0.60 (Bagozzi & Yi, 1988). Taking all together, the analyses showed that validity of measures was established.

Table 3 summarizes outcomes of separate regression analyses for each country to test our hypotheses. Results indicate that demographic variables (H1) accounted for near zero variance in mediators (attitude toward counterfeiting, decision/intention to obtain). Although studies have found that different demographic combinations may help explain consumer complicity, none has been replicated (Bian & Veloutsou, 2007; Prendergast et al., 2002; Wee et al., 1995), and others have observed little or no relationship between demographic variables and complicity (Bloch, Bush, & Campbell, 1993). Results also showed Indian and Chinese consumers expressed less positive attitudes toward counterfeiting than Brazilian and Russian counterparts, suggesting limited support for (H2). In contrast, product quality (H3) and product price (H4) accounted for meaningful and significant variance in attitude toward counterfeiting and intentions to purchase, and consumers' judgments are likely to be related to usage of these counterfeit goods.

The relationship between shopping experience and attitude toward counterfeiting (H5) is positive and significant for Brazilian and Russian consumers; however, the same relationship is insignificant for Chinese and Indian consumers. However, the relationship between the shopping experience and intention to obtain is positive and significant for Chinese

Table 2. Construct Measurement.

Measure	Brazil[a]		China[b]		India[c]		Russia[d]	
	Standard Loading	*t*-Value	Standard Loading	*t*-Value	Standard Loading	*t*-Value	Standard Loading	*t*-Value
Perceived product quality ($\alpha_a = 83$, $\alpha_b = 81$, $\alpha_c = 87$, $\alpha_d = 84$)								
Counterfeit products have a similar quality to the legal version.	0.87	11.39	0.82	10.78	0.84	11.02	0.89	11.64
Counterfeit products are as reliable as the legal version.	0.78	10.24	0.80	10.41	0.81	10.37	0.77	9.87
Product price								
How important is price as a cue to inform you that a product is counterfeit?	0.79	10.29	0.76	9.64	0.91	11.83	0.80	10.40
Shopping experience ($\alpha_a = 87$, $\alpha_b = 91$, $\alpha_c = 92$, $\alpha_d = 86$)								
I would shop for counterfeits at a market, not because I had to, but because I wanted to.	0.80	10.29	0.82	10.51	0.87	11.48	0.81	10.54
Compared to other things I could do, the time spent shopping for a counterfeit product would be enjoyable.	0.79	10.25	0.75	9.73	0.82	10.74	0.80	10.40
I would have a good time shopping for counterfeit goods because I would be able to act freely.	0.67	7.80	0.70	8.62	0.61	6.96	0.69	8.12
While shopping for counterfeits, I would feel a sense of adventure.	0.90	11.89	0.77	9.82	0.90	11.91	0.91	12.03
I would not worry about legal prosecution since I would use cash to pay for the counterfeit goods.								

Table 2. (*Continued*)

Measure	Brazil[a]		China[b]		India[c]		Russia[d]	
	Standard Loading	*t*-Value	Standard Loading	*t*-Value	Standard Loading	*t*-Value	Standard Loading	*t*-Value
Attitudes towards counterfeiting ($\alpha_a = 77$, $\alpha_b = 82$, $\alpha_c = 85$, $\alpha_d = 78$)								
Products counterfeiting infringes on intellectual property rights.	0.78	10.23	0.75	9.78	0.78	10.19	0.80	10.42
Products counterfeiting damages the industry.	0.76	9.74	0.77	9.85	0.81	10.36	0.80	10.42
Obtaining counterfeit products is illegal.	0.91	12.02	0.87	11.45	0.82	10.71	0.85	11.16
Obtaining counterfeit products is unethical.	0.92	12.21	0.85	11.17	0.84	11.05	0.86	11.24
Decision/intention to obtain ($\alpha_a = 86$, $\alpha_b = 80$, $\alpha_c = 84$, $\alpha_d = 82$)								
I would encourage friends to obtain counterfeit products.	0.82	10.77	0.77	9.84	0.68	7.98	0.66	7.36
I would consider giving counterfeit products to a friend.	0.67	7.79	0.71	8.72	0.61	6.97	0.71	8.75
I would obtain counterfeit products from the Internet.	0.85	11.15	0.81	10.52	0.72	9.12	0.82	10.81
I would obtain counterfeit products from a vendor.	0.91	12.04	0.85	11.14	0.80	10.39	0.92	12.23
Obtaining counterfeit ($\alpha_a = 75$, $\alpha_b = 75$, $\alpha_c = 83$, $\alpha_d = 74$)								
Have you ever obtained a counterfeit product?	0.74	9.54	0.69	8.17	0.64	7.16	0.71	8.83
In the past two years, how many times have you obtained a counterfeit good?	0.93	14.55	0.73	9.31	0.83	10.98	0.82	10.82

Table 3. Multiple Regression Results.

Construct	Brazil[a]		China[b]		India[c]		Russia[d]	
	Standard Beta	*p*-Value	Standard Beta	*p*-Value	Standard Beta	*p*-Value	Standard Beta	*p*-Value
Dependent variable: Attitudes towards counterfeiting								
Demographics	0.067	0.526	0.034	0.683	0.166	0.121	0.060	0.497
Perceived product quality	0.275	0.031	0.245	0.052	0.237	0.067	0.272	0.039
Product price	0.716	0.001	0.364	0.044	0.356	0.028	0.542	0.017
Shopping experience	0.212	0.036	0.252	0.258	0.263	0.148	0.248	0.076
Dependent variable: Decision/intention to obtain								
Demographics	0.371	0.276	0.364	0.116	0.182	0.307	0.179	0.365
Perceived product quality	0.604	0.007	0.287	0.024	0.386	0.019	0.724	0.001
Product price	0.418	0.027	0.567	0.008	0.414	0.030	0.522	0.012
Shopping experience	0.173	0.367	0.232	0.082	0.296	0.074	0.048	0.437
Dependent variable: Obtaining counterfeit								
Attitudes toward counterfeiting	0.256	0.201	0.119	0.437	0.196	0.327	0.267	0.035
Decision/intention to obtain	0.723	0.001	0.513	0.015	0.247	0.039	0.786	0.001

[a] $R^2 = 0.447$, $F = 4.790$ ($p = 0.005$).
[b] $R^2 = 0.704$, $F = 12.661$ ($p = 0.001$).
[c] $R^2 = 0.521$, $F = 8.348$ ($p = 0.001$).
[d] $R^2 = 0.525$, $F = 8.589$ ($p = 0.001$).

and Indian consumers, but it is insignificant for Brazilian and Russian consumers. Further, Russian and Brazilian consumers expressed stronger intention to purchase counterfeits than others, and Russians, Brazilians, and Chinese reported they obtained more counterfeits than Indian consumers. To test for interaction effects among price, perceived quality, shopping experience, and attitude toward counterfeiting, product by country comparisons for these variables were regressed on intention to buy and actual purchase. Two minor but significant interactions were observed (out of the 16 examined), each accounting for less than ½ % of the variance. Thus, while there are different levels of complicity by country, relationships among study variables did not vary to a meaningful degree.

IMPLICATIONS FOR FUTURE RESEARCH AND LIMITATIONS

The results suggest that consumers in emerging markets are more alike than different in their reasons, but not necessarily likelihood, for complicity, and thus may have similar reactions to anti-counterfeiting tactics regardless of cultural ties. Although more study is required to validate our results, these findings support the need to profile the global complicit consumer according to a few key characteristics. In contrast to the framework proposed by Eisend and Schuchert-Güler, this multicountry investigation shows that antecedents with the greatest impact upon consumer attitudes toward and decision/intentions to obtain fakes are product (i.e., price and quality) and purchase situation/mood. In contrast, demographics and national origin were poor predictors of complicity. Although each country market has different sized segments of complicit consumers, the predictors remained the same across demographics of age, gender, education level, income, and country markets. This result is consistent with counterfeiting studies that have been inconclusive and unable to replicate demographic/individual and social/cultural differences as predictors.

Fig. 1 is a revised frame to guide research and actions that may help reduce consumer complicity and the purchase of counterfeit products. Our model supports and enhances Eisend and Schuchert-Güler's framework in the following ways: (1) "person" and "cultural context" predictors are eliminated because of a lack of meaningful impact, (2) purchase situation, price, perceived quality, attitudes toward counterfeits, decisions/intentions, and purchase behavior are reorganized to support continued theory development, (3) conceptualizations of processes that color or frame consumers' thinking – cognitive dissonance and expected utility – are refined to better capture their potential influences, and (4) best ways to reduce complicity are suggested as increasing dissonance or lowering capacity to diminish it, and increasing associated costs and/or decreasing-associated benefits with acquisition of counterfeits. As always, additional scholarship to clarify/amend these findings is warranted.

Therefore, the most viable anti-counterfeiting strategies may involve manipulation of value perceptions of real and fake items along with demarketing tactics that may significantly affect the quality of the shopping experience. This research suggests cognitive dissonance can be used to lessen repetitive complicit consumer behavior, through a decrease in the ability to use cognitive dissonance reduction to rationalize illicit purchases. This

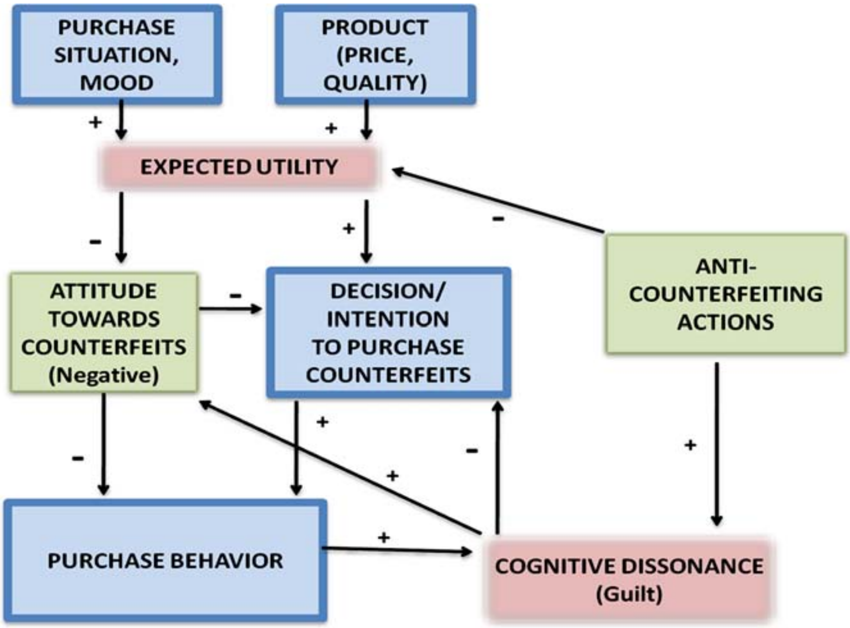

Fig. 1. Consumer Complicity with Counterfeit Products.

practice has meaningful implications for marketers working to protect brands. Further, our framework suggests that managers and policy makers must focus more on consumer behavior across national markets, and decisions and intentions to engage in the illicit trade, as opposed to evaluating significance of cultural underpinnings and moral reasoning that may (or may not) shape attitudes toward counterfeits.

Fig. 1 shows potential viability of cognitive dissonance to lessen repetitive complicit consumer behavior. For example, a direct way to lower expected utility of counterfeit products is price reductions for *legitimate* goods and services, which may not be the first choice of managers who seek price-based brand exclusivity and/or a certain levels of profitability. Thus, utility altering nonprice strategies can be crafted that increase value of legitimate products, thereby increasing perceived quality at given prices and modifying price differentials between legitimate and counterfeit items. Consider that current electronic distributions channels in the movie industry from legitimate suppliers have driven down price even further and make consumers question the value of the fake movie. The key is to find the right price/quality tradeoff so that consumers no longer choose illicit alternatives, especially in a virtual environment.

Of course, studies have limitations and our use of self-reports rather than actual behavior suggests the possibility of a social-acceptability bias that may have led to underreporting of complicity. Nonetheless, the use of non-university subjects from diverse countries and more than one product may yield greater external validity than most previous investigations. Also, BRIC countries are important national trading markets and important markets within the world's population, yet significant diversity exists elsewhere and replication using consumers living in Europe, Canada, Australia, and Africa would provide a fuller indication of the value of these findings. These investigations might also employ alternative goods and industries to ensure that nuances associated with these two products did not drive findings because of unrecognized similarities or differences in characteristics or shopping experiences.

ACKNOWLEDGMENTS

The names of authors are listed alphabetically and contributed equally to the research. We wish to thank the Editor for his helpful and supportive comments about this chapter. The first and third authors greatly appreciate financial assistance from the endowment of the Fred J. Springer Chair in Business Leadership. The second author thanks the Naclerio family for their continuing support of his research.

REFERENCES

ACG. (2004). *What consumers really think about fakes*. Retrieved from http://www.a-cg.org/guest/pdf/Why_you_should_care.pdf. Accessed on July 1, 2009.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, *103*(3), 411–423.

Babin, B. J., Darden, W. R., & Griffin, M. (1994). Work and/or fun: Measuring hedonic and utilitarian shopping. *Journal of Consumer Research*, *20*(4), 644–656.

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation modeling. *Journal of the Academy of Marketing Science*, *16*(Spring), 74–94.

Bian, X., & Moutinho, L. (2009). An investigation of determinants of counterfeit purchase consideration. *Journal of Business Research*, *62*(3), 368–378.

Bian, X., & Veloutsou, C. (2007). Consumers' attitudes regarding non-deceptive counterfeit brands in the UK and China. *Journal of Brand Management*, *14*(3), 211–222.

Bloch, P. H., Bush, R. F., & Campbell, L. (1993). Consumer 'accomplices' in product counterfeiting: A demand-side investigation. *Journal of Consumer Marketing*, *10*(4), 27–36.

Chaudhry, P., & Stumpf, S. (2010). Consumer complicity with counterfeit products. *Journal of Consumer Marketing*, *28*(2), 139–151.

Chaudhry, P., & Zimmerman, A. (2009). *The economics of counterfeit trade: Governments, consumers, pirates and intellectual property rights*. Germany: Springer.

Chiou, J., Huang, C., & Lee, H. (2005). The antecedents of music piracy attitudes and intentions. *Journal of Business Ethics*, *57*(2), 161–174.

Commuri, S. (2009). The impact of counterfeiting on genuine-item consumers' brand relationships. *Journal of Marketing*, *73*(3), 86–98.

Cordell, V. V., Kieschnick, R. L., Jr., & Wongtada, N. (1996). Counterfeit purchase intentions: Role of lawfulness attitudes and product traits as determinants. *Journal of Business Research*, *35*(1), 41–53.

Counterfeit Drugs and China. (2011). Retrieved from http://www.cmpi.org/in-the-news/testimony/counterfeit-drugs-and-china-new/. Accessed on January 18.

Counterfeiting Intelligence Bureau. (2011). Retrieved from http://www.icc-ccs.org/home/cib

Currim, I. S., & Sarin, R. K. (1983). A procedure for measuring and estimating consumer preferences under uncertainty. *Journal of Marketing Research*, *20*(3), 249–256.

Desarbo, W. S., Kim, J., Choi, S. C., & Spaulding, M. (2002). A gravity-based multidimensional scaling model for deriving spatial structures underlying consumer preference/choice judgments. *Journal of Consumer Research*, *29*(1), 91–100.

Ding, M. (2007). A theory of intraperson games. *Journal of Marketing*, *71*(2), 1–11.

Eisend, M., & Schuchert-Güler, P. (2006). Explaining counterfeit purchases: A review and preview. *Academy of Marketing Science Review,* 2006(12). Retrieved from www.amsreview.org/articles/eisend12-2006.pdf

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.

Gentry, J. W., Putrevu, S., & Shultz, C. J., II. (2006). The effects of counterfeiting on consumer search. *Journal of Consumer Behaviour*, *5*(3), 245–256.

Hauser, J. R., & Urban, G. L. (1979). Assessment of attribute importance and consumer utility functions: Von Neumann-Morgenstern theory applied to consumer behavior. *Journal of Consumer Research (Pre-1986)*, *5*(4), 251–262.

Hofstede, G. (1980). *Culture's consequences, international differences in work-related values* (*Cross Cultural Research and Methodology*). Newbury Park, CA: Sage.

Holbrook, M. B., & Hirschman, E. C. (1982). The experiential aspects of consumption: Consumer fantasies, feelings, and fun. *Journal of Consumer Research*, *9*(2), 132–140.

KPMG. (2006). *IP strategies for consumer companies*. Retrieved from http://www.kpmg.cz/czech/images/but/0607_CONS_IP_Strategies.pdf. Accessed on September 1.

Kwong, K. K., Yau, O. H. M., Lee, J. S. Y., Sin, L. Y. M., & Tse, A. C. B. (2003). The effects of attitudinal and demographic factors on intention to buy pirated CDs: The case of Chinese consumers. *Journal of Business Ethics*, *47*(3), 223–235.

Leisen, B., & Nill, A. (2001). Combating product counterfeiting: An investigation into the likely effectiveness of a demand-oriented approach. *American Marketing Association. Conference Proceedings*, *12*, 271–277.

Mehta, N., Chen, X., & Narasimhan, O. (2008). Informing, transforming, and persuading: Disentangling the multiple effects of advertising on brand choice decisions. *Marketing Science*, *27*(3), 334–357.

Nia, A., & Zaichkowsky, J. L. (2000). Do counterfeits devalue the ownership of luxury brands? *Journal of Product and Brand Management*, *9*(7), 485–497.

Nill, A., & Shultz, C. J., II. (1996). The scourge of global counterfeiting. *Business Horizons*, *39*(6), 37–42.

O'Shaughnessy, J., & O'Shaughnessy, N. J. (2005). Considerations of equity in marketing and Nozick's decision-value model. *Academy of Marketing Science Review*, *10*, 1–21.

Peace, A. G., Galletta, D. F., & Thong, J. Y. L. (2003). Software piracy in the workplace: A model and empirical test. *Journal of Management Information Systems*, *20*(1), 153–177.

Penz, E., & Stöttinger, B. (2004). Forget the "real" thing-take the copy! An exploratory model for the volitional purchase of counterfeit products. *Advances in Consumer Research*, *32*, 568–575.

Penz, E., & Stöttinger, B. (2008). Corporate image and product similarity-assessing major demand drivers for counterfeits in a multi-country study. *Psychology & Marketing*, *25*(4), 352–381.

Poison Pills. (2010). Retrieved from http://www.economist.comwww.economist.com. Accessed on January 10, 2011.

Prendergast, G., Chuen, L. H., & Phau, I. (2002). Understanding consumer demand for non-deceptive pirated brands. *Marketing Intelligence & Planning*, *20*(7), 405–416.

Putler, D. S. (1992). Incorporating reference price effects into a theory of consumer choice. *Marketing Science*, *11*(3), 287–309.

Robertson, C., Gilley, K., Crettenden, V., & Crittenden, W. (2008). An analysis of the predictors of software piracy within Latin America. *Journal of Business Research*, *61*(6), 651–656.

Schwartz, S. (1999). A theory of cultural values and some implications for work. *Applied Psychology: An International Review*, *48*(1), 23–47.

Steenkamp, J. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*(1), 78–90.

Treise, D., Weigold, M. F., Conna, J., & Garrison, H. (1994). Ethics in advertising: Ideological correlates of consumer perceptions. *Journal of Advertising*, *23*(3), 56–59.

Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.

Voss, K. E., Spangenberg, E. R., & Grohmann, B. (2003). Measuring the hedonic and utilitarian dimensions of consumer attitude. *Journal of Marketing Research*, *40*(3), 310–320.

Wang, F., Zhang, H., Zang, H., & Ouyang, M. (2005). Purchasing pirated software: An initial examination of Chinese consumers. *Journal of Consumer Marketing*, *22*(6), 340–351.

Wee, C., Tan, S., & Cheok, K. (1995). Non-price determinants of intention to purchase counterfeit goods. *International Marketing Review*, *12*(6), 19–46.

Wilcox, K., Kim, H., & Sen, S. (2009). Why do consumers buy counterfeit luxury brands? *Journal of Marketing Research*, *46*(2), 247–259.