Vocino, Andrea 2009, The effects of clustered data on standard error estimates in covariance structure analysis : a field data application, *Asia Pacific journal of marketing and logistics*, vol. 21, no. 1, pp. 7-18.

# The effects of clustered data on standard error estimates in covariance structure analysis: A field data application

## A field data application

### The Authors

**Andrea Vocino,** ***School of Management and Marketing, Deakin University, Burwood, Australia***

### Acknowledgements

The author would like to thank Harmen Oppewal (Monash University), Laura Stapleton (University of Maryland, Baltimore County), Lynd Bacon (Loma Buena Associates) and Wagner Kamakura (Duke University) for the valuable comments made on an earlier draft of the manuscript which was presented at the 37th European Marketing Academy Conference in Brighton, UK in May 2008. Special thanks to Allison Ringer (Deakin University) for providing editorial support in the preparation of the article. The author also benefited from consulting with the SEMNet (the Structural Equation Modeling Discussion Network) community and in particular with Edward Rigdon (Georgia State University), Cameron McIntosh (Statistics Canada) and Stas Kolenikov (University of Missouri, Columbia) to whom the author extends his sincere gratitude and appreciation for their continuous support. However, any omissions or errors in the article remain the sole responsibility of the author.

### Abstract

**Purpose** – The purpose of this article is to present an empirical analysis of complex sample data with regard to the biasing effect of non-independence of observations on standard error parameter estimates. Using field data structured in the form of repeated measurements it is to be shown, in a two-factor confirmatory factor analysis model, how the bias in SE can be derived when the non-independence is ignored.

**Design/methodology/approach** – Three estimation procedures are compared: normal asymptotic theory (maximum likelihood); non-parametric standard error estimation (naïve bootstrap); and sandwich (robust covariance matrix) estimation (pseudo-maximum likelihood).

**Findings** – The study reveals that, when using either normal asymptotic theory or non-parametric standard error estimation, the SE bias produced by the non-independence of observations can be noteworthy.

**Research limitations/implications** – Considering the methodological constraints in employing field data, the three analyses examined must be interpreted independently and as a result taxonomic generalisations are limited. However, the study still provides "case study" evidence suggesting the existence of the relationship between non-independence of observations and standard error bias estimates.

**Originality/value** – Given the increasing popularity of structural equation models in the social sciences and in particular in the marketing discipline, the paper provides a theoretical and practical insight into how to treat repeated measures and clustered data in general, adding to previous methodological research. Some conclusions and suggestions for researchers who make use of partial least squares modelling are also drawn.

## Keyword(s):

Structural analysis; Modelling; Samples; Surveys.

### Introduction and research background

There has been a growing attention in recent years in analysing models fitting data collected from either longitudinal or more general surveys involving repeated measures that use complex sample designs (CSDs). A CSD typically refers to data acquired by stratification (often on the basis of geography), cluster sampling and/or sampling with an unequal probability selection (Kish, 1965; Lee and Forthofer, 2005; Skinner *et al.*, 1989). Complex survey data are also referred to as multi-level or hierarchical data. Such designs, however, render the statistical analysis more complicated since the observations are not independent and identically distributed (IID).

The conventional methods for estimating SEs included in most statistical software packages rely on a simple random samples (SRS) design (Lee and Forthofer, 2005; Skinner *et al.*, 1989). While independence of sample constituents is normally assumed, it is seldom realised in the procedures of practical survey work (Kish and Frankel, 1974). The assumption of independent observations, while not entirely essential for the estimation of parameters (e.g. path coefficients), is central for the SEs of those parameters (Frenkel and Frenkel, 1977; Lee and Forthofer, 2005; Skinner *et al.*, 1989). Since conventional estimation of SEs assumes that the correlation of the errors across individuals is zero, a researcher using clustered data (e.g. data in the form of repeated measures) may underestimate the SE. Failure to use the appropriate statistical analysis leads, therefore, to an increased probability of committing Type I errors (erroneously rejecting the null hypothesis), underestimation of SEs and misleadingly positive test results (Kish and Frankel, 1974; Scariano and Davenport, 1987; Stapleton, 2006).

In a Monte Carlo simulation study, Julian (2001) reported that covariance models that do not account for the multi-level structure of the data present estimation problems in the $\chi^2$ statistic, parameter estimates and SEs (when the intraclass correlations are $>0.05$). The author states that, in such a case, the effect of disregarding the data dependence can no longer be overlooked. Furthermore, when the group/member ratio decreases, the consequence of not accounting for the multi-level data structure will become more severe on the quality of estimation (Julian, 2001). However, other researchers such as Shackman (2001) and Maas and Hox (2005) revealed that in multi-level modelling the concern is not so much the intraclass correlation $\rho$ (or rho), but the design effect (deff) which Shackman classified as a correction that should be used to determine sample size (Kish, 1965).

The Monte Carlo method is “an empirical method for evaluating statistics” (Paxton *et al.*, 2001, p. 289) and facilitates the investigator to examine the performance of a given test statistic across a number of random samples (Rubenstein, 1981). However, Monte Carlo studies have also been the object of various criticisms since they may lack strong theoretical underpinning guiding the design and analysis of the simulation and because of their questionable external validity and generalisability. Without strong theoretical grounding, simulation studies are often considered “to be akin to randomly looking for a needle in a haystack” (Paxton *et al.*, 2001, p. 290).

In this article, an analysis of field data is used to investigate whether the theories proposed by Julian (2001), Shackman (2001) and Maas and Hox (2005) hold in an applied field such as marketing. As far as the writer knows, there is no empirical research in the realm of marketing that has investigated the extent of the problems resulting from the application of nonhierarchical covariance structure modelling techniques which assume IID to complex sample data. It is argued that it is an empirical question as to whether or not interdependence between multi-level measures needs to be accommodated in models.

Conventional wisdom in covariance structure analysis suggests that when data are non-normally distributed, three different possible strategies are available to accommodate such an anomaly: (a) asymptotically distribution free[1] estimation (b) Satorra-Bentler scaled $\chi^2$ with robust standard errors and (c) non-parametric naïve bootstrapping. Most times, however, the use of one or the other procedure is dictated by the functionalities of the software package employed by the researcher.

Of particular interest to this study is the bootstrapping procedure since relative to normal asymptotic estimation for SEs, bootstrapping has the advantage of calculating distribution-free estimates of SEs and/or empirical confidence intervals for statistics having unknown sampling distributions (e.g. Diaconis and Efron, 1983; Efron, 1979; Efron and Tibshirani, 1993). In non-parametric (or even called naïve) bootstrapping, the sample at hand is treated as a pseudo-population. Numerous units are drawn with replacement from the available dataset in order to engender a number of new datasets (bootstrap samples) that typically have the same number of units to the original. Sampling with replacement implies that in any new generated dataset, the same unit can materialise more than once and some units may not even be included. However, the structure of units will vary fairly across the created bootstrap samples. When iterated manifold times, non-parametric bootstrapping simulates the drawing of many samples from a population. SE are then computed as the SD of the separate estimates across all the new created samples.

Indeed, it could be argued that when dealing with repeated measures, the use of naïve bootstrapping would be a suitable solution to breaking down the intraclass correlation providing perhaps more appropriate SE estimates than the ones computed using normal asymptotic theory. Notwithstanding, Hox and Mass (2001) contend that a complication may arise in using such a procedure since the naïve bootstrap sampling process should reflect the multi-level structure of the data. In other words, as in this case, the researcher should resample individuals (e.g. clusters of subjects) and then resample repeated measures (e.g. brand observations) within individuals.

Thus, the current study offers an evaluation of three estimation procedures when observations are not IID. Using a two-factor confirmatory factor analysis (CFA) model, this article provides an analysis of: (a) normal asymptotic theory (maximum likelihood), (b) non-parametric SE estimation (naïve bootstrap) and (c) sandwich (robust covariance matrix) estimation (pseudo-maximum likelihood).

To render the examination relatively simple and easy to interpret, using CFA, the psychometric properties and association of only two constructs were analysed. Cross-sectional data were used in the form of repeated measurements of the same respondents across different brand ratings on a set of items. The items reflected two dimensions (i.e. brand investments and brand quality) that are part of the brand signalling framework proposed by Erdem and Swait (1998). Since the research question at hand is not hierarchical or multi-level in nature, it was employed "aggregated analysis of complex survey" data as suggested by Stapleton (2006), (also Muthén and Satorra, 1995), benefiting from the analysis of complex designs built in LISREL 8.8[2] (Jöreskog and Sörbom, 2006). The consequence for disregarding the dependencies among observations intrinsic to the multi-level data for their estimated SEs was appraised. In so doing, the analysis offered in this study is diverse from the one proposed by Julian (2001) for three reasons: (1) Julian (2001) used synthetic computer generated data while this study uses real field data, (2) Julian (2001) used a disaggregated analysis [with a multi trait multi method model], whereas this study proposes an aggregated analysis of complex survey data and (3) in the current study, the naïve bootstrapping estimation procedure is also included, whereas such an analysis was not offered by Julian (2001).

The findings suggest that when using either normal asymptotic theory or non-parametric SE estimation, the SE bias produced by the nonindependence of observations can be extensive.

**Complex designs and clustering**

Repeated measures occur in many areas of research. It is not uncommon either in commercial or academic research to have repeated measures as a result of "stacking up" observations from different survey cycles. In the marketing discipline and particularly in the brand literature (e.g. Aaker, 1997; Dillon *et al.*, 2001; Erdem and Swait, 1998, 2004; Erdem *et al.*, 2006) respondents are frequently asked to rate a set of items across different brands. In cross-sectional data, repeated measures can be analysed as a two-level structure where measurement occurrences are level-1 units and respondents are level-2 units. Hence, these repeated observations can be regarded as additional clusters within the larger design. However, for a set sample size, cluster designs are subject to larger SEs. Because the calculation of the SEs entails the sample size in the denominator of the equation:(see equation 1)when the observations are not independent, the effective sample size is not the mere number of cases in the dataset and an adjustment needs to be made in order to avoid the SEs being biased downwards. The discrepancy in the accuracy of the estimates generated by a CSD in relation to a SRS is known as the deff (Skinner *et al.*, 1989). In samples characterised by clusters, the deff is the fraction of the actual variance, under the sampling method used, to the variance computed according to the postulation of SRS (Muthén and Satorra, 1995) and can be expressed as follows:(see equation 2)where deff is the design effect, $\rho$ is the intraclass correlation for the variable under consideration and $m$ is the cluster size. The intraclass correlation $\rho$ can be expressed as a simple case of variance decomposition. It is comparable to a one-way analysis of variance (ANOVA) with random effects. In order to determine $\rho$, the outcome variability at each of the levels of the hierarchy needs to be established. As this study uses repeated measures in the form of brand ratings on a set of items, let the subscript $i$ refer to the $i$th level-2 unit, in this instance, the $i$th respondent (who has completed the questionnaire). The subscript $j$ refers to the rating of $j$th brand observation (cf. one of the brand names observed in a specific item) for the *i*th respondent. Using this notation, the one-way ANOVA model can be written as:(see equation 3)where $y_{ij}$ denotes the *j*th brand observation (within a particular item) for respondent *i*, $\beta_0$ represents the intercept of the fixed part of the model, $u_{0i}$ denotes the random variation in intercepts of level-2 of the model and $\varepsilon_{ij}$ represents the random variation of level-1 of the model. It is assumed that $u_{ij}$ has a value of 0 and variance of $\sigma_b^2$ . The variance $\sigma_b^2$ may be described as the "between-group" variability. At the same time, it is assumed that $\varepsilon_{ij}$ is $N(0,\sigma_w^2)$ distributed. Thus, $\sigma_w^2$ may be described as the "within-group" variability. Such a model is also described as a fully unconditional model (Bryk and Raudenbush, 1992) as there are no predictors specified at either level of the hierarchy. ρ exhibits the degree of correspondence within each group and Shrout and Fleiss (1979, p. 423) expressed it as follows:(see equation 4)Hence, ρ indicates the fraction of total variability $\sigma_b^2+\sigma_w^2$ that can be attributed to the variability within cluster group $\sigma_w^2$. When data are independent, ρ coefficients ought to be equal to zero. The scale of ρ hinges on the relation between the characteristics of the groups (i.e. structure, size and function) and the variable measured (Julian, 2001; also Kish, 1965).

## Application to signalling data

This study follows previous work (Erdem and Swait, 1998) in signalling theory which developed an information economics perspective on the value (or equity) attributed to brands by consumers. Signalling theory explains how high-quality firms can differentiate themselves from the lower quality ones. Based on this theory, it has been argued that brand signalling inferences emerge from dissipative signals (Rao *et al.*, 1999), which derive from an *ex ante* expenditure comprising investment in building a reputation that could be lost, should the promised product quality not correspond to the actual quality delivered (Erdem and Swait, 1998). A brand incorporates and represents a firm's past and present marketing mix activities and brand investments (Erdem and Swait, 1998). Firms spend resources on their brands to guarantee that promises are maintained. In addition, firms make brand investments to exhibit commitment to their brands (Klein and Leffler, 1981). Brand investments contribute to credibility by signalling. When a branded product fails to fulfil the promise expressed in the brand signal, the brand compromises the expected returns on these brand investments as well as its reputation for delivering on its promises (Erdem and Swait, 1998). The literature suggests that investments in brand strategies and activities (e.g. brand logo, sponsorship or an influential advertising campaign) are sunk costs that cannot be recuperated (Ippolito, 1990). If the credibility of the brand is compromised the firm cannot command the premium associated with its reputation and brand investment (Erdem and Swait, 1998).

Using CFA, the model examined the relationship between brand investment and brand quality. Both constructs are reflective latent variables in the brand signalling framework proposed by Erdem and Swait (1998). Brand ratings obtained from 239 members of an online panel in Australia were used in order to assess the two dimensions of the signalling framework in the jeans (121 surveys) and digital cameras (118 surveys) product categories. Each respondent rated eight brands; therefore, there were a total of 1,912 individual observations at the brand level (968 for jeans and 944 for digital cameras). Scales similar to the ones employed by Erdem and Swait (1998, 2004) and Erdem *et al.* (2006) were used. Consistent with Erdem and Swait (1998, 2004) and Erdem *et al.* (2006) all items were measured on 9 point Likert scales. Following the same methodology used in Erdem and Swait (1998) and Erdem *et al.* (2006), data were stacked in one single matrix in order to produce a pooled model across different brands and the two product categories. The data were balanced, that is, there were the same number of brand observations per respondent. The data lent themselves to be modelled via complex sample analysis since they were structured in the form of repeated measurements of the same respondents across different brand ratings on a set of items.

**Data analysis**

First, it was assessed the two factor CFA pooled model making use of LISERL 8.8 (Jöreskog and Sörbom, 2006), employing design base adjustment of the likelihood ratio test (LRT) statistic which automatically rescales $\chi^2$ value and estimates "robust" SE (for a detailed estimation procedure, Scientific Software International Inc., 2005). All $\lambda$ and $\varphi$ were left free to vary/covary, while the variance of each latent variable was set to 1.

The model converged very well resulting in $\chi^2 = 4.36$, df = 4, $p = 0.35945$, root mean square error of approximation (RMSEA) = 0.007 (90 per cent confidence interval for RMSEA = 0.000 – 0.0358 and $p$-value for test of close fit RMSEA < 0.05 = 0.997) with standardised $\varphi_{invest,qual} = 0.88$ ($p = 0.000$). All the $\lambda$s were statistically significant ($p = 0.000$). The average variance explained was >0.5 and the composite reliability (CR) was >0.7 suggesting that convergent validity was achieved (e.g. Dillon and Goldstein, 1984; Fornell and Larker, 1981). On the other hand, discriminant validity was assessed by constraining the correlation parameter between the two latent variables to 1. If the $\Delta\chi^2$ between the constrained and unconstrained models is statistically significant, it is likely that the correlation for the two latent variables is indeed not 1 (cf. Anderson and Gerbing, 1988; Bagozzi *et al.*, 1991). However, because the rescaled $\chi^2$ does not follow a normal $\chi^2$ distribution the $\Delta\chi^2$ between the nested models was estimated following the correction given by Satorra and Bentler (2001). The estimation procedure is reported in the Appendix.

The $\Delta\chi^2$ resulted in a value of 13.528 for 1 df ($p = 0.000$) suggesting that even in this instance the test was positive, indicating that discriminant validity was achieved. A good fitting model allowed the study to progress to the next phase of the research. In order to observe the effects of $\rho$ and deff on the SEs estimation, this study considered three respondent/repeated measures configurations for the multi-level data: (a) 239 respondents with eight brand observations each, (b) 239 respondents with five brand observations each and (c) 239 respondents with three brand observations each. Two of the above configurations ((b) and (c)) were achieved by subdividing the dataset into two further data subsets. The number and nature of brands chosen for this experiment were selected randomly from the available dataset and they were consistent across all respondents (cf. all the brand names were constant across individuals). The number of brand observations per individual was manipulated with the expectation of having an incremental/decremental effect on $\rho$. On the other hand, as stated previously, it is known that such a manipulation is, *ceteris paribus*, likely to influence deff estimates since the number of brand observations per individual is a function of deff itself (equation 2). However, the above manipulation resulted in unequal sample sizes in the three configurations as in (a) $n = 1{,}912$, in (b) $n = 1{,}195$ and in (c) $n = 717$.

Notably, it could be argued that the three models may not comparable because the variability from one data subset to another dictates that the parameter estimates will vary according to what data are observed as well as the sample sizes. Yield estimates from a particular model depend on the parameter estimates, and so the yield estimates themselves are variable. However, because field data are used, one cannot specify a priori parameter values for these models to produce the desired $\rho$ (as one would normally do in a Monte Carlo simulation). As an alternative solution, this study presents the analysis of nine models whereby, following the three configurations proposed above, the SEs of three models in (a), (b) and (c) obtained from normal asymptotic theory (maximum likelihood estimation) and via naïve bootstrapping [in this case AMOS 7 (Arbuckle, 2006) was employed in order to produce 2,000 bootstrap samples] are compared with their respective models calculated with the "sandwich" estimator (pseudo maximum likelihood estimation as portrayed in Pfeffermann *et al.*, 1998) to generate robust estimates using LISREL 8.8 (Jöreskog and Sörbom, 2006). In summary, this study evaluates the SEs of:

- model (a) with 239 respondents with eight brand observations each obtained from normal asymptotic theory and model ($a^{'}$) which used naïve bootstrapping vs. model ($a^{''}$) which utilised the sandwich estimator;
- model (b) having 239 respondents with five brand observations each obtained from normal asymptotic theory with model ($b^{'}$) which used naïve bootstrapping vs model ($b^{''}$) which utilised the sandwich estimator;
- model (c) having 239 respondents with three brand observations each obtained from normal asymptotic theory with model ($c^{'}$) which used naïve bootstrapping vs model ($c^{''}$) which utilised the sandwich estimator.

Since the SEs of the models ($^{''}$) are properly specified and "robust", the SE estimates of these are taken as baselines to be compared with the estimates (and consequent misspecifications) of the models computed with normal asymptotic theory and naïve bootstrap. The percentages $\Delta$ in SEs between the three models represents the SE misspecifications given by the SRS design.

**Findings**

Table I shows $\rho$, deff and SEs of the nine models for the $\lambda$ the $\psi$ as well as the $\varphi_{invest,qual}$ parameter estimates $\rho_8$, $deff_8$ and $SEs_8$ refer to both models (a, $a^{'}$ and $a^{''}$) with eight repeated measures $\rho_5$, $deff_5$ and $SEs_5$ to models (b, $b^{'}$ and $b^{''}$) with five repeated measures and $\rho_3$, $deff_3$ and $SEs_3$ to models (c, $c^{'}$ and $c^{''}$) with 3 repeated measures per respondent.

Bias per cent measures the bias of the asymptotic (ML) estimation vs the sandwich estimation and the bias of the bootstrap estimation vs the sandwich estimation which are calculated as follows:(see equation 5)The findings suggest a relationship between the deff and the SEs bias. If, for instance, one observes the deff coefficients in all the $\lambda$ estimations, it can be noted that in all the parameters $deff_8 > deff_5 > deff_3$, respectively, $bias_8 > bias_5 > bias_3$ for both asymptotic and naïve bootstrapping estimation, suggesting the larger the deff the larger the bias. It is also interesting to observe that relative to $\varphi_{invest,qual}$, $bias_8 > bias_5 > bias_3$. However, it is somewhat surprising to note that there is little difference in $\rho$ relative to the number of observations per individual. In fact, relative to the $\lambda$ and $\psi$ parameters, $\rho_3 \cong \rho_5 \cong \rho_8$.

**Conclusion**

One conclusion that can be drawn from this study is that taxonomic outcomes based on relationships between $\rho$, deff and the SEs misspecifications cannot be rigorously derived. Admittedly, the biggest limitation of this research is that the three case studies examined therein must be interpreted independently since, in order to manipulate $\rho$ and deff, three different data subsets were used. However, considering the methodological constraints in employing field data and the impossibility to link the three independent analyses, the study still provids some "case study" evidence that might suggest the existence of such a relationship. Using real data, this article confirms the findings of Julian (2001), Shackman (2001) and Maas and Hox (2005), that is: (a) when $\rho > 0.05$ the multi-level structure of the data cannot be disregarded in estimating a proper SE and (b) as previously suggested there would seem to be a relationship between the deff and SE estimates misspecifications.

A further important finding reported in this article, as pointed out by Hox and Mass (2001), is that when data are clustered, even non-parametric SE estimations do not seem to help. While the bias produced by naïve bootstrapping is smaller than the one produced by normal asymptotic theory estimation, it is worth noting that such a bias is still considerable. This is particularly relevant for researchers who use, for instance, partial least squares modelling (Wold, 1985), which employs nonparametric SE estimation. As of this writing, the writer is not aware of any partial least squares software package that has the ability to undertake CSD computations and as a result researchers who are dealing with clustered or more general complex sample data would need to write a specific program to perform such estimations.

**Notes**

1. Different variants of robust weighted least squares procedures are implemented in different software packages.
2. All the complex design estimations have also been reconciled using M*plus* 5.1 (Muthén and Muthén, 2008).

| | $\lambda_{qual\_02}$ | $\lambda_{qual\_03}$ | $\lambda_{qual\_04}$ | $\lambda_{invest\_01}$ | $\lambda_{invest\_02}$ | $\psi_{qual\_02}$ | $\psi_{qual\_03}$ | $\psi_{qual\_04}$ | $\psi_{invest\_01}$ | $\psi_{invest\_02}$ | $\phi_{invest,qual}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\rho_8$ | 0.223 | 0.243 | 0.287 | 0.264 | 0.354 | 0.523 | 0.542 | 0.583 | 0.563 | 0.641 | |
| $deff_8$ | 2.558 | 2.700 | 3.008 | 2.850 | 3.477 | 4.659 | 4.797 | 5.084 | 4.939 | 5.486 | |
| $SE_8$ | | | | | | | | | | | |
| Asymptotic (ML) | 0.038 | 0.038 | 0.039 | 0.052 | 0.046 | 0.041 | 0.039 | 0.048 | 0.127 | 0.103 | 0.013 |
| Bootstrap | 0.034 | 0.033 | 0.039 | 0.054 | 0.045 | 0.064 | 0.054 | 0.086 | 0.143 | 0.118 | 0.017 |
| Sandwich estimation (PML) | 0.064 | 0.065 | 0.078 | 0.093 | 0.083 | 0.128 | 0.092 | 0.174 | 0.257 | 0.187 | 0.029 |
| $bias_8$ % [asymptotic] | −39.94% | −41.55% | −50.38% | −44.10% | −44.44% | −68.20% | −57.92% | −72.64% | −50.58% | −44.92% | −55.17% |
| $bias_8$ % [bootstrap] | −46.38% | −48.37% | −50.51% | −42.60% | −45.77% | −49.92% | −41.43% | −50.80% | −44.40% | −37.01% | −41.38% |
| $\rho_5$ | 0.188 | 0.211 | 0.261 | 0.241 | 0.308 | 0.470 | 0.489 | 0.529 | 0.513 | 0.567 | |
| $deff_5$ | 1.753 | 1.843 | 2.043 | 1.965 | 2.234 | 2.880 | 2.955 | 3.116 | 3.054 | 3.266 | |
| $SE_5$ | | | | | | | | | | | |
| Asymptotic (ML) | 0.048 | 0.048 | 0.049 | 0.064 | 0.057 | 0.050 | 0.047 | 0.057 | 0.150 | 0.119 | 0.015 |
| Bootstrap | 0.045 | 0.043 | 0.048 | 0.064 | 0.056 | 0.083 | 0.067 | 0.103 | 0.176 | 0.137 | 0.021 |
| Sandwich estimator (PML) | 0.068 | 0.069 | 0.081 | 0.099 | 0.087 | 0.131 | 0.097 | 0.169 | 0.266 | 0.190 | 0.026 |
| $bias_5$ % [asymptotic] | −28.40% | −30.36% | −39.33% | −34.72% | −34.56% | −62.21% | −51.50% | −66.39% | −43.61% | −37.37% | −42.31% |
| $bias_5$ % [bootstrap] | −33.43% | −36.64% | −40.69% | −35.23% | −35.60% | −36.41% | −30.88% | −39.29% | −33.91% | −27.74% | −21.15% |
| $\rho_3$ | 0.228 | 0.268 | 0.355 | 0.218 | 0.401 | 0.487 | 0.511 | 0.564 | 0.481 | 0.593 | |
| $deff_3$ | 1.456 | 1.536 | 1.709 | 1.436 | 1.803 | 1.974 | 2.021 | 2.128 | 1.963 | 2.186 | |
| $SE_3$ | | | | | | | | | | | |
| Asymptotic (ML) | 0.062 | 0.062 | 0.065 | 0.084 | 0.076 | 0.065 | 0.065 | 0.080 | 0.200 | 0.168 | 0.021 |
| Bootstrap | 0.058 | 0.055 | 0.062 | 0.085 | 0.074 | 0.113 | 0.096 | 0.144 | 0.222 | 0.201 | 0.030 |
| Sandwich estimator (PML) | 0.076 | 0.074 | 0.085 | 0.110 | 0.095 | 0.141 | 0.119 | 0.196 | 0.302 | 0.237 | 0.035 |
| $bias_3$ % [asymptotic] | −18.90% | −16.35% | −23.76% | −23.55% | −20.40% | −53.76% | −45.63% | −58.98% | −33.77% | −29.11% | −40.00% |
| $bias_3$ % [bootstrap] | −23.36% | −25.95% | −26.71% | −22.73% | −22.29% | −20.00% | −18.99% | −26.58% | −26.59% | −15.19% | −14.29% |

***Table I.****SE misspecifications according to SRS design*

$$\mathrm{SE}(\bar{x}) = \frac{\sigma_x}{\sqrt{n}} \tag{1}$$

*(see equation 1)*

$$\text{deff} = 1 + \rho(m-1) \tag{2}$$

*(see equation 2)*

$$y_{ij} = \beta_0 + u_{0i} + \varepsilon_{ij} \tag{3}$$

*(see equation 3)*

$$\rho = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2} \tag{4}$$

*(see equation 4)*

$$\text{bias asymp}\,\% = \frac{\text{SE}_{\text{asympt(ML)}} - \text{SE}_{\text{sandwich est.}}}{\text{SE}_{\text{asympt(ML)}}} \tag{5}$$

$$\text{bias bootstr}\,\% = \frac{\text{SE}_{\text{bootstrap}} - \text{SE}_{\text{sandwich est.}}}{\text{SE}_{\text{bootstrap}}} \tag{6}$$

*(see equation 5)*

$$\text{cM} = \frac{\text{NM}\chi^2}{\text{ScM}\chi^2} = \frac{113.553}{20.199} = 5.622$$

$$\text{cL} = \frac{\text{NL}\chi^2}{\text{ScL}\chi^2} = \frac{23.376}{4.360} = 5.361$$

*(see equation 6)*

$$\text{dts} = \frac{(\text{dfM} * \text{cM}) - (\text{dfL} * \text{cL})}{(\text{dfM} - \text{dfL})} = \frac{(5 * 5.622) - (4 * 5.361)}{(5-4)} = 6.666$$

*(see equation 7)*

$$\mathrm{Sc}\Delta\chi^2 = \frac{\mathrm{NM}\Delta\chi^2 - \mathrm{NL}\Delta\chi^2}{\mathrm{dts}} = \frac{113.553 - 23.376}{6.666}$$
$$= 13.528 \quad \text{for 1 df } (p = 0.000)$$

*(see equation 8)*

## References


Aaker, J.L. (1997), "Dimensions of brand personality", *Journal of Marketing Research,* Vol. 34 No. 3 pp.347-56.

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

Arbuckle, J.L. (2006), *AMOS (version 7.0),* SPPS Inc, Chicago, .

Bagozzi, R.P., Youjae, Y., Phillips, L.W. (1991), "Assessing construct validity in organizational research", *Administrative Science Quarterly*, Vol. 36 No.3, pp.421-58.

Bryk, A.S., Raudenbush, S.W. (1992), *Hierarchical Linear Models: Applications and Data Analysis Methods,* Sage Publications, Newbury Park, .

Diaconis, P., Efron, B. (1983), "Computer-intensive methods in statistics", *Scientific American*, Vol. 248 No.5, pp.116-30.

Dillon, W.R., Goldstein, M. (1984), *Multivariate Analysis: Methods and Applications,* Wiley, New York, .

Dillon, W.R., Madden, T.J., Kirmani, A., Soumen, M. (2001), "Understanding what's in a brand rating: a model for assessing brand attribute effects and their relationship to brand equity", *Journal of Marketing Research*, Vol. 38 No.4, pp.415-29.

Efron, B. (1979), "Bootstrap methods: another look at the jackknife", *The Annals of Statistics*, Vol. 7 No.1, pp.1-26.

Efron, B., Tibshirani, R. (1993), *An Introduction to the Bootstrap*, Chapman and Hall, London, .

Erdem, T., Swait, J. (1998), "Brand equity as a signaling phenomenon", *Journal of Consumer Psychology*, Vol. 7 No.2, pp.131-57.

Erdem, T., Swait, J. (2004), "Brand credibility, brand consideration, and choice", *Journal of Consumer Research*, Vol. 31 No.1, pp.191-98.

Erdem, T., Swait, J., Valenzuela, A. (2006), "Brands as signals: a cross-country validation study", *Journal of Marketing*, Vol. 70 No.1, pp.34-49.

Fornell, C., Larker, D.F. (1981), "Evaluating structural equation models with unobservable variables and measurement error", *Journal of Marketing Research*, Vol. 18 No.3, pp.39-50.

Frenkel, M.R., Frenkel, L.R. (1977), "Some recent developments in sample survey design", *Journal of Marketing Research*, Vol. 14 No.3, pp.280-93.

Hox, J.J., Maas, C.J.M. (2001), "The accuracy of multi-level structural equation modeling with pseudobalanced groups and small samples", *Structural Equation Modeling: A Multidisciplinary Journal*, Vol. 8 No.2, pp.157-74.

Ippolito, P.M. (1990), "Bonding and nonbonding signals of product quality", *Journal of Business*, Vol. 63 No.1, pp.41-60.

Jöreskog, K.G., Sörbom, D. (2006), *LISREL 8 (version 8.8)*, Scientific Software International Inc, Chicago, IL, .

Julian, M.W. (2001), "The consequences of ignoring multi-level data structures in nonhierarchical covariance modeling", *Structural Equation Modeling: A Multidisciplinary Journal*, Vol. 8 No.3, pp.325-52.

Kish, L. (1965), *Survey Sampling*, John Wiley, New York, .

Kish, L., Frankel, M.R. (1974), "Inference from complex samples", *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, Vol. 36 No.1, pp.1-37.

Klein, B., Leffler, K.B. (1981), "The role of market forces in assuring contractual performance", *Journal of Political Economy*, Vol. 89 No.4, pp.615-41.

Lee, E.S., Forthofer, R.N. (2005), *Analyzing Complex Survey Data*, 2nd ed., Sage, Newbury Park, .

Maas, C.J.M., Hox, J.J. (2005), "Sufficient sample sizes for multi-level modeling", *Methodology*, Vol. 1 No.3, pp.85-91.

Muthén, L.K., Muthén, B.O. (2008), *Mplus (version 5.1)*, Muthén & Muthén, Los Angeles, .

Muthén, B.O., Satorra, A. (1995), "Complex sample data in structural equation modeling", *Sociological Methodology*, Vol. 25 pp.267-316.

Paxton, P., Curran, P.J., Bollen, K.A., Kirby, J., Chen, F. (2001), "Monte Carlo experiments: design and implementation", *Structural Equation Modeling: A Multidisciplinay Journal*, Vol. 8 No.2, pp.287-312.

Pfeffermann, D., Skinner, C.J., Holmes, D.J., Goldstein, H., Rasbash, J. (1998), "Weighting for unequal selection probabilities in multi-level models", *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, Vol. 60 No.1, pp.23-40.

Rao, A.R., Qu, L., Reuekert, R.W. (1999), "Signaling unobservable product quality through a brand ally", *Journal of Marketing Research*, Vol. 36 No.2, pp.258-68.

Rubenstein, R.Y. (1981), *Simulation and the Monte Carlo Method*, John Wiley, New York, .

Satorra, A., Bentler, P.M. (2001), "A scaled difference chi-square test statistic for moment structure analysis", *Psychometrika*, Vol. 66 No.4, pp.507-14.

Scariano, S.M., Davenport, J.M. (1987), "The effects of violations of independence assumptions in the one-way ANOVA", *The American Statistician*, Vol. 41 No.2, pp.123-29.

Scientific Software International Inc (2005), "Analysis of structural equation models for continuous random variables in the case of complex survey data", available at, .

Shackman, G. (2001), "Developing sample sizes for new surveys: estimating the design effect adjustment", in", in Bailar, J. (Eds),*Conference Application in Statistics: What's Hot / What's Next*, American Statistical Association, Albany, Vol. VI.

Shrout, P.E., Fleiss, J.L. (1979), "Intraclass correlation: uses in assessing rater reliability", *Psychological Bulletin*, Vol. 86 No.2, pp.420-28.

Skinner, C.J., Holt, D., Smith, F.T.M. (1989), *Analysis of Complex Surveys*, John Wiley, Chichester, .

Stapleton, L.M. (2006), "An assessment of practical solutions for structural equation modeling with complex sample data", *Structural Equation Modeling: A Multidisciplinary Journal*, Vol. 13 No.1, pp.28-58.

Wold, H. (1985), "partial least squares", in Kotz, S. and Johnson, N.L. (Eds)", *Encyclopedia of Statistical Sciences*, Wiley, New York, Vol. 6 pp.581-91.

## Appendix. $\chi^2$

#### difference testing using the scaled

$\chi^2$

In order to compute $\Delta\chi^2$ testing, the following estimates are needed:

- scaled $\chi^2$ (ScM $\chi^2$) for the MORE constrained model = 20.199;
- normal $\chi^2$ (NM $\chi^2$) for the MORE constrained model = 113.553;
- degrees of freedom (dfM) for the MORE constrained model = 5;
- scaled $\chi^2$ (ScL $\chi^2$) for the LESS constrained model = 4.360;
- normal $\chi^2$ (NL $\chi^2$) for the LESS constrained model = 23.376; and
- degrees of freedom (dfL) for the LESS constrained model = 4.

The scaling correction factors of both models are first computed:(see equation 6)Second, it is estimated the difference test scaling (dts) correction where dfM is the degrees of freedom in the more constrained model and dfL is the degrees of freedom in the less constrained model:(see equation 7)

Third, the scaled $\chi^2$ difference test (Sc$\Delta\chi^2$) is computed as follows:(see equation 8)

## About the author

Andrea Vocino is a Lecturer at the School of Management and Marketing at Deakin University in Burwood, Australia and PhD candidate in the Department of Marketing, at Monash University, Melbourne, Australia. His thesis is on the relationships between store names and product brand names as brand signals. His research interests also include marketing research methods and marketing science. Andrea Vocino can be contacted at: andrea.vocino@deakin.edu.au