Structural equation modelling in marketing and business research

Critical issues and practical recommendations

Francisco J. Martínez-López
Department of Management, University of Granada, Granada, Spain
Juan C. Gázquez-Abad
Department of Marketing, University of Almería, Almería, Spain, and
Carlos M.P. Sousa
Department of Marketing, University College Dublin, Dublin, Ireland

Received 10 February 2010
Revised 6 July 2010
28 October 2010
Accepted 14 January 2011

Abstract

Purpose – Structural equation modelling (SEM) is a method that is very frequently applied by marketing and business researchers to assess empirically new theoretical proposals articulated by means of complex models. It is, therefore, a logical thought that the quality of the new advances in marketing and business theory depends, in part, on how well SEM is applied. This study aims to conduct an extensive review and empirical analysis of a broad variety of classic and recent controversies and issues related with the use of SEM, in order to identify problematic questions and prescribe a compendium of solutions for its suitable application.

Design/methodology/approach – The main analyses were conducted on a sample of 191 SEM-based papers and 472 applications, i.e. all the SEM-based studies published in four leading marketing journals during the period 1995-2007.

Findings – Despite the maturity of SEM, its application in marketing research still has notable room for improvement. This is a general conclusion based on numerous problems detected and discussed here.

Practical implications – The study provides plausible solutions to the problems identified, a useful guide that is easy to follow and to apply adequately to present SEM issues in marketing and business studies.

Research limitations/implications – The sample of SEM-based papers and applications is limited to four publication outlets. A wider set or/and other journals different to those analyzed here may be preferred.

Originality/value – This is a valuable and timely study of the application of SEM in marketing and business research, and is also useful as a guiding framework for good practice. Likewise, as the problems discussed here presumably occur in other areas of social science, this paper should be welcome beyond the borders of the business disciplines.

Keywords Structural equation modelling, Critical review, Controversies, Good practice, Marketing, Research

Paper type Research paper

The authors sincerely thank the Editor-in-Chief, Professor Nick Lee, and the anonymous reviewers for their insights and valuable guidance. They also acknowledge the constructive advice of Professor Hans Baumgartner (The Pennsylvania State University Smeal College of Business) and Professor Pradeep K. Chintagunta (The University of Chicago Graduate School of Business, USA) at an earlier stage of this research.

Dr Carlos M.P. Sousa is now a Reader in Marketing at Durham Business School, Durham University. Prior to joining Durham, he was Lecturer in Marketing at the UCD Michael Smurfit Graduate Business School at University College Dublin (Ireland).

European Journal of Marketing
Vol. 47 No. 1/2, 2013
pp. 115-152

0309-0566
DOI 10.1108/03090561311285484

1. Introduction

> . . . while the SEM practices of many social scientists may be less than optimal, the use of SEM as a powerful data analysis and causal modelling tool is here to stay (Barret, 2007, p.820).

As proposed theories increase in complexity, so too must the causal structures which represent them, and SEM seems to be the preferred method of estimation by academics (Shook *et al.*, 2004). The potentialities offered by SEM to empirically analyse theoretical relations are the product of a sum of multidisciplinary contributions, which started in the seventies (see, for more detail: Aigner and Goldberger, 1977; Blalock, 1971; Goldberger and Duncan, 1973; Jöreskog and Wold, 1982), though the origins of SEM are placed decades before. SEM is based on three main pillars (Bollen, 1989; Hair *et al.*, 2005):

(1) the path analysis;

(2) the synthesis of latent variables and measurement models; and

(3) methods to estimate the parameters of structural models.

These combined features make SEM a method with a singular philosophy of application which differs from the others used in marketing modelling (Bagozzi, 1994). In sum, SEM is a powerful research tool for theory testing (Steenkamp and Baumgartner, 2000).

Notwithstanding, the sole application of SEM does not guarantee reliable theoretical findings. SEM-based procedures provide researchers with more benefits and flexibility than precedent, first-generation techniques, for the interplay between theory and data, but to be precise, they must be correctly applied (Chin, 1998a). It would not, therefore, be risky to argue that the quality of marketing knowledge generated, and in particular that knowledge proceeding from marketing studies using SEM, is dependent on how well researchers apply this modelling methodology. This is a commanding reason to justify the interest in studying how SEM is being applied in marketing, in order to indentify inappropriate practices and provide useful recommendations for future applications. Also, though this statistical method has been subjected to multiple evolutions and improvements since its introduction, it is still affected by "some very old and familiar problems, constraints and misconceptions" (Tomarken and Waller, 2005, p. 56). In fact, recent reviews on its use have revealed serious flaws (see Brannick, 1995; Chin, 1998a; Shook *et al.*, 2004). The marketing arena is not immune to them (see: Baumgartner and Homburg, 1996; Hulland *et al.*, 1996; Steenkamp and Baumgartner, 2000; Steenkamp and van Trijp, 1991). For this reason, it is convenient to periodically review the modellers' common assumptions and practices in respect of SEM in our discipline (Chin *et al.*, 2008).

Here, the reader will meet a discussion about, and plausible responses to many SEM themes:

- Questions which appear with relative recurrence in the specific literature (e.g. assessment of measures, distinction between theory construction and testing, validity of results, etc.).

- Questions apparently solved in the past, but now revived (e.g. the logic of model testing with SEM, approaches to evaluate the fit of models, cut-off values for Approximate Fit Indices (AFIs), etc.).
- New questions with a relatively recent life time (e.g. how to act when the chi-square fit test does not work, statistical power to assess models' fit; cross-validation of models, etc.).

This is not the first attempt to analyse the use of SEM. Several studies, with more or less levels of detail in the analyses, have appeared during the last two decades either in marketing or in other scientific disciplines; the renowned study of Baumgartner and Homburg (1996) has been taken, in this regard, as a main reference to set the starting framework of analysis. There are, however, diverse strengths which make this paper an interesting contribution. Much of the SEM classic problems and controversies are treated, incorporating recent views on them. But, beyond these theoretical questions, special attention is paid to the applied orientation of SEM. We have undertaken an extensive literature review on a variety of themes, many of them purely technical, in order to illustrate critical analyses of issues associated with the specification, evaluation and validation of SEM-based models. A large sample of SEM-based papers and applications published in four major marketing journals provides detailed and data-supported information on the analysed issues. Finally, a summary of practical recommendations is offered.

2. SEM issues analysed in previous review studies

SEM in its current state has evolved, in its diverse compounding parts, through decades. Since the late eighties, varying research efforts from several scientific disciplines, mainly belonging to the social sciences, have attempted to analyse the use of SEM, in order to identify incorrect and best practices, and prescribe recommendations to improve its application; other similar studies have also been conducted in disciplines as management (see: Chin, 1998a; Garver and Mentzer, 1999; James and James, 1989; Medsker *et al.*, 1994; Shook *et al.*, 2004) and psychology (see Brannick, 1995; Breckler, 1990; MacCallum and Austin, 2000). In Table I, a synthesis of significant SEM-review papers on marketing themes is shown. Only two of them (see: Baumgartner and Homburg, 1996; Hulland *et al.*, 1996) presented empirical analyses of SEM-based applications to support the theoretical issues treated; empirical support is particularly important, if a more exact view on the SEM issues in marketing is wanted.

Among these reviews on SEM, that by Baumgartner and Homburg (1996) deserves particular attention for several reasons. First, these authors discussed and critically reviewed a broad variety of issues associated with SEM methodology. This is helpful, especially when compared with other SEM reviews, for its explicit, clear and structured presentation of many areas of analysis. Nevertheless, most of the other reviews, while undoubtedly comprising a rich database on SEM issues reviewed, opted to discuss those issues without providing much quantitative information in detail. This feature limits the utility and projection of those studies for readers. Second, the wide time period that Baumgartner and Homburg (1996) consider, starting from the late seventies, which marks the beginning of the use of SEM in marketing, and extending until the mid-nineties (years 1977-1994) is valuable. Third, and no less interesting, the impact of this study in the marketing academic community has been substantial, it

Table I.
Previous reviews on the use of SEM in marketing

Studies – in alphabetic order	Journals analyzed	SEM based papers	Period of time	Controversies (in detail)	Areas of analysis in SEM applications – issues related with: Model specification	Modelling approach	Types of SEM models	Modelling strategies	Sample size	Degrees of freedom	Model estimation	SEM software packages	Model fit assessment
Baumgartner and Homburg (1996)	4 journals: *Journal of Marketing*, *Journal of Marketing Research*, *International Journal of Research in Marketing*, and *Journal of Consumer Research*	149 papers – 184 SEM applications	1977-1994		◆		◆		◆	◆	◆		◆
Chin *et al.* (2008)	Paper without empirical analysis of SEM-based applications	–	–		◆						◆		◆
Hulland *et al.* (1996)	11 journals: *J. of Marketing Research*, *J. of Marketing*, *J. of Consumer Research*, *Management Science*, *Operations Research*, *Marketing Science*, *J. of Business*, *J. of Advertising Research*, *Journal of Retailing*, *J. of Business Research*, and *International J. of Research in Marketing*	186 papers – 343 SEM applications	1980-1994		◆	◆			◆		◆		◆
Steenkamp and Baumgartner (2000)	This study analyzes three journals: *International Journal of Research in Marketing*, *Journal of Marketing*, *Journal of Marketing Research*. But, just with the purpose of approaching to the number of papers using SEM over the total. Hence, the SEM issues discussed here are not supported with *ad hoc* empirical analysis of SEM-based applications	22	1999	◆	◆								
Our study	Four journals: *Journal of Marketing*, *Journal of Marketing Research*, *International Journal of Research in Marketing*, and *Journal of Consumer Research*	191 papers – 466 SEM applications	1995-2007	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆

(continued)

Studies – in alphabetic order	Journals analyzed	SEM based papers	Period of time	Areas of analysis in SEM applications – issues related with: Chi test (in detail)	Equivalent models	Power analysis	AFIs (in detail)	Reliability constructs	Predictive accuracy (R^2)	Model modification	Cross validation models	Reporting of models	Other SEM issues
Baumgartner and Homburg (1996)	4 journals: *Journal of Marketing*, *Journal of Marketing Research*, *International Journal of Research in Marketing*, and *Journal of Consumer Research*	149 papers – 184 SEM applications	1977-1994	◆			◆	◆	◆	◆	◆		◆
Chin *et al.* (2008)	Paper without empirical analysis of SEM-based applications	–	–							◆		◆	◆
Hulland *et al.* (1996)	11 journals: *J. of Marketing Research*, *J. of Marketing*, *J. of Consumer Research*, *Management Science*, *Operations Research*, *Marketing Science*, *J. of Business*, *J. of Advertising Research*, *Journal of Retailing*, *J. of Business Research*, and *International J. of Research in Marketing*	186 papers – 343 SEM applications	1980-1994	◆	◆			◆					◆
Steenkamp and Baumgartner (2000)	This study analyzes three journals: *International Journal of Research in Marketing*, *Journal of Marketing*, *Journal of Marketing Research*. But, just with the purpose of approaching to the number of papers using SEM over the total. Hence, the SEM issues discussed here are not supported with *ad hoc* empirical analysis of SEM-based applications	22	1999										◆
Our study	Four journals: *Journal of Marketing*, *Journal of Marketing Research*, *International Journal of Research in Marketing*, and *Journal of Consumer Research*	191 papers – 466 SEM applications	1995-2007	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆

Table I.

being one of the most cited papers (i.e. the second) in the history of the IJMR (see Stremersch and Lehmann, 2008, p. 145).

It is the issues analysed by these authors that form the backdrop to our study. We focus firstly on updating the application of SEM in marketing over the last decade (years 1995-2007) in detail and, secondly on making comparative analyses of the results for both periods, in order to evaluate how the diverse SEM issues have evolved over time. Clearly, the second focus will have significant appeal for marketing researchers and modellers. However, our study should not merely be seen as an updating of such an excellent and thought-provoking paper, at the time of publication. On the contrary, as already noted, much has changed, old and new problematic issues are in vogue and there is tension in the SEM literature currently. Hence, this paper also provides a timely and added-value critical review. Moreover, as can easily be seen from Table I, the present study covers the broadest variety of topics, when compared with the other reviews of SEM.

3. Research methodology

Regardless of the particular themes treated within each of them, the analysis is structured into the following main blocks of issues:

- Evolution of the use of SEM in marketing.
- Type of SEM-based models.
- SEM-based modelling strategies.
- Issues related to the specification of models.
- Issues related to sample size.
- Assessment of models.
- Validation of models.
- Reporting.

These blocks of themes have been also specified after reviewing the diversity of issues treated by previous reviews on SEM, in order to work with a congruent, exhaustive and appealing structure for readers. But, together with these issues that will be discussed in the section of the results (Section 4), diverse controversies in SEM, also supported with empirical data, are tackled (Section 5).

As already mentioned, the sample of SEM-based papers and applications has been extracted from four publication outlets: *JM*, *IJRM*, *JCR* and *JMR*. The number of marketing journals considered here could have been wider, but this strategy would have introduced some "noise" into one of our research purposes; i.e. assessing the evolution in SEM issues treated by Baumgartner and Homburg (1996) in their study, and searching for benefits arising from practising specific comparative analyses between the periods considered by both studies; i.e. 1977-1994 and 1995-2007. Besides, these are traditionally well ranked in the editions of the Journal of Citation Reports (SSCI/ Business) published by Thomson. Also, three of the set of four journals considered in this study are the most preferred (1st: *JM*; 2nd: *JMR*; 3rd: *JCR*) by marketing journal readers (*IJRM* is in the 14th position), in a list of 110 academic journals in marketing and other related areas (see, for more detail: Hofacker *et al.*, 2009).

In sum, 1,300 of all the papers published by these publications for the period of years 1995-2007 have been analysed. This figure represents the number of papers which worked with any kind of causal approach, with or without empirical support, to a marketing research problem. From them, a specific detailed review was undertaken, which involved checking the SEM-related issues considered in this study within all those papers containing any kind of SEM application. Overall, our database was finally composed of 191 SEM-based papers and 472 SEM applications.

The analyses conducted to reach the first draft stage were undertaken by three judges/coders, all being PhD holders and marketing lecturers. In order to ensure inter-coder reliability, prior to the analysis of the full sample, several sessions were held to unify criteria in order to design a procedure for the analysis. Then, a pilot-test was carried out on a sample of 30 papers. The reliability of agreement in the codification of SEM issues was assessed, and the levels of agreement were optimal for most issues. Notwithstanding, the conclusions of this test were applied to refine the coding criteria for a few of the issues where the agreement was not complete; i.e. a minority of aspects which could have a more subjective nature, such as determining the type of SEM models or the modelling strategy followed by the authors. Finally, the study was developed for the full sample.

4. Descriptive results and first critical thoughts

4.1 Overall view on the evolution of SEM in marketing over time

Figure 1 integrates our results with those from the study by Baumgartner and Homburg, in order to obtain a full period of more than three decades – i.e. 31 years (1977-2007) – using SEM in the four marketing journals considered here. The rise in SEM use for causal modelling purposes in marketing is evident from the late 1970s until the early 1990s, during which a long learning process is observed. Beyond those years, a clear decrease in the overall number of papers per year occurs until the year 2000, with the exception of 1998. Finally, in the last decade, the application of SEM seems to have revitalized. However, when relative data are observed – i.e. the use of SEM with respect to the total number of causal-based papers with empirical applications per year – rather than a revitalisation, the use of SEM in the whole set of causal testing techniques applied in the sample of journals produces stability or even a decrease. This evidence from these four leading marketing journals suggests that the use of SEM in marketing modelling has clearly achieved a phase of maturity.

4.2 Type of SEM-based models

The database has been firstly structured as the following classic typology of three SEM-based models, also used in previous reviews on the use of SEM (see Baumgartner and Homburg, 1996; MacCallum and Austin, 2000): confirmatory measurement models (Type I); single-indicator structural models (Type II), associated with path analysis; and integrated measurement/latent variable models (Type III). This typology allows us to discriminate the analyses we produce, although a general approach, without making any distinction between types of models, is also offered.

According to the typology of SEM models analysed, integrated measurement/latent variable models (Type III) are the most applied in our sample of SEM-based papers (69 per cent). Next, in a more distant second place, is the confirmatory SEM (Type I), or the application of SEM to analyse the measurement structure subjacent to a set of observed variables (26.3 per cent); usually, this type is used to analyse the reliability and validity of measurement scales, as well as dimensional structures with no latent

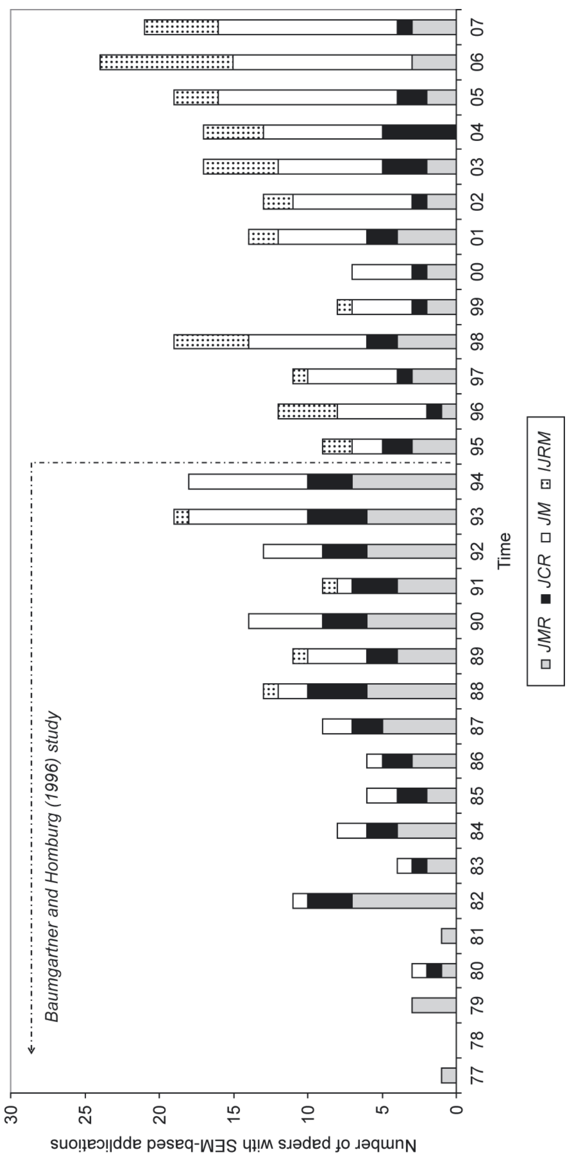

Figure 1.
Number of papers using SEM per year and journals for the whole sample of publications (31 years)

endogenous variables. Finally, single indicator structural models (Type II) have been marginally applied (7.7 per cent). Here, we would like to make two remarks[1], in order to clarify eventual readers' questions in this regard:

(1) structural models not completely compounded by one-single indicator constructs, but rather having just some of them, have been classified as Type III; and
(2) structural models being originally type III, with all or part of the constructs based on multi-item scales, but totally transformed into one-single indicator constructs (e.g., Chaudhuri and Holbrook, 2001; Kempf and Smith, 1998; Price *et al.*, 1995), have been also coded as Type II; i.e. full structural models would be converted in path models.

This practice aims to both reduce the models' complexity, and work with acceptable variable-sample size ratios, thus using aggregated (composite) measures, such as summary- or average-item constructs, when estimating the structural coefficients (Calantone *et al.*, 1996; Cavusgil and Zou, 1994; Homer and Yoon, 1992). That said, although they share some sought benefits, the said practice should not be confused with the so-called item-parcelling technique (Bengtson *et al.*, 2005, p. 477) (see section 4.4).

The clear transference of papers from Types I and II to Type III is observed, if compared to the sample of SEM-based papers analysed in the study by Baumgartner and Homburg (1996). Also, the Type II models, more applied in the first decade of SEM applications in marketing modelling, present a marginal representation (less than one paper per year) in our sample of journals. This is logical if one considers the increasing attention paid by researchers and editorial boards (editors and reviewers) to working with multiple observed variables per construct, with the aim of assuring good-quality measurement models. In this context, Type II models are not adequate and should be avoided. This recommendation is also suggested for originally-designed full structural models that, for practical reasons (see above), are transformed in Type II to obtain the path coefficients. Hence, though unlike pure Type II models this resource[2] starts from a measurement model of base, it finally implies a clear loss of technical rigour for a practicality gain, as the structural coefficients are also based on a path analysis (see Lee and Calantone, 1998).

4.3 SEM-based modelling strategies (formulation modes)

Three types of strategies (Jöreskog and Sörbom, 1993) have been considered here:

(1) Strictly confirmatory/confirmatory modelling that tests a theoretical model with no changes in, or modifications to, the original model.
(2) Competitive modelling/model comparison that analyses alternative or "rival" models with the aim of selecting the most valid.
(3) Model development/generating that estimates a model initially specified, and then makes subsequent re-specifications with the aim of achieving a final model with better fit.

The primary modelling strategy applied is the confirmatory one, representing about 70 per cent of our sample of SEM-based papers; competitive modelling and model development strategies are secondary alternatives to researchers, with about 16 per cent and 14 per cent of papers respectively. There are two plausible, related reasons to explain these percentages. First, competitive modelling and model development strategies are more complex when applied and explained than confirmatory modelling. This question leads to the second reason, that being the necessarily lengthier manuscripts, which are inconvenient when considering article space constraints and the number of words per manuscript. Curiously, due to the particularities of SEM, Babin *et al.* (2008) recently suggested that journal editors should allow more space for SEM-based papers if necessary.

However, these two less used strategies are clearly more coherent in terms of the applied vision of marketing modelling, which is more dynamic and focused on the implementation of models (Naert and Leeflang, 1978; Lilien and Kotler, 1990). In turn, strictly confirmatory strategy is more respectful of the classic vision of the scientific research process (Bunge, 1967), more related to the traditional vision of the marketing modelling process (Lilien *et al.*, 1992; Wierenga and Van Bruggen, 2000), and focused on specifying, estimating (parameterisation stage) and evaluating the model. Consequently, based on certain claims regarding the current suitability of a dynamic orientation in the marketing modelling process (Leeflang *et al.*, 2000), modellers should make greater use of competitive modelling and, especially, model development strategies.

Notwithstanding, researchers should be rigorous and cautious when opting for a model generating strategy (see Hughes *et al.*, 1986). In particular, two main precautions should be taken:

(1) Re-specification should be based on theoretical and content considerations (Anderson and Gerbing, 1988), otherwise modifications would respond to data-driven considerations that may lack validity (MacCallum, 1986). In respect of our data, about 75 per cent of applications which used a model development strategy were aided by statistical analyses (e.g. the Wald test), although 45 per cent of such applications did not theoretically justify re-specifications.

(2) Cross-validation of models is even more necessary in this scenario (Chin, 1998a; Kelloway, 1995), though our data reflect that just 22 per cent of applications following this particular strategy were validated in some way.

In sum, past SEM studies during the 1980s and 1990s had already detected these faults in the application of generating model strategies, and as yet they still remain.

4.4 Issues related to the specification of models

In SEM, a measurement model allows the modeller to set the relationships between observed variables (i.e. indicators) and their respective unobserved variables (i.e. latent variables or constructs) by defining a particular structural model (Bollen, 1989). How such a model is designed, and thus, how many indicators per construct are set, has been said to highly influence the degree to which the structural model is well identified. In fact, Anderson and Gerbing (1988) advocated a two-stage approach in SEM, where the measurement model was developed before, and independently from, the structural model. We have observed that this is now a usual practice in marketing, being adopted in 87 per cent of the SEM-based papers analysed (see Table II).

On the basis of their results, Baumgartner and Homburg (1996) were critical of the ratio observed/latent variables found in the models analysed, specifically for the full structural (Type III) models, which resulted in a median ratio of 2; based on Bollen

Table II. Issues related to the specification of the models (overall and by type of model): 1995-2007

	Total ($n = 472$)	Type I ($n = 144$)	Type (II $n = 13$)	Type (III $n = 315$)
Measurement model specification issues				
Number of observed variables-indicators[a]	24 (13, 36)	21 (12, 32)	6 (3, 7)	26 (15, 37)
Number of latent variables/constructs[a]	7 (5, 10)	6 (3, 9)	6 (3, 7)	8 (6, 11)
Ratio of observed variables to latent variables[a]	3,2 (2,4, 4,4)	3,8 (2,6, 5)	1,0 (1,0,1,0)	3 (2,4, 4,2)
Percentage of models containing at least one single-indicator construct	21,6	17,4	100	23,8
Percentage of models lacking simple structure	14	14	0	14
Percentage of models containing at least one second-order construct	27,5	30	0	27,6
Percentage of models with measurement model graphically provided	38	10,7	23	53,7
Sample size issues				
Sample size[a]	259 (154, 419)	252 (172, 362)	137(108, 193)	280 (156, 452)
Percentage of models where there is no information about the sample size	0,2	0,7	0	0
Number of parameters estimated[a]	63 (35, 93)	50 (29, 92)	21 (4, 29)	65 (44, 94)
Ratio of sample size to number of parameters estimated[a]	4,4 (2,7, 7,8)	4,8 (32, 8,6)	8 (4,6, 37,6)	4 (2,5, 6,7)
Percentage of models with a ratio of sample size to number of free parameters smaller than 5	51,8	50,7	23	54
Percentage of models with a ratio of sample size to number of free parameters smaller than 10	70	74,5	53,8	69,4
Percentage of models with adequate sample size to achieve power analysis of 80 per cent (Test of close fit - Test of exact fit) – MacCallum *et al.* (1996)	(68 - 66,4)	(75,7 - 75,7)	(16,6 - 16,6)	(66,6 - 64,3)
Percentages of models with transversal samples	92	97	92	91
Degrees of freedom issues				
Percentage of models where degrees of freedom are indicated	52,7	48,6	46,2	53,5
Number of degrees of freedom[a] [244/ 70/ 6/ 168]	110 (24, 359)	221 (53, 428)	9 (7, 11)	107 (21, 341)

Notes: [a]Entries are medians (with first and third quartile in parentheses, unless a percentage (or frequency of use) value is indicated. Numbers in brackets [Total/Type I/ Type II/Type III] are the sample sizes on which the norms for the magnitude of various goodness-of-fit measures are based

(1989), it was recommended that each construct be measured by at least three indicators. Our results have evidenced a clear advance in this question. Generally, the number of observed variables has considerably increased, in overall terms, from a median of 11 for the period 1977-1994 to 24 for the last period we have analysed. Thus, while the number of latent variables has also risen (overall median of 7), its increase is proportionally smaller than that observed for the indicators. Obviously, this fact has produced a very positive effect on the median ratio. The explanation for such significant growth, however, is worthy of comment.

First, a general commentary: our sample of SEM-based applications, more than double the number analysed by Baumgartner and Homburg over a longer period of time, presents a notable difference in the distribution of the SEM-based model types. Now, Type III models are predominant, representing about 70 per cent of the models analysed, while the application of Type I is secondary and Type II marginal. The latter question is not strange if the serious weaknesses of Type II models in contributing to a reliable measurement process are taken into account; see also the evolution from the operational definition philosophy (followed by Type II models) to the partial interpretation philosophy (Bagozzi, 1984), currently predominant in the marketing modelling discipline. According to this view, the measurement of constructs on one-single item is discouraged due to its incapacity to gather abstract constructs, and instead the use of multi-item measures is generally favoured (Steenkamp and Baumgartner, 2000). Therefore, the overall growth of the observed/latent variables relationship is logical, considering the greater representation of Type III models in the overall median ratio, as they usually contain the most complex measurement models of all three types. Second, the clear increase in the overall median ratio is not only due to the higher contribution of Type III models to the aggregate figure (for instance, the median number of indicators has increased by more than 100 per cent, from 12 to 26, between both periods of analysis), but also to the contribution of the Type I models.

Certainly, the necessity for basing the measurement of constructs on multi-item scales is one of the mantras that researchers have in mind when applying SEM procedures nowadays. Notwithstanding the foregoing, recent studies have paradoxically shown particular measurement scenarios where multi-item scales (Bergkvist and Rossiter, 2007; Rossiter, 2002), and an excessive number of items per construct (Bagozzi and Edwards, 1998; Little *et al.*, 2002), might be unnecessary or generate inefficiencies; e.g. a higher number of items (i.e. free parameters to estimate) are needed to define a construct hence requiring longer and more expensive surveys, higher chance for variability in representation (i.e. face validity) and residuals correlations, a loss of quality in the sample size-free parameters ratio, etc. It is, therefore, pertinent to close this sub-section with a brief comment in this regard. In particular, the following paragraphs are focused on two[3] notable practices that are used to alleviate eventual problems associated with such scenarios: the adequate use of one single-indicator constructs; and, the item-parcelling technique.

In general, as previously remarked, a measurement philosophy of constructs based on single-indicator scales is not recommended; an extreme case would lead to Type II models, the use of which has been discouraged. The selective use of this kind of construct in full structural models is said to generate diverse problems too (see, for more detail: Bollen, 1989; Ding *et al.*, 1995), so researchers have been traditionally warned against its consideration (Churchill, 1979; Peter, 1979). Based on our data, the use of this type of construct has fallen by about 28 per cent overall. This decrease has been exclusively due to the strong reduction in Type III modellers' use of one single-indicator constructs – down to 23.8 per cent of cases, from 71 per cent in the period 1977-1994. This could be seen as a positive advance (see Baumgartner and Homburg, 1996). Notwithstanding, the use of single-indicator constructs should not be demonized in any event. On the contrary, Rossiter (2002) demonstrates how this can be an equally valid alternative to multi-item measures, when applied to constructs whose object is concrete and singular. Similar findings can be observed in a recent study by Bergkvist and Rossiter (2007)[4], where the use of single-indicator scales is advocated for doubly (i.e. object and attribute) concrete constructs;, e.g. attitude (concrete attribute) toward a brand (a concrete singular object). Therefore, we are prone to extraordinarily recommend a selective and well-justified use of single-indicator measures, as a rigorous alternative to avoid eventual inconvenience related to multi-item scales, when modellers deal with no abstract, but doubly concrete constructs.

On the other hand, the item-parcelling technique searches for a reduction in the number of indicators per latent variable (Kishton and Widaman, 1994; Little *et al.*, 2002; MacCallum *et al.*, 1999; Marsh *et al.*, 1998; Yuan *et al.*, 1997), but without paying a trade-off as a consequence of eliminating indicators that might contribute to the explanation of a construct (Nasser and Takahashi, 2003). Parcelling logic consists of reducing the raw indicators that compound certain multi-item scales by forming groups (i.e. parcels, also called testlets or miniscales) of two or more indicators, usually aggregated by their summation or average, which are then used as the lowest-indicator variables (Bandalos and Finney, 2001). Though this practice is not free of controversy (see Meade and Kroustalis, 2006), it may provide diverse benefits, e.g. reduction in both the number of parameters to be estimated and measurement errors, palliate problems derived from eventual non-normality of raw indicators, improvement of the sample size/indicator ratio, a better model fit, etc.

Nevertheless, item-parcelling is not really applied by marketing researchers, probably due to the technique not being widely known yet. The number of papers that explicitly report dealing with parcels in our database is marginal (e.g. Ahearne *et al.*, 2007; Homburg *et al.*, 2007; Rapp *et al.*, 2006; Sandvik and Sandvik, 2003). However, based on our review of authoritative technical studies, we regard parcelling as a plausible solution to some of these shortcomings that researchers might find when estimating structural models. That said, the technique must be applied with rigour. Basically, we highlight three main points (see, for an extensive and detailed prescription, Bandalos and Finney, 2001):

(1) Assess, by exploratory or confirmatory factor analyses, the dimensionality of the constructs to be parcelled. Though it can be applied to constructs with several facets (in this case, every parcel of items should have a correspondence with every facet), it is highly recommendable only when dealing with unidimensional constructs.

(2) Select a proper parcelling method or strategy, as the way in which the indicators are grouped might be important (see also, Hall *et al.*, 1999; Little *et al.*, 2002).

(3) Avoid using this technique in research whose main object is the development, refinement or testing of scales.

4.5 Issues related to sample size

The size of the sample is important in terms of the generalisation of results, the reliability of the parameters' estimation of the model, and the power analysis of model testing. SEM modellers must ask themselves what the right sample size is. Two main streams of thought can be identified in this respect, depending on whether the size is considered in isolated terms or whether it is evaluated in relation to the number of parameters to be estimated.

Several sample size recommendations can be found for the first stream, either as general recommendations regardless of the particular characteristics in the data/variables, or as a minimum size recommendation for certain situations in the data/variables distribution or methods of estimation applied. For instance, Ding *et al.* (1995) suggest the use of a minimum of 100-150 individuals; others recommend at least 200 in order to reduce eventual biases in the model estimation (see: Kline, 2005; Loehlin, 1998); there are also recommendations depending on the SEM method of estimation applied, for example, for certain asymptotically distribution-free methods of estimation, a sample size in the range of 1,000-2,000 is suitable (see: Boomsma and Hoogland, 2001; Hoogland and Boomsma, 1998; Hoyle, 1995).

However, the second stream of thought seems more appropriate when this issue is considered in overall terms (Baumgartner and Homburg, 1996). Specifically in Table II, based on the suggestions of Bentler (1995) and Bentler and Chou (1987) about the ratio value, differences can be seen between models with ratios as low as 5 (i.e. trustworthy parameter estimates) or 10 (i.e. suitable significant tests) individuals per parameter.

Considering the above, a clear and general conclusion can be extracted: sample sizes used to estimate the SEM-based models are now less appropriate than those used during the period 1977-1994. The median ratio of sample size to number of free parameters is about 4:1, with Type III models showing the lowest median ratio (4). About 52 per cent and 70 per cent of the models analysed were below ratios of 5:1 and 10:1 respectively, compared to the 41 per cent and 73 per cent observed by Baumgartner and Homburg (1996). These results are worrying, as one of their main recommendations was to improve these ratios, by strongly urging modellers to study the appropriate sample size beforehand. Thus, although modellers are more aware of the importance of working with larger sample sizes than in the past (our results show an overall improvement regarding this), an increase in the complexity of models has also been noted, especially in Type III models. Due to this fact, the improvement of sample sizes for models is not enough to justify the increase in the number of parameters to be estimated. In conclusion, modellers must still pay attention to this issue in future research in order to achieve truthful conclusions.

Also, based on the values of reference provided by MacCallum *et al.* (1996) in terms of the models' degrees of freedom, the percentage of SEM applications with a large enough sample size to achieve power analysis of 80 per cent in the chi-square test is detailed by type of model (this issue is treated with more detail in section 5.1.2). The conclusion to be reached in this regard is that there are a significant number of SEM applications where this test is used with inadequate sample sizes, these being about 25 per cent (for Type I) and 35 per cent (for Type III).

4.6 Assessment of models

Diverse controversial questions related to this issue are addressed in section 5, taking as a backdrop the full sample of SEM applications. Previously, some notable and descriptive results by type of models are discussed (detailed results are in Table III).

4.6.1 Assessment of model fit. Another attraction of SEM is its provision of diverse ways of evaluating the model fit. Together with the popular chi-square test, an arsenal of goodness-of-fit indices, also called approximate fit indices, has been growing since its origins; the 1980s were especially prolific in this regard. A common classification (Hair *et al.*, 2005) distinguishes between:

- overall model fit indices;
- incremental fit indices; and
- parsimonious fit indices.

If necessary, the reader can find an introductory explanation on these indices in any good guide to SEM.

Without discriminating by type of model, the percentage of models which used at least one stand-alone fit index (91.5 per cent) is similar to that in the previous period 1977-1994 analysed by Baumgartner and Homburg (92 per cent). However, modellers tend to show, either overall or by type, the same hierarchy of preferences in their use: the chi-square test (88 per cent), GFI (43 per cent) and RMSR (28 per cent). Likewise, one application increasingly used is the RMSEA (66 per cent of models); in part, this increase also explains the reduction in the use of the RMSR. Type III models are where this index is applied more (69.8 per cent); here, the application of the chi-square test has been reduced by 10 per cent. Finally, unlike the RMSEA, no application for the McDonald (1989) measure of centrality (MC) has been noted. Hence, though Baumgartner and Homburg (1996) predicted that these two indices would become widespread, our results support it only for the RMSEA. With respect to the magnitudes observed, the medians for GFI, AGFI, RMSR and RMSEA were 0.91, 0.91, 0.05 and 0.06.

The significant increase in the use of incremental fit indices is one of the most remarkable points of this period. In general, the number of applications where at least one incremental fit index was reported have more than doubled the figure obtained in the period analysed by Baumgartner and Homburg (from 38 per cent to 89 per cent). All the types of models have experienced a minimum growth of 35-40 per cent; especially remarkable is the increase for Type I models (from 31 per cent to a 96 per cent). The main reason is the boost in CFI use, up from 13 per cent during 1977-1994, to 80 per cent in the period 1995-2007. This index, along with the RMSEA, was recommended in the late 1990s instead of others like the GFI and AGFI (Hu and Bentler, 1999), as being less affected by sample size (Fan *et al.*, 1999). Likewise, the growth in use of "other incremental fit indices" has been also significant; these include TLI/NNFI, RNI or IFI. The medians for NFI/BBI, RFI and CFI were 0.92, 0.97 and 0.945.

Regarding the parsimonious fit indices, the data show a residual application; when used, the Parsimonious Normed-Fit-Index (PNFI) or the Akaike Information Criterion (AIC) are the ones preferred; no use of the Parsimonious Goodness-of-fit Index (PGFI) and Critical N has been evidenced. This is especially worrying for those models which follow competitive modelling and model development strategies, where these indices are more convenient to use. On the contrary, modellers clearly tend to base their

Table III.
Issues related to the estimation, assessment and validation of models (overall and by type of model): 1995-2007

	Total ($n = 472$)	Type I ($n = 144$)	Type (II $n = 13$)	Type (III $n = 315$)
Model estimation issues				
Percentage of models estimated by ML	14,8	12,5	7,7	16,2
Percentage of models estimated by GLS	2,9	4,2	0,0	2,5
Percentage of models estimated by WLS	1,2	2,0	7,7	0,6
Percentage of models where the estimation method is not reported	81,1	81,3	84,6	80,09
Issues in assessing overall model fit				
Chi-square test				
Percentage of models for which at least one stand-alone fit index was reported	91,5	97,2	76,9	89,5
Percentage of use of χ^2 test	88,1	88,8	76,9	88,3
Percentage of models where the χ^2 statistic is indicated[a]	59,4	54,7	70	61,3
Percentage of models where p-value related to χ^2 test is indicated[a] (average of p-value)	37,8	32,1	70	39,2
Percentage of models with significant results[a]	49	50,8	10	49,7
Percentage of models with significant χ^2 test, where results are: Completely ignored/ Justified with AFI's/ Justified with data distribution	55,4/ 24,5/ 5,8	55,4/ 26,2/ 6,2	100/ 0/ 0	55,1/ 23,9/ 5,8
Percentage of models using the Satorra-Bentler χ^2 test[a] [5/ 3/ – / 2]	1,2	2,3	0	0,7
Global Approximate-Fit-Indexes				
Percentage of models[a] with ratio χ^2/ d.f equal or below 3 – Jöreskog and Sörbom (1993)	40,6	32,8	50	43,8
Percentage of use of GFI	43,4	32	38,5	48,9
Percentage of use of AGFI	19	13,2	30,76	21,3
Percentage of use of RMSR	28,2	19,4	38,5	31,7
Percentage of use of RMSEA	66,7	61,8	46,2	69,8
Magnitude of χ^2/ d.f[b] [244/ 70/ 6/ 168]	2,14 (1,57, 3,84)	2,35 (1,75, 3,95)	1,12 (0,67, 1,44)	2,11 (1,50, 3,48)
Magnitude of GFI10 [197 / 45/ 5/147]	0,91 (0,88, 0,95)	0,89 (0,86, 0,93)	0,986 (0,98, 0,99)	0,91 (0,88, 0,96)
Magnitude of AGFI[b] [89 / 18/ 4/ 67]	0,91 (0,85, 0,94)	0,87 (0,77, 0,91)	0,955 (0,92, 0,965)	0,91 (0,86, 0,94)
Magnitude of RMSR[b] [125 / 28/ 5/ 92]	0,051 (0,036, 0,08)	0,046 (0,03, 0,06)	0,04 (0,036, 0,044)	0,054 (0,041, 0,093)
Magnitude of RMSEA[b] [306 / 88/ 6/ 212]	0,06 (0,045, 0,074)	0,060 (0,048, 0,08)	0,054 (0,03, 0,08)	0,059 (0,04, 0,073)
Incremental Approximate-Fit-Indexes				

(*continued*)

	Total ($n = 472$)	Type I ($n = 144$)	Type (II $n = 13$)	Type (III $n = 315$)
Percentage of models for which at least one incremental fit index was reported	89,6	95,8	69	87,9
Percentage of use of NFI/BBI	20	16,6	23	21,7
Percentage of use of RFI	0,42	0,7	0,76	0
Percentage of use of CFI	80	86,8	53,8	78,34
Percentage of use of other incremental fit indexes	42,6	41	53,8	43
Magnitude of NFI/BBI[b] [95 / 23/ 3/ 69]	0,92 (0,90, 0,98)	0,93 (0,90, 0,95)	0,98 (0,94, 0,99)	0,92 (0,90, 0,98)
Magnitude of RFI[b] [1 /–/ 1/–]	0,97 (–, –)	– – –	0,97 (–, –)	– –-
Magnitude of CFI[b] [371/ 125/ 7/239]	0,945 (0,914, 0,978)	0,95 (0,92, 0,974)	0,99 (0,96, 0,99)	0,94 (0,91, 0,977)
Percentage of models for which at least one parsimonious fit index was reported	7,2	0,8	0	5
Percentage of use of AIC	3,8	0	7,7	3,2
Percentage of use of PNFI	1,5	–	–	2,2
Magnitude of AIC [12/–/1/11]	86,90 (-22, 103,75)	–	83,9 (–, –)	89,9 (-64,3,107,2)
Magnitude of PNFI [7/–/–/7]	0,53 (0,41, 0,71)	–	–	0,53 (0,41, 0,71)
Issues in assessing the measurement model				
Percentage of models for which at least some reliability information (composite reliability or extracted variance)	32	44	15.	27
Percentage of models for which at least one type of construct validity is analyzed – in parenthesis % of models analyzing a minimum of three types, nomological validity included.	54,6 (6,1)	61,1 (9,7)	23 (0)	53,2 (4,7)
Issues in assessing the latent variable model				
Percentage of models for which R^2 for structural equations was reported	26	4.	38	35
Model modification issues				
Percentage of models that were cross-validated	10	3	8	13
Type of sample for cross-validation – frequency total positive cases and % type on total where: [No information reported/ Split-sample procedure/ Independent sample]	45 [86,8/ 6,6/ 6,6]	4 [50/ 50/ 0]	1 [0/ 0/ 100]	40 [92,5/ 2,5/ 5]
Percentage of use of Expected Cross Validation Index	0,2	–	–	0,32
Magnitude [1/–/–/1]	2,88 (–, –)	–	–	2,88 (–, –)

Notes: [a] Models using χ^2 test as norm; [b] Table entries are medians (with the first and third quartiles in parentheses). Unless a percentage (or frequency of use) value is indicated

Table III.

models' fit evaluations and selections on global and incremental fit indices. This implies omitting the important premise in modelling of finding the right balance between models' fit and parsimony, of what it contributes to its generalizability (Preacher, 2006). Therefore, parsimony and goodness-of-fit indices should be conjointly applied when evaluating a model's degree of adjustment. Consequently, modellers should pay much more attention to this aspect of models' evaluation in the future.

4.6.2 Assessment of the measurement and structural model. The focus here is on measures of construct reliability, construct validity, and R^2 for each structural equation.

In marketing (and social science research), researchers routinely apply the Cronbach's alpha to assess the internal consistency of measures. However, this coefficient has diverse weaknesses (e.g. it is not a true index for assessing unidimensionality) that motivate the additional utilisation of more accurate construct reliability indices (see Miller, 1995; Raykov, 1998; Schmitt, 1996). The composite reliability (CR) and the average variance extracted (AVE) are the most adequate (Hair *et al.*, 2005) for this purpose. Nevertheless, these two measures are poorly applied in practice; just 26 per cent of the SEM applications in the database used, as a minimum, one of these two measures. So, based on this result, there should be a logical concern about the true reliability of constructs used in marketing models. In order to solve this problem, modellers need to be systematic and attend to the following questions: a complementary utilisation and interpretation of the coefficient alpha (cut-off ≥ 0.7 – Nunnally and Bernstein, 1994), altogether with the CR (cut-off ≥ 0.7 – Steenkamp and van Trijp, 1991); and an analysis of the indicators' loadings for every construct (cut-off ≥ > 0.5 – Fornell and Larcker, 1981).

Moreover, a more complete diagnosis of the validity of constructs must follow the reliability analyses. This implies diverse additional criteria (see Bagozzi, 1980), some referring to the analyses of semantic meanings (i.e. content or face validity), others involving empirical observations, such as convergent, discriminant and nomological validity. About half of the applications in our database show at least one type of validity in constructs, although only 6 per cent of models presented a detailed analysis, considering a minimum of three types of validity (nomological, the final criterion, included); full structural models are those where construct validity is more neglected. It is, therefore, necessary that researchers, especially when introducing new concepts, dedicate time to discuss the validity of constructs within their papers. In refering to the validity criteria with a purest base on empirical analyses, next we provide some useful rules. Regardless of other rigorous but more complex procedures based on the Campbell and Fiske' (1959) multitrait-multimethod approach, the existence of convergent validity could be admitted in scenarios of reliabilities of 0.8 or higher and AVE superior to 0.5 (see Ping, 2004); also, it would not be risky to assume discriminant validity when the square root the AVE for every construct is greater than the maximal correlation between said construct and the rest of constructs forming the model (see Chin, 1998b).

Finally, the application of the R^2 during the period analysed has not shown any significant amelioration with respect to the period 1977-1994. On the contrary, its application has even lessened in the Type III models, from 45 per cent (in Baumgartner and Homburg) to 35 per cent, in which it is even more crucial to know the reliability of every structural equation. This is a critical problem that should be radically improved, as R^2 provides relevant information on the reliability of equations; i.e. how well the endogenous elements of the model are explained by their predictor constructs.

4.7 Validation of models

Just 10 per cent of models were cross-validated, about half the percentage observed for the period 1977-1994. This reduction is similar or smaller when analysed by type. Especially remarkable is the drop for Type I, of which only 3 per cent of models were cross-validated. Again, it is necessary to reiterate that the issues related to the adjustment of the model are important, but modellers should not tackle them without paying attention to the generalisation of the proposed theoretical models. As a minimum, if modellers cannot work with two independent samples as recommended by Bagozzi and Yi (1988) to assure validity generalisation, they should work with large enough samples so as to split them into estimation and validation samples.

4.8 Reporting

Diverse reviews on SEM (Chin, 1998a; Chin *et al.*, 2008; MacCallum and Austin, 2000; Shook *et al.*, 2004) have suggested a variety of questions that studies should consider and report on in papers. These include a complete presentation and analysis of results, but other questions which are frequently ignored should also be addressed (Chin, 1998a) e.g. the applied input matrix, software package and version, distribution of the data, method of estimation and a graphical representation of the measurement model. If all or some of these questions were reported, readers would improve their understanding on the applied SEM process. In analyzing our database of SEM applications, we report that.

Input matrix. Almost 60 per cent gave information on the input matrix; in particular, 21 per cent used a covariance matrix and 38 per cent a correlation matrix. However, 41 per cent did not report this aspect of their work, and this is a failing since the type of matrix is influential in the result. In this regard, conventional estimation methods in SEM have usually been designed to work with a covariance matrix. Hence, treating a correlation matrix as a covariance matrix in the parameterisation stage of SEM is expected to introduce error in the parameter estimates (Cudeck, 1989), thereby producing unreliable results. The main problem here is how to deal with the kind of data that typically characterises the observed variables (i.e. indicators) of structural models in marketing, i.e. rating scales (e.g. Likert-type) that are ordinal, not continuous, in rigour (Jöreskog, 1990; Rigdon and Ferguson, 1991). From the mid-1990s, a few SEM software programs began to include options to properly treat this issue (MacCallum and Austin, 2000). Basically, instead of variance-covariance or product-moment (Pearson) correlation matrices, to assure more accurate (i.e. unbiased) parameter estimates, the use of a polychoric or polyserial (correlations) matrix is highly recommended (see, for greater detail, Babakus *et al.*, 1987; Jöreskog and Sörbom, 1988); although, this fact also has implications in respect of which estimation procedure to apply (see above). In any event, we have observed that modellers do not confirm, when using a correlation matrix as input, that they have followed the adequate procedures or selected certain SEM software to assure its correct treatment. This is a definite shortcoming in their reporting, which impedes a judgement about whether modellers worked with the proper matrix for their raw measures.

SEM Statistical Packages. LISREL, with an aggregated figure of 93 papers, is the one most used. This is logical if we consider that it is based on what is colloquially known as the "JKW model" (Jöreskog, Keesling and Wiley) in reference to those who first developed structural models with latent variables and measurement models

(Bentler, 1980; Bollen, 1989). However, despite its supremacy, if a longitudinal analysis is conducted (see Figure 2), it is seen that LISREL has gradually lost ground to others that appeared later. In this sense, it especially highlights the rise of EQS and, in the final period of years analysed, AMOS and PLS. These results contrast with the LISREL hegemony observed by Baumgartner and Homburg (1996) in 1977-1994. They suggested however, that LISREL was enjoying its 'first-mover advantage', and that new competitors could threaten its position. Our results seem to confirm this suspicion.

Methods of estimation. If model conditions were ideal, the diverse adjusting methods would be expected to perform similarly and their election would lose relevance (see Olsson *et al.*, 2000). For instance, the Maximum Likelihood (ML) method (applied in a 95 per cent of models in the period 1977-1994[5]) is very accurate when variables are continuous and normally distributed (Schumacker and Lomax, 2004). However, these theoretical conditions are habitually violated, so researchers should decide with rigour which method is the most appropriate.

In our sample of papers, a wide variety of methods have been analysed (see Table III). ML is first once again (14.8 per cent overall), followed by GLS (2.9 per cent). In any event, the most notable point to stress here is the considerable percentage of instances where the modellers have not explicitly reported the method used (81 per cent overall). Hence, it is difficult to reach any precise conclusion about the real application of these estimation methods in marketing structural modelling. This is not a trivial question as, depending on characteristics like type of measures, sample size, complexity of the measurement model, data distribution, etc., the application of a certain method of estimation may provide accurate or imprecise model estimates. In this regard, for instance, modellers usually find it difficult to work with a multivariate normality distribution (Micceri, 1989), although paradoxically, only 6.5 per cent of models reported analysis of the data distribution. A scenario of non-normality would discourage the use of multivariate normality-based methods – e.g. ML or Generalized Least Squares (GLS) – and makes other asymptotically distribution-free methods of estimation convenient – e.g. weighted least squares (WLS) or unweighted least squares (ULS) – (see Jöreskog, 1990).

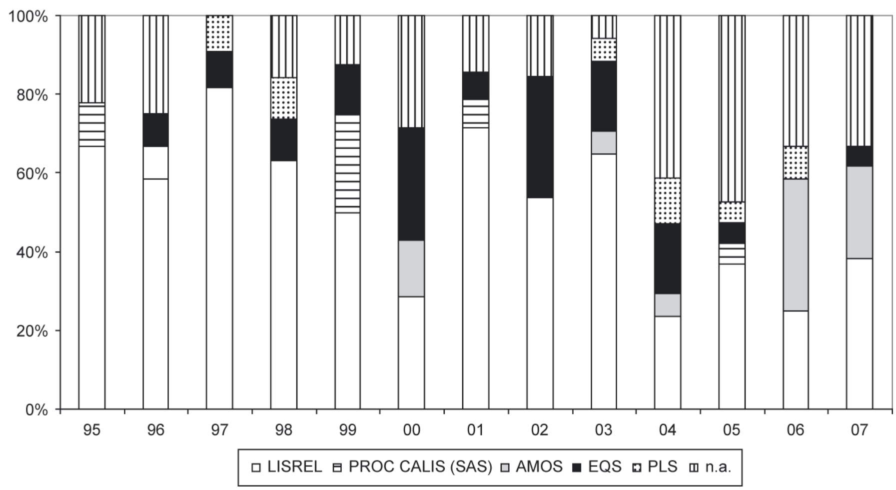

Figure 2.
Percentage of papers using certain SEM statistical packages per year

Moreover, as commented above, it is necessary that modellers be aware of the typical measures characterizing the observed variables (i.e. ordinal). This fact makes it more suitable to work with a polychoric/polyserial matrix, instead of the habitual variance-covariance or the product-moment correlation matrices; a conventional ML-based adjusting procedure, however, would not be appropriate to directly process a polychoric matrix either (Babakus *et al.*, 1987). With these model conditions, Rigdon and Ferguson (1991) have particularly recommended following a WLS fitting function as a primary option, because it provides parameter estimates that are unbiased, although, they also highlight that WLS might overestimate the statistics of model fit (see also, Babakus *et al.*, 1987). Nevertheless, recent studies highlight the strengths of using robust estimation methods to solve these shortcomings (see Lei, 2009). In particular, a robust WLS with ordinal scales has been demonstrated to provide proper solutions with parameter estimates, test statistics and standard errors, showing also a good tolerance to variations in model complexity and sample size (see Flora and Curran, 2004).

Picture of the measurement model. A total of 38 per cent (of SEM applications analysed provide graphical representation of the measurement model, although this proportion is 53.7 per cent for Type III models. Graphical representation of models helps the reader to obtain a quick view of the constructs, their measures and interconnections to be analysed. Hence, future improvement in this area would facilitate a first sight global understanding of papers.

5. Some insights into classic and recent model testing-related controversies

In spite of the utility of SEM as a tool for investigating and developing theory (Hayduk *et al.*, 2007), diverse weaknesses have attracted criticism over the years, calling into question its true potential. Such comments have emerged from academics within several disciplines (e.g. statistics, psychology, management, marketing, etc.) during the last decades (see, as, e.g. Anderson and Gerbing, 1988; Bagozzi and Yi, 1988; Barret, 2007; Chin, 1998a; Fornell, 1983; Hu and Bentler, 1999; Kelloway, 1995; Marsh *et al.*, 2004; MacCallum *et al.*, 1996; Sivo *et al.*, 2006). Now, there is a good opportunity to compile and review these controversies, incorporating recent advances in respect of the issues raised. For this reason, regardless of the general view already offered in Section 4, where some punctual controversies have been introduced, now the most significant model testing-related controversies are discussed[6].

5.1 Controversies around the chi-square test

5.1.1 Interpretation. Let us structure this discussion in two scenarios, i.e. non-significant versus significant test. Researchers routinely apply the test to see if "models fit" the data (Barret, 2007). But, how should a non-significant result be interpreted? This is the kind of result desired by modellers, though it is not a conclusive result at all simply to say that a "model is accepted"; however, this scenario has been found in almost 22 per cent of non-significant tests in our database of SEM applications, which is a worrying proportion. Likewise, authors frequently interpret a non-significant result as a sign of model fit. As Steiger (2007) recently highlights, this is also a wrong conclusion to reach, which corresponds with the classic "accept-support" fallacy. Rigorously speaking, a non-significant result only means that one cannot reject the idea that a covariance matrix associated with the hypothesised model is equivalent to the true model (Raykov and Penev, 1999). So, it would not imply that a model is correct, but merely that such model is one of the potential different causal models which are consistent with the data (Bullock *et al.*, 1994; Hayduk *et al.*, 2007). Notwithstanding, our results show that in just 11 per cent of non-significant cases, there was an explicit comment made by the authors on the plausible existence of equivalent models.

On the other hand, modellers may have to deal with a significant test. This eventual result means rejecting the null hypothesis of the test, which indicates a bad fit of the model; this situation represents about 50 per cent of the models using this test in the database, which is a considerable proportion. However, though this result suggests a rejection of the model, authors usually ignore it or try to look for some justification in the limitations of this test, and/or report acceptable values of some goodness-of-fit indices (Bagozzi and Yi, 1988). These wrong practices are clearly observed in our data of SEM applications in marketing; researchers completely ignore it (this means making no commentaries at all to justify this result) in 55.4 per cent of significant cases; or try to justify it arguing values over broadly accepted cut-offs for some AFIs, in 24.5 per cent of models with significant chi-square tests. It is also widely known that the traditional chi-square test of exact fit presents several shortcomings which reduce its reliability (e.g. violations of multivariate normal distribution of data, sample size and model complexity, among others) (Chou and Bentler, 1995; Chou *et al.*, 1991; Satorra and Bentler, 1990; Schermelleh-Engel *et al.*, 2003). This fact is difficult to resolve in practice, and some SEM experts (see Jöreskog and Sörbom, 1993) have suggested not using it as a formal test, but as a descriptive index of fit (i.e. χ^2/d.f.). In any event, our data reveal not only how normal it is to work with or to ignore significant results without scientific rigour, but also that the tests' significance are inaccurately reported (see, also: Markland, 2007); however, chi-square statistic, degrees of freedom and significance of the test should always be reported.

Nevertheless, this apparent generalised bad praxis should be readdressed in order to be more rigorous. An interesting protocol is proposed in Barret (2007, p. 821), with the following stages:

N: Examine the distribution of data. If they are not normally distributed, the chi-square test may be inflated, so it is recommended to do a transformation of data, rescale or even exclude variables in order to work with data which are respectful of the assumptions of the estimation methods. However, just 11.3 per cent of SEM applications with significant results in our database analyse the multivariate-normality of data, although, most of such applications look for solutions in the way suggested by Barret (2007).

On the contrary, if data are normally distributed, researchers should inform their readers that the model fails, and should subsequently discuss the implications for the subjacent theoretical framework on which the model is based.

However, researchers could also explore the residual matrix (this is a more complex process) and search for clues to detect the seeds of misfit, amend it and then refit it again. Based on our results, these are practices not generally followed by marketing researchers. Firstly, just a minority of applications (about 7 per cent) explicitly checks normality in the distribution of variables and, second, in the unusual cases where multivariate-normality is positively checked, none really proceeds to report that the model fails and then halts the analysis of results.

But, researchers may decide to ignore the test, which is one of the usual behaviours observed in our database (see above). In this case, they have to clearly justify why.

At this point, Barret (2007) points out an important warning against the use of typical excuses such as the sensitivity of the test to diverse factors (e.g. sample size or violations in data distribution) or the alternative use of goodness-of-fit indices to counteract a bad result of the test. On the contrary, more rigorous alternative actions are recommended, where analyses to determine the cross-validated predictive accuracy and parsimony (via AIC indices) of the model are highlighted. These are alternatives which are particularly necessary to follow when SEM is applied with theoretical purposes (Hayduk *et al.*, 2007). Notwithstanding, such alternatives are marginally used in marketing, as can be observed from our analyses; i.e. just 10 per cent of applications are cross-validated and about 7 per cent analyse the models' parsimony (less than 4 per cent work with the AIC).

5.1.2 Exact fit versus close fit. The chi-square test has been classically applied with an approach of "exact fit", meaning that no significant differences between covariance matrices are allowed. However, Steiger (2007), for instance, questions its direct value on two grounds:

(1) a perfect fit hypothesis is irrelevant, as models are restrictive, so the probability of perfect fit is practically zero; and

(2) even when testing results are non-significant, it is not necessarily good evidence of the model fit, as it could be due to a lack of ability to detect model misspecification (Chin, 1998a).

This is a problem of power analysis, particularly worrying for the approach of exact fit, where it is needed (Millsap, 2007). A few authors have analysed this issue in SEM in detail (e.g. Saris and Satorra, 1993; Satorra and Saris, 1985), and in this respect the contribution by MacCallum *et al.* (1996) is worth considering. They argue against the exact fit approach and advocate the logic of close-fit in SEM to finally propose a test of non-close fit, based on the RMSEA index, especially designed to work with power analysis in testing discrepancies of matrices. However, although this option is available in some SEM software packages from the nineties, its use has not been evidenced in our sample of SEM applications.

Finally, the tables (in function of sample size and degrees of freedom) proposed by MacCallum *et al.* (1996) to achieve a power analysis of 0.8 have been applied to the database of models. The results are clear in this respect, as more than 30 per cent of SEM applications do not present enough power for the chi-square test, regardless of the alternative (exact versus close-fit) which is followed. This fact plays against the reliability of the test in these cases.

5.1.3 Other tests of fit. The chi-square test, though traditionally emphasised as the only true test of overall fit in SEM –mainly when distinguishing it from the goodness-of-fit indices– is not the only test available. Recently, Bentler (2007) observed that the variety of possible model tests has been expanded, so nowadays it would not be correct to talk about "the" chi-square test; for instance, the EQS 6 provides about a dozen alternatives, which are interesting because of their potential to overcome some of the shortcomings of the traditional chi-square test, commented on above; additionally, they reduce the strong dependence on just one test, which is one of the main limitations of SEM (see Tomarken and Waller, 2003). Of these alternatives, the Satorra-Bentler scaled chi-square is more robust with smaller sample sizes or violations in the multivariate normality assumptions (Kunnan, 1998; McIntosh, 2007), but its use has only been explicitly declared in 1.2 per cent of applications.

5.2 Approximate/goodness-of-fit indices

5.2.1 Interpretation of AFIs with an acceptance-rejection approach. Although AFIs were created to provide an alternative source of information to the chi-square test, researchers have increasingly interpreted them with a hypothesis-testing approach (Marsh *et al.*, 2004). In other words, based on certain threshold values of reference, it is habitually concluded whether or not a model presents an adequate fit; it is also usual to use the results of AFIs to explicitly or implicitly justify a significant result of the chi-square test (24.5 per cent in our database of SEM-based models). However, as already discussed, this kind of conclusion is even questioned when a statistical test (generally the chi-square test) is used. Therefore, researchers should not use broadly accepted "rules of thumb" for AFIs with this "acceptable/non-acceptable" spirit when evaluating the model fit. On the contrary, AFIs must be regarded as helpful relative measures of model fit/misfit (Yuan, 2005), but not as indices to support conclusions on it.

5.2.2 The utility of threshold values. From the 1980s until now, there have been several attempts to provide SEM users with cut-off values of reference to evaluate the fit of models. The first such study was that of Bentler and Bonett (1980), although probably the more rigorous and influential study is that authored by Hu and Bentler (1999); here, researchers evaluate the adequacy of the "rule of thumb" conventional threshold values and propose new alternatives for other fit indices. The final result is a set of recommended cut-off values for several AFIs, broadly adopted by SEM users. However, diverse rigorous studies published in recent years have warned against two main problematic issues (Beauducel and Wittman, 2005; Fan and Sivo, 2005; Marsh *et al.*, 2004; Sivo *et al.*, 2006: Yuan, 2005): the application of such "rules of thumb" as if they were "golden rules" to accept or reject models, generically commented on above; and/or the use of threshold values with a "universal" approach, regardless of data and the conditions of the models.

Undoubtedly, the aim of Hu and Bentler (1999) was not to fix values for AFIs acting as "golden rules", though their application by SEM users may have distorted their truest purpose in that particular incorrect direction. In this regard, Markland (2007) currently advocates normal adoption of the Hu and Bentler (1999) cut-off criteria, as they are more stringent than previous recommendations, and always useful to support substantive interpretation of the data. Likewise, later studies which partially replicated that of Hu and Bentler (1999) also recognised the utility of their cut-off values in evaluating the degree of models' fit/misfit (see: Marsh *et al.*, 2004; Yuan, 2005). So, it would not be irresponsible for researchers to continue using such cut-off values as a reference.

However, the influence of factors such as the model complexity (e.g. number of indicators and constructs), or the impact of the sample size in the appropriate threshold values for AFIs should not be ignored (Chen *et al.*, 2008; Sharma *et al.*, 2005). Hence, a contingent approach regarding the use of threshold cut-offs to assess model fit is recommended, avoiding, whenever possible, universal cut-off values, in order to be more accurate in the assessment of model fit. In this respect, one of the most recent and rigorous replicas of the Hu and Bentler (1999) study is that by Sivo *et al.* (2006). Here, the authors demonstrate by Monte Carlo simulation that sample size is the most influential factor in determining the adequate threshold values for AFIs; other factors, such as the complexity[7] of models and type of models of data distribution were found not to be influential. In particular, with small sample sizes cut-off values should be relaxed, though they should be increased as sample sizes get higher. Besides, especially interesting is a proposal of optimal index values, without rejecting any correct model (i.e. no Type I error), for 13 AFIs used with six sample sizes. In Table IV, we conjointly show the results of applying both the traditional approach of universal cut-offs and a contingent approach of cut-offs in function of the SEM-based models' sample size proposed by Sivo *et al.* (2006); the set of selected AFIs responds to all those from our database with observations in the diversity of sample size ranges considered here, although these AFIs practically coincide also with those most frequently used by modellers. It can be observed how the contingent approach tends to admit a higher proportion of correct models than the traditional approach in a scenario of small samples, when cut-offs are relaxed, though this is reversed as sample size ranges increase.

A final reflection: both the chi-square test and AFIs are very useful for assessing how a model fits the data, although modellers must consider other questions when accurately evaluating the results offered by the SEM process in relation to a certain theoretical model of reference (Jöreskog, 1993). It is also the responsibility of researchers going further, to analyse such questions related to the measurement and structural models as (Bagozzi and Yi, 1988; Barret, 2007): the estimated parameters, the predictive accuracy (R^2) of the model, the reliability of constructs, etc. Notwithstanding, based on our data, while the analyses and interpretation of structural parameters is generalised, marketing researchers pay more attention to assessing model fit than to these ulterior and fundamental questions (see sections 4.6 and 4.7).

6. Final remarks and practical recommendations

This study has provided a modern and refreshed view on the use of SEM as a research tool to aid the development and generation of knowledge in scientific disciplines, paying special attention to its application within the marketing discipline. That said, many of the themes treated have benefited from a general and eclectic approach, regardless of particular references to our discipline as a way to accurately illustrate them.

SEM is a powerful method for theory testing which has reached a stage of maturity, in terms of its adoption and application, in marketing. In this paper, the significant supporting role of SEM in the generation of new knowledge in marketing theory has been highlighted. Nowadays, in research scenarios where theories are increasingly more complex, with more elements and interconnections to consider, SEM is one of the most preferred methods by marketing academics. It should, therefore, not be a matter of debate that SEM ought to be properly applied in order to assure a good quality and reliable process of theory development. Here, based on an extensive and miscellaneous literature review, a plethora of traditional, renewed and recent controversies around the use of SEM has been presented. All of these issues have been empirically supported with specific data in marketing, thereby allowing for improved vision and diagnosis within this discipline.

Table IV.
Comparison of methods in the assessment of models based on AFI's: universal versus contingent cut-off values

	Traditional approach to evaluate AFIs		New protocol, based on the study of Sivo *et al.* (2006), discriminating the cut-off values of AFIs by sample size of SEM applications % of models, for every AFI and particular sample restrictions (Optimal cut-off values without rejecting any correct model –Type I error = 0– in parenthesis) Sample size					
AFIs	Cut-off[a]	% models satisfying cut-off	≤150	151-250	251-500	501-1000	1000-2500	>2500
GFI	0,95	24,4	55,5 (0,89)	27,45 (0,93)	3,9 (0,96)	0,04 (0,98)	16,6 (0,99)	0 (0,99)
Valid *n*		197	34	49	76	24	6	9
AGFI	0,95	19,1	53,3 (0,87)	19 (0,91)	15,6 (0,95)	0 (0,97)	25 (0,99)	0 (0,99)
Valid *n*		89	15	21	31	17	4	1
RMSEA	0,06	47,3	45,45 (0,06)	40,8 (0,05)	6,3 (0,03)	3,8 (0,03)	33 (0,02)	0 (0,01)
Valid *n*		306	53	70	108	51	9	15
RMSR	0,08	68,4	76,9 (0,12)	72,1 (0,10)	44 (0,07)	23,1 (0,05)	33 (0,03)	0 (0,03)
Valid *n*		125	24	41	23	24	6	7
CFI	0,95	40,47	52 (0,95)	32,3 (0,97)	13,6 (0,98)	1,7 (0,99)	18,2 (0,99)	0 (0,99)
Valid *n*		371	72	97	116	58	11	17
NFI	0,95	33,7	66 (0,88)	69 (0,92)	30 (0,96)	0 (0,99)	50 (0,99)	0 (0,99)
Valid *n*		95	9	29	30	12	2	13

Note: [a] Conventional/Universal cut-off values for: GFI, AGFI and NFI are taken from Schumacker and Lomax (2004), RMSEA, RMSR and CFI are taken from Hu and Bentler (1999)

Table V. Brief reminders of practical recommendations based on the problematic issues identified and discussed in SEM

Issue in SEM	Description of the problem identified	Practical recommendations
Type of SEM-based models	Although a residual representation of Type II models has been observed, these are obsolete due to the fact that they do not follow the predominant measurement philosophy in marketing, which implies working with several measures (items) per construct (as Type I and III models do)	Avoid using Type II models in SEM-based studies
SEM-based modelling strategies	Excessive use of strictly confirmatory studies. This is a modelling strategy which matches the "classic vision" of the scientific research process, although, it is not so coherent with the current dynamic orientation in marketing modelling based on the implementation of models	Higher use of competitive modelling and generating model strategies. This fact may imply lengthier manuscripts, so journals' editors and members of editorial boards are recommended to address this issue, allowing authors extra words if the paper requires it
	Competitive modelling and model development are modelling strategies where the study of models' parsimony is very necessary. However, the analysis of parsimony is far from being routinely analyzed in these cases	Always make analyses of parsimony when opting for any of these two modelling strategies. Indices as PNFI, PGFI or AIC are recommended
	Models obtained after following a model generating strategy are not usually re-specified according to theoretical considerations, but changes are data-driven	Regardless of the assistance provided by statistical analyses, re-specification of models should be always supported by underlying theory
Issues related to the specification of models	Although a high proportion of SEM-models following a two-stage approach is observed, some models, albeit in a small proportion, still do not differentiate the assessment of the measurement model (e.g. analysis of validity of constructs) and structural model	A generalized adoption of a two-step approach is suggested. However, in those cases where the presentation of these two stages may require too much space in papers, modellers are recommended to briefly present the issues related with validity, but also to make a detailed presentation of the issues related to the structural model

(*continued*)

Table V.

Issue in SEM	Description of the problem identified	Practical recommendations
	There are still many single-indicator constructs, although this problem has been significantly improved when compared with results obtained by previous reviews on SEM in marketing	In general, follow the predominant philosophy of multi-item (a minimum of three for convention) measurement for every construct, so avoiding the inclusion of single-indicator constructs. However, exceptionally, for the case of doubly concrete constructs, the use of single-indicator constructs can be methodologically accepted, if well justified
Issues related to sample size	Again, the old problem of high proportions of SEM applications with samples that are not big enough to obtain reliable parameters and valid testing is identified	Make aprioristic estimations of the number of free parameters to finally work with adequate sample sizes
Assessment of model fit		
Chi-square test	This test presents a high proportion of cases where its use, interpretation and reporting is inadequate	Follow the protocols, based on Barret (2007), described in section 5.1.1
	Inadequate sample size to achieve enough power analysis	Base on the aprioristic estimation of free parameters to calculate the expected degrees of freedom, using the formula: 0,5[(p + q)(p + q + 1)] − t, where "p" = exogenous indicators, "q" = endogenous indicators and "t" is the number of estimated coefficients in the model. Then, with that information, search in tables proposed by MacCallum *et al.* (1996) for the minimum sample size necessary
	There is high dependence of the chi-square test, in spite of the fact that it has several limitations and scenarios where it fails	Use alternative tests in situations where the classic chi-square test fails, e.g. violations of multivariate normality of data. One of the plausible solutions is the Satorra-Bentler scaled test. Other alternative tests, however, may require working with certain SEM software packages

(*continued*)

Issue in SEM	Description of the problem identified	Practical recommendations
Approximate Fit Indices	The use of AFIs with an acceptance-rejection approach	Regard AFIs as helpful relative measures of model fit/misfit, but not as indices to extract conclusions on it
	Delimitation of threshold values of AFIs with a traditional approach. Recently, several studies have demonstrated that using universal cut-off for AFIs, without considering the influence of other factors as sample size or the specification of model in such cut-off value, is inaccurate to assess the model fit/misfit	Apply a contingent approach when working with the threshold of AFIs. In this regard, it is recommended using the cut-off values discriminated by sample size for AFIs, as provided by Sivo *et al.* (2006), who concluded sample size as the truest influential factor. In Table IV, an adaptation has been done for six AFIs
	Some AFIs had been demonstrated to be less reliable than others with regard to variations in sample sizes	If the modeller still decides to apply a traditional approach with cut-off values, avoid using, or use secondarily, indices like GFI or AGFI and intensify the application of more reliable indices, e.g. RMSEA and CFI
Assessment of measurement and structural models		
Measurement model	Low application of analyses of constructs' reliability through composite reliability and/or the indicators' loadings for every construct	These measures of reliability based on the estimated model parameters are better than traditional measures based on Cronbach's alpha. So, modellers should routinely report and complementarily analyze the coefficient alpha (cut-off $\geq$ 0.7), altogether with the CR (cut-off $\geq$ 0.7), and an analysis of the indicators' loadings (cut-off $\geq$ 0.5)

(*continued*)

Table V.

Issue in SEM	Description of the problem identified	Practical recommendations
	Scarce analysis of constructs' validity in detail	Modellers have to treat, even synthetically, in papers several types of validity: content or face (this implies a theoretical analysis); convergent (e.g. reliabilities $\geq$ 0.8 and AVE $>$ 0.5); discriminant (e.g. AVE $>$ maximal correlation between said construct and the other constructs); and, ideally, nomological
Structural model	Few reports of the R^2 of structural equations. This scenario is even worse for the case of Type III, full structural models. This is a problem already identified by previous reviews which has shown no improvement in the last years	Routinely report and analyze R^2 of structural equations, especially in Type III models
Validation of models	Cross-validation of models is necessary, although the poor proportion of cross-validated models observed in this study, also when compared with previous reviews, suggests that this is a chronic problem of SEM-based studies in marketing. Also, as indicated above, this is a procedure especially necessary to follow when using model generating strategies.	Models should be generally cross-validated. This is especially necessary, to avoid "capitalization on chance", for model generating strategies, where it is recommended to regularly cross-validate models with and independent sample
Reporting	Diverse SEM-related questions are not routinely nor properly reported. This lack of information makes it more difficult for the reader to achieve an adequate understanding on the full SEM process applied	Apart from the variety of questions already recommended, some others usually ignored should be also reported and properly justified. In particular: the applied input matrix, software package and version, distribution of the data, method of estimation and a graphical representation of the measurement model

Table V.

In essence, despite its limitations (e.g. number of journals analysed, presentation of contents with a "low technical charge", etc.), the study does demonstrate, with a deep theoretical discussion and accurate empirical support, a basic idea: the current application of SEM in marketing has much room for improvement; a conclusion extendable with no risk to other social sciences disciplines. Considering the role of this method in the generation of new knowledge for the theory of marketing, it is very convenient that researchers, referees and readers of SEM-based studies pay attention to the themes discussed here. Numerous current issues related with the application of SEM to marketing research have been identified, treated and provided with a solution. Hopefully, the study will be a useful guide for researchers wanting to adequately apply SEM and, above all, generate reliable and solid new knowledge for marketing theory; also, by extension, such knowledge may be applied to other business areas, and even more distant scientific disciplines, which use SEM to test their theoretical proposals. Finally, for the reader's convenience, in Table V the problematic issues detected in this study together with their corresponding practical recommendations are synthesized; the content structure is that defined in Section 4.

Notes

1. The authors thank one of the reviewers for this suggestion.
2. In any event, if modellers prefer working with some kind of aggregated measures to reduce the number of indicators per construct, latent variable models based on item-parcelling seem a more convenient intermediate alternative than path analysis models with total scale scores (Coffman and MacCallum, 2005).
3. The authors thank one of the reviewers for this suggestion.
4. Readers interested in a detailed and extensive discussion of pros and cons of multiple- vs. single-item scales can consult this article.
5. Its predominance was attributed to the default option of most SEM software packages, as well as to its normality assumption in the distribution of the data (see Baumgartner and Homburg, 1996)
6. The authors are aware of the interest of other key controversial themes (e.g. the more abstract, even metaphysical debate on the causal nature of SEM models), but these have been consciously omitted to avoid an excessively long manuscript.
7. A study by Sharma *et al.* (2005) also concludes the special influence of sample size in cut-off values. However, it was also observed that such values are also affected by an interaction between the sample size and the total number of indicators for certain AFIs (i.e. GFI, RNI, TLI and NNCP).

References

Ahearne, M., Hughes, D.E. and Schillewaert, N. (2007), "Why sales reps should welcome information technology: measuring the impact of CRM-based IT on sales effectiveness", *International Journal of Research in Marketing*, Vol. 24 No. 4, pp. 336-49.

Aigner, D.J. and Goldberger, A.S. (Eds) (1977), *Latent Variables in Socio-economic Models*, North-Holland, Amsterdam.

Anderson, J.C. and Gerbing, D.W. (1988), "Structural equation modelling in practice: a review and recommended two-step approach", *Psychological Bulletin*, Vol. 103 No. 3, pp. 411-23.

Babakus, E., Ferguson, C.E. and Joreskog, K.G. (1987), "The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions", *Journal of Marketing Research*, Vol. 24 No. 2, pp. 222-8.

Babin, B.J., Hair, J.F. and Boles, J.S. (2008), "Publishing research in marketing journals using structural equation modelling", *Journal of Marketing Theory and Practice*, Vol. 16 No. 4, pp. 279-85.

Bagozzi, R.P. (1980), *Causal Models in Marketing*, Wiley, New York, NY.

Bagozzi, R.P. (1984), "A prospectus for theory construction in marketing", *Journal of Marketing*, Vol. 48, pp. 11-29.

Bagozzi, R.P. (1994), "Measurement in marketing research", in Bagozzi, R.P. (Ed.), *Principles of Marketing Research*, Blackwell, Cambridge, MA, pp. 1-49.

Bagozzi, R.P. and Edwards, J.R. (1998), "A general approach for representing constructs in organizational research", *Organizational Research Methods*, Vol. 1 No. 1, pp. 45-87.

Bagozzi, R.P. and Yi, Y. (1988), "On the evaluations of structural equations models", *Journal of the Academy of Marketing Science*, Vol. 16 No. 1, pp. 74-94.

Bandalos, D. and Finney, J.S. (2001), "Item parceling issues in structural equation modeling", in Marcoulides, A.G. and Schumacker, E.R. (Eds), *New Developments and Techniques in Structural Equation Modeling*, Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 269-95.

Barret, P. (2007), "Structural equation modelling: adjudging model fit", *Personality and Individual Differences*, Vol. 42, pp. 815-24.

Baumgartner, H. and Homburg, C. (1996), "Applications of structural equations modelling in marketing and consumer research: a review", *International Journal of Research in Marketing*, Vol. 13, pp. 139-61.

Beauducel, A. and Wittman, W. (2005), "Simulation study on fit indices in confirmatory factor analysis based on data with slightly distorted simple structure", *Structural Equation Modeling*, Vol. 12, pp. 41-75.

Bengtson, V.L., Acock, A.C., Allen, K.R., Dilworth-Anderson, P. and Klein, D. (2005), *Source Book of Family Theory and Research*, Sage Publications, Thousand Oaks, CA.

Bentler, P.M. (1980), "Multivariate analysis with latent variables: causal modelling", *Annual Review of Psychology*, Vol. 31, pp. 419-56.

Bentler, P.M. (1995), *EQS Structural Equations Program Manual*, Multivariate Software, Encino, CA.

Bentler, P.M. (2007), "On tests and indices for evaluating structural models", *Personality and Individual Differences*, Vol. 42, pp. 825-9.

Bentler, P.M. and Bonett, D.G. (1980), "Significance tests and goodness-of-fit in the analysis of covariance structures", *Psychological Bulletin*, Vol. 88, pp. 588-606.

Bentler, P.M. and Chou, C.-P. (1987), "Practical issues in SEM", *Sociological Methods and Research*, Vol. 16, pp. 78-117.

Bergkvist, L. and Rossiter, J.R. (2007), "The predictive validity of multiple-item versus single-item measures of the same constructs", *Journal of Marketing Research*, Vol. 44 No. 2, pp. 175-84.

Blalock, H.M. (1971), "Causal models involving unmeasured variables in stimulus-response situations", in Blalock, H.M. (Ed.), *Causal Models in the Social Sciences*, Aldine-Atherton, Chicago, IL, pp. 335-47.

Bollen, K.A. (1989), *Structural Equations with Latent Variables*, Wiley Interscience, New York, NY.

Boomsma, A. and Hoogland, J.J. (2001), "The robustness of LISREL modeling revisited", in Cudeck, R., du Toit, S. and Sörbom, D. (Eds), *Structural Equation Modeling: Present and Future*, Scientific Software International, Inc, Chicago, IL, pp. 139-68.

Brannick, M.T. (1995), "Critical comments on applying covariance structure modeling introduction", *Journal of Organizational Behavior*, Vol. 16, pp. 201-13.

Breckler, S.J. (1990), "Applications of covariance structure modeling in psychology: cause for concern?", *Psychological Bulletin*, Vol. 107, pp. 260-73.

Bullock, H.E., Harlow, L.L. and Mulaik, S.A. (1994), "Causation issues in structural equation modeling research", *Structural Equation Modeling*, Vol. 1 No. 3, pp. 253-67.

Bunge, M. (1967), *Scientific Research: The Search for Information*, Springer, New York, NY.

Calantone, R.J., Schmidt, J. and Song, X.M. (1996), "Controllable factors of new product success: a cross-national comparison", *Marketing Science*, Vol. 15 No. 4, pp. 341-58.

Campbell, D.T. and Fiske, D.W. (1959), "Convergent and discriminant validation by the multitrait-multimethod matrix", *Psychological Bulletin*, Vol. 56, pp. 81-105.

Cavusgil, S.T. and Zou, S. (1994), "Marketing strategy-performance relationship: an investigation of the empirical link in export market ventures", *Journal of Marketing*, Vol. 58 No. 1, pp. 1-21.

Chaudhuri, A. and Holbrook, M.B. (2001), "The chain of effects from brand trust and brand affect to brand performance: the role of brand loyalty", *Journal of Marketing*, Vol. 65 No. 2, pp. 81-93.

Chen, F., Curran, P.J., Bollen, K.A., Kirby, J. and Paxton, P. (2008), "An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models", *Sociological Methods Research*, Vol. 36 No. 4, pp. 462-94.

Chin, W.W. (1998a), "Issues and opinion on structural equation modelling", *MIS Quarterly*, Vol. 22 No. 1, pp. vii-xvi.

Chin, W.W. (1998b), "The partial least squares approach to structural equation modelling", in Marcoulides, G.A. (Ed.), *Modern Methods in Business Research*, Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 295-336.

Chin, W.W., Peterson, R.A. and Brown, S.P. (2008), "Structural equation modelling in marketing: some practical reminders", *Journal of Marketing Theory and Practice*, Vol. 16 No. 4, pp. 287-98.

Chou, C.-P. and Bentler, P. (1995), "Estimates and tests in SEM", in Hoyle, R. (Ed.), *Structural Equation Modeling*, Sage Publications, Thousand Oaks, CA, pp. 37-55.

Chou, C.P., Bentler, P. and Satorra, A. (1991), "Scaled test statistics and robust standard errors for non-normal data in covariance structure analysis: a Monte Carlo study", *British Journal of Mathematical and Statistical Psychology*, Vol. 44, pp. 345-57.

Churchill, G.A. (1979), "A paradigm for developing better measures of marketing constructs", *Journal of Marketing Research*, Vol. 16 No. 1, pp. 64-73.

Coffman, D.L. and MacCallum, R.C. (2005), "Using parcels to convert path analysis models into latent variable models", *Multivariate Behavioral Research*, Vol. 40 No. 2, pp. 235-59.

Cudeck, R. (1989), "Analysis of correlation matrices using covariance structure models", *Psychological Bulletin*, Vol. 105, pp. 317-27.

Ding, L., Velicer, W.F. and Harlow, L.L. (1995), "Effects of estimation methods, number of indicators per factor, and improper solutions on SEM fit indices", *Structural Equation Modeling*, Vol. 2, pp. 119-43.

Fan, X. and Sivo, S.A. (2005), "Sensitivity of fit indices to misspecified structural or measurement model components: rationale of two-index strategy revisited", *Structural Equation Modeling*, Vol. 12, pp. 343-67.

Fan, X., Thompson, B. and Wang, L. (1999), "Effects of sample size, estimation method, and model specification on structural equation modeling fit indexes", *Structural Equation Modeling*, Vol. 6, pp. 56-83.

Flora, D.B. and Curran, P.J. (2004), "An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data", *Psychological Methods*, Vol. 9 No. 4, pp. 466-91.

Fornell, C. (1983), "Issues in the application of covariance structure analysis: a comment", *Journal of Consumer Research*, Vol. 9, pp. 443-8.

Fornell, C. and Larcker, D.F. (1981), "Evaluating structural equation models with unobservable variables and measurement error", *Journal of Marketing Research*, Vol. 18 No. 1, pp. 39-50.

Garver, M.S. and Mentzer, J.T. (1999), "Logistics research methods: employing structural equation modeling to test for construct validity", *Journal of Business Logistics*, Vol. 20 No. 1, pp. 33-58.

Goldberger, A.S. and Duncan, O.D. (Eds) (1973), *Structural Equation Models in the Social Sciences*, Academic Press, New York, NY.

Hair, J.F., Black, W.C., Babin, B., Anderson, R.E. and Tatham, R.L. (2005), *Multivariate Data Analysis*, 6th ed., Prentice Hall, Englewood Cliffs, NJ.

Hall, R.J., Snell, A.F. and Foust, M.S. (1999), "Item parceling strategies in SEM: investigating the subtle effects of unmodeled secondary constructs", *Organizational Research Methods*, Vol. 2 No. 3, pp. 233-56.

Hayduk, L., Cummings, G., Boadu, K., Pazderka-Robinson, H. and Boulianne, S. (2007), "Testing! Testing! One, two, three – testing the theory in structural equation models!", *Personality and Individual Differences*, Vol. 42, pp. 841-50.

Hofacker, C.F., Gleim, M.R. and Lawson, S.L. (2009), "Revealed reader preference for marketing journals", *Journal of the Academy of Marketing Science*, Vol. 37 No. 2, pp. 238-47.

Homburg, C., Grozdanovik, M. and Klarmann, M. (2007), "Responsiveness to customers and competitors: the role of affective and cognitive organizational systems", *Journal of Marketing*, Vol. 71 No. 3, pp. 18-38.

Homer, P.M. and Yoon, S.-G. (1992), "Message framing and the interrelationships among ad-based feelings, affect and cognition", *Journal of Advertising*, Vol. 21 No. 1, pp. 19-33.

Hoogland, J.J. and Boomsma, A. (1998), "Robustness studies in covariance structure modeling: an overview and a meta-analysis", *Sociological Methods and Research*, Vol. 26, pp. 329-67.

Hoyle, R.H. (Ed.) (1995), *Structural Equation Modeling*, Sage Publications, Thousand Oaks, CA.

Hu, L. and Bentler, P.M. (1999), "Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives", *Structural Equation Modeling*, Vol. 6 No. 1, pp. 1-55.

Hughes, M.A., Price, R.L. and Marrs, D.W. (1986), "Linking theory construction and theory testing: models with multiple indicators of latent variables", *Academy of Management Review*, Vol. 11 No. 1, pp. 128-44.

Hulland, J., Chow, Y.H. and Lam, S. (1996), "Use of causal models in marketing research: a review", *International Journal of Research in Marketing*, Vol. 13 No. 2, pp. 181-97.

James, L.R. and James, L.A. (1989), "Causal modelling in organizational research", in Cooper, C. and Robertson, I. (Eds), *International Review of Industrial and Organizational Psychology*, Wiley, Chichester, pp. 371-404.

Jöreskog, K.G. (1990), "New developments in LISREL: analysis of ordinal variables using polychoric correlations and weighted least squares", *Quality and Quantity*, Vol. 24, pp. 387-404.

Jöreskog, K.C. (1993), "Testing structural equations models", in Bollen, K.A. and Long, J.S. (Eds), *Testing Structural Equation Models*, Sage Publications, Newbury Park, CA, pp. 294-316.

Jöreskog, K.G. and Sörbom, D. (1988), *PRELIS: A Program for Multivariate Data Screening and Data Summarization. A Preprocessor for LISREL*, 2nd ed., Scientific Software, Inc, Mooresville, IN.

Jöreskog, K.G. and Sörbom, D. (1993), *LISREL VIII: Structural Equation Modelling with the SIMPLIS Command Language*, Scientific Software, Inc, Hillsdale, NJ.

Jöreskog, K.G. and Wold, H. (1982), *Systems under Indirect Observation, Parts I and II*, North-Holland, Amsterdam.

Kelloway, E.K. (1995), "Structural equation modelling in perspective", *Journal of Organizational Behavior*, Vol. 16, pp. 215-24.

Kempf, D.S. and Smith, R.E. (1998), "Consumer processing of product trial and the influence of prior advertising: a structural modeling approach", *Journal of Marketing Research*, Vol. 35 No. 3, pp. 325-38.

Kishton, J.M. and Widaman, K.F. (1994), "Unidimensional versus domain representative parceling of questionnaire items: an empirical example", *Educational and Psychological Measurement*, Vol. 54 No. 3, pp. 757-65.

Kline, R. (2005), *Principles and Practice of Structural Equation Modeling*, 2nd ed., Guilford, New York, NY.

Kunnan, A.J. (1998), "An introduction to structural equation modelling for language assessment research", *Language Testing*, Vol. 15 No. 3, pp. 295-332.

Lee, T. and Calantone, R.J. (1998), "The impact of market knowledge competence on new product advantage: conceptualization and empirical examination", *Journal of Marketing*, Vol. 62 No. 3, pp. 13-29.

Leeflang, P.S.H., Wittink, D.R., Wedel, M. and Naert, P.A. (2000), *Building Models for Marketing Decisions*, Kluwer Academic, Dordrecht.

Lei, P.-W. (2009), "Evaluating estimation methods for ordinal data in structural equation modeling", *Quality & Quantity*, Vol. 4 No. 3, pp. 495-507.

Lilien, G.L. and Kotler, P. (1990), *Toma de decisiones en mercadotecnia. Un enfoque a la construcción de modelos*, Compañía Editorial Continental, Mexico City.

Lilien, G.L., Kotler, P. and Moorthy, K.S. (1992), *Marketing Models*, Prentice Hall International, Hemel Hempstead.

Little, T.D., Cunningham, W.A., Shahar, G. and Widaman, K.F. (2002), "To parcel or not to parcel: exploring the question, weighing the merits", *Structural Equation Modeling*, Vol. 9 No. 2, pp. 151-73.

Loehlin, J.C. (1998), *Latent Variable Models: An Introduction to Factor, Path, and Structural Analysis*, Lawrence Erlbaum Associates, Mahwah, NJ.

McDonald, R.P. (1989), "An index of goodness-of-fit based on noncentrality", *Journal of Classification*, Vol. 6, pp. 97-103.

McIntosh, C.N. (2007), "Rethinking fit assessment in structural equation modelling: a commentary and elaboration on Barrett (2007)", *Personality and Individual Differences*, Vol. 42, pp. 859-67.

MacCallum, R.C. (1986), "Specification searches in covariance structure analysis", *Psychological Bulletin*, Vol. 100, pp. 107-20.

MacCallum, R.C. and Austin, J.T. (2000), "Applications of structural equation modeling in psychological research", *Annual Review of Psychology*, Vol. 51, pp. 201-26.

MacCallum, R.C., Browne, M.W. and Sugawara, H.M. (1996), "Power analysis and determination of sample size for convariance structure modelling", *Psychological Methods*, Vol. 1, pp. 130-49.

MacCallum, R.C., Widaman, K.F., Zhang, S. and Hong, S. (1999), "Sample size in factor analysis", *Psychological Methods*, Vol. 4 No. 1, pp. 84-99.

Markland, D. (2007), "The golden rule is that there are no golden rules: a commentary on Paul Barrett's recommendations for reporting model fit in structural equation modelling", *Personality and Individual Differences*, Vol. 42, pp. 851-8.

Marsh, H.W., Hau, K.T. and Wen, Z. (2004), "In search of golden rules: comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings", *Structural Equation Modeling*, Vol. 11 No. 3, pp. 320-41.

Marsh, H.W., Hau, K.T., Balla, J.R. and Grayson, D. (1998), "Is more ever too much? The number of indicators per factor in confirmatory factor analysis", *Multivariate Behavioral Research*, Vol. 33 No. 2, pp. 181-220.

Meade, A.W. and Kroustalis, C.M. (2006), "Problems with item parcelling for confirmatory factor analytic tests of measurement invariance", *Organizational Research Method*, Vol. 9 No. 3, pp. 369-403.

Medsker, G.M., Williams, L.J. and Holohan, P. (1994), "A review of current practices for evaluating causal models in organizational behaviour and human resources management research", *Journal of Management*, Vol. 20, pp. 439-64.

Micceri, T. (1989), "The unicorn, the normal curve, and other improbable creatures", *Psychological Bulletin*, Vol. 105 No. 1, pp. 156-66.

Miller, M.B. (1995), "Coefficient alpha: introduction from the perspectives of classical test theory and structural equation modelling", *Structural Equation Modeling*, Vol. 2 No. 3, pp. 255-73.

Millsap, R.E. (2007), "Structural equation modeling made difficult", *Personality and Individual Differences*, Vol. 42, pp. 875-81.

Naert, P. and Leeflang, P. (1978), *Building Implementable Marketing Models*, Martinus Nijhoff, Leiden.

Nasser, F. and Takahashi, T. (2003), "The effect of using item parcels on ad hoc goodness-of-fit indexes in confirmatory factor analysis: an example using Sarason's reactions to tests", *Applied Measurement in Education*, Vol. 16 No. 1, pp. 75-97.

Nunnally, J. and Bernstein, I. (1994), *Psychometric Theory*, 3rd ed., McGraw-Hill, New York, NY.

Olsson, U.H., Foss, T., Troye, S.V. and Howell, R.D. (2000), "The performance of ML, GLS, and WLS estimation in structural equation modeling under conditions of misspecification and nonnormality", *Structural Equation Modeling*, Vol. 7 No. 4, pp. 557-95.

Peter, P.J. (1979), "Reliability: a review of psychometric basics and recent marketing practices", *Journal of Marketing Research*, Vol. 16 No. 1, pp. 6-17.

Ping, R.A. (2004), "On assuring valid measures for theoretical models using survey data", *Journal of Business Research*, Vol. 57 No. 2, pp. 125-41.

Preacher, K.J. (2006), "Quantifying parsimony in structural equation modelling", *Multivariate Behavioural Research*, Vol. 41 No. 3, pp. 227-59.

Price, L.L., Arnould, E.J. and Tierney, P. (1995), "Going to extremes: managing service encounters and assessing provider performance", *Journal of Marketing*, Vol. 59, April, pp. 83-97.

Rapp, A., Ahearne, M., Mathieu, J. and Schillewaert, N. (2006), "The impact of knowledge and empowerment on working smart and working hard: the moderating role of experience", *International Journal of Research in Marketing*, Vol. 23 No. 3, pp. 279-93.

Raykov, T. (1998), "Coefficient alpha and composite reliability with interrelated nonhomogeneous items", *Applied Psychological Measurement*, Vol. 22 No. 4, pp. 375-85.

Raykov, T. and Penev, S. (1999), "On structural equation model equivalence", *Multivariate Behavioral Research*, Vol. 34, pp. 199-244.

Rigdon, E.E. and Ferguson, C.E. (1991), "The performance of the polychoric correlation coefficient and selected fitting functions in confirmatory factor analysis with ordinal data", *Journal of Marketing Research*, Vol. 28, pp. 491-7.

Rossiter, J.R. (2002), "The C-OAR-SE procedure for scale development in marketing", *International Journal of Research in Marketing*, Vol. 19 No. 4, pp. 305-35.

Sandvik, I.L. and Sandvik, K. (2003), "The impact of market orientation on product innovativeness and business performance", *International Journal of Research in Marketing*, Vol. 20 No. 4, pp. 355-76.

Saris, W.E. and Satorra, A. (1993), "Power evaluations in structural equation models", in Bollen, K.A. and Long, J.S. (Eds), *Testing Structural Equation Models*, Sage Publications, Thousand Oaks, CA, pp. 181-204.

Satorra, A. and Bentler, P.M. (1990), "Model conditions for asymptotic robustness in the analysis of linear relations", *Computational Statistics and Data Analysis*, Vol. 10, pp. 235-49.

Satorra, A. and Saris, W.E. (1985), "The power of the likelihood ratio test in covariance structure analysis", *Psychometrika*, Vol. 50, pp. 83-90.

Schermelleh-Engel, K., Moosbrugger, H. and Müller, H. (2003), "Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures", *Methods of Psychological Research Online*, Vol. 8 No. 2, pp. 23-74.

Schmitt, N. (1996), "Uses and abuses of coefficient alpha", *Psychological Assessment*, Vol. 8 No. 4, pp. 250-353.

Schumacker, R.E. and Lomax, R.G. (2004), *A Beginner's Guide to Structural Equation Modeling*, 2nd ed., Lawrence Erlbaum Associates, Mahwah, NJ.

Sharma, S., Mukherjee, S., Kumar, A. and Dillon, W. (2005), "A simulation study to investigate the use of cutoff values for assessing model fit in covariance structure models", *Journal of Business Research*, Vol. 58 No. 7, pp. 863-1008.

Shook, C.L., Ketchen, D.J., Hult, G.T.M. and Kacmar, K.M. (2004), "An assessment of the use of structural equation modelling in strategic management research", *Strategic Management Journal*, Vol. 25, pp. 397-404.

Sivo, S.A., Fan, X., Witta, E.L. and Willse, J.T. (2006), "The search for 'optimal' cutoff properties: fit index criteria in structural equation modelling", *Journal of Experimental Education*, Vol. 74 No. 3, pp. 267-88.

Steenkamp, J.B. and van Trijp, H. (1991), "The use of LISREL in validating marketing constructs", *International Journal of Research in Marketing*, Vol. 8 No. 4, pp. 283-99.

Steenkamp, J.-B.E.M. and Baumgartner, H. (2000), "On the use of structural equation models for marketing modelling", *International Journal of Research in Marketing*, Vol. 17, Special Issue, pp. 195-202.

Steiger, J.H. (2007), "Understanding the limitations of global fit assessment in structural equation modelling", *Personality and Individual Differences*, Vol. 42, pp. 893-8.

Stremersch, S. and Lehmann, D.R. (2008), "25 years of *IJRM*: reflections on the past and future", *International Journal of Research in Marketing*, Vol. 25 No. 3, pp. 143-8.

Tomarken, A.J. and Waller, N.G. (2003), "Potential problems with 'well fitting' models", *Journal of Abnormal Psychology*, Vol. 112 No. 4, pp. 578-98.

Tomarken, A.J. and Waller, N.G. (2005), "Structural equation modelling: strengths, limitations, and misconceptions", *Annual Review of Clinical Psychology*, Vol. 1, pp. 31-65.

Wierenga, B. and Van Bruggen, G.T. (2000), *Marketing Management Support Systems: Principles, Tools and Implementation*, Kluwer Academic, Dordrecht.

Yuan, K.H. (2005), "Fit indices versus test statistics", *Multivariate Behavioral Research*, Vol. 40, pp. 115-48.

Yuan, K.H., Bentler, P.M. and Kano, Y. (1997), "On average variables in a confirmatory factor analysis model", *Behaviormetrika*, Vol. 24 No. 1, pp. 71-83.

Further reading

West, S.G., Finch, J.F. and Curran, P.J. (1995), "Structural equation modeling with nonnormal variables: problems and remedies", in Hoyle, R.H. (Ed.), *Structural Equation Modeling: Concepts, Issues, and Applications*, Sage Publications, Thousand Oaks, CA, pp. 56-75.

Corresponding author

Francisco J. Martínez-López can be contacted at: fjmlopez@ugr.es