
A primer for the estimation of structural equation models in the presence of missing data: Maximum likelihood algorithms

Received (in revised form): 9th January, 2002

Cara Lee Okleshen Peters

is an assistant professor of marketing from the University of Georgia. She received her PhD from the University of Nebraska. Dr Peters' research is in the areas of customer communities, electronic commerce and research methods. She has published papers in the *Journal of Consumer Psychology* and *Consumption, Markets, and Culture*, among other journals. She teaches courses in electronic commerce, marketing communication and qualitative research.

Craig Enders

is an assistant professor of educational research at the University of Miami. His research interest is structural equation modelling methodology, in particular issues related to missing data.

Abstract Maximum likelihood algorithms have undergone substantial development, yet marketers have been slow to adopt these techniques to address missing data. This study familiarises marketers with the available maximum likelihood estimators, reviews missing data theory and research, and presents a structural equation modelling simulation study to demonstrate the advantages of maximum likelihood estimation versus other techniques. Results indicate that the full information maximum likelihood and expectation-maximisation outperform traditional techniques with respect to parameter estimate bias, model fit and parameter estimate efficiency. Marketers should be aware of the potential impact of missing data assumptions and decrease their reliance on *ad hoc* methods in favour of maximum likelihood estimators.

INTRODUCTION

Marketing researchers typically treat missing observations using one of the classic *ad hoc* techniques that attempt to fix the incomplete data matrix prior to analysis. Listwise deletion discards all cases with missing values and, therefore, could waste a great deal of usable data, resulting in lower power. Pairwise deletion computes each variance/covariance matrix element separately using all available data for a bivariate pair, and the individual covariance matrix elements are

aggregated into one matrix. Pairwise deletion results in a covariance matrix that may not be positive definite due to the use of different subsets of cases. Mean imputation replaces missing observations with the variable mean. Mean imputation underestimates the variance of the variable and thus the covariance with other variables. These techniques may be practical due to their widespread availability as default options in statistical software. Yet, because of their availability, marketers may assume that these *ad hoc* methods possess

Cara Lee Okleshen Peters
Department of Marketing
and Distribution, Terry
College of Business,
University of Georgia,
Athens, GA 30602-6258,
USA

Tel: +1 706 542 3766;
Fax: +1 706 542 3738;
e-mail: okleshen@terry.uga.
edu

desirable statistical qualities and will not bias parameter estimates. In many cases, this is far from the truth.

During the past 20 years, maximum likelihood missing data estimators have undergone substantial development.¹ Malhotra introduced marketers to one maximum likelihood method, expectation-maximisation,² but few marketers adopted the approach. In 1987, pre-packaged software was not widely available, which may have contributed to a lack of use. Now maximum likelihood estimators are widely available in statistical software and are easy to implement. While the lack of software may have contributed to the reliance on *ad hoc* methods, it may also be the case that the theoretical benefits of using maximum likelihood estimators are not widely understood. The goals of this study are to: familiarise marketers with the available maximum likelihood estimators, review missing data theory and research, and present a structural equation model simulation study to demonstrate the advantages of maximum likelihood estimation versus other techniques.

The simulation is based on a structural equation model from an actual marketing study. This is appropriate since much of the recent missing data literature is in the field of structural equation modelling; many of the implementations of maximum likelihood estimators are found in structural equation modelling software; and structural equation modelling is frequently used in marketing research. The ideas presented in this paper are not, however, limited to structural equation model analyses and are more generally applicable to the family of linear models, most of which are subsumed under the structural equation modelling framework.^{3,4} Thus, even though maximum likelihood estimators are frequently found in structural equation

modelling software, traditional linear model analyses (eg regression, ANOVA) can be performed with these packages.

MAXIMUM LIKELIHOOD MISSING DATA ESTIMATION

Three approaches for obtaining maximum likelihood estimates have been discussed in the statistical literature: the multiple-group approach, full information maximum likelihood and expectation-maximisation. While these estimators differ in their mathematical approach, they are built on the premise that the covariation among variables can be used to infer, or even estimate, probable values for the missing data. Although an over-simplification, information is borrowed from the complete data (via the covariation among variables) to assist in the estimation of parameters that involve missing values. Partially complete cases are not discarded but are incorporated into the estimation process. This serves two purposes. Incorporating partially complete cases can increase the precision (ie lower sampling variability) of parameter estimates, and the inclusion of these observations can decrease parameter estimate bias.

Hartley and Hocking⁵ provided a maximum likelihood missing data estimator that was applied to structural equation modelling analyses by Muthén, Kaplan and Hollis.⁶ This method is referred to as the multiple-group approach, as its mathematical form is similar to the LISREL estimator used for multiple-group structural equation modelling analyses. This method has not experienced widespread use due to its programming complexity and the fact that it is practically restricted to situations that have a small number of distinct missing data patterns.⁷

The full information maximum likelihood approach was originally

outlined by Finkbeiner⁸ for use with factor analysis and is similar to the multiple-group approach, except that the likelihood function is comprised of N components, each of which contain the available data for a given case, rather than group, level. For this reason, the full information maximum likelihood approach has been referred to as raw maximum likelihood estimation.⁹ AMOS,¹⁰ Mx,¹¹ MPLUS and LISREL currently offer full information maximum likelihood estimation routines.

Assuming multivariate normality, each case's contribution to the likelihood function is

$$\log L_i = K_i - \frac{1}{2} \log |\Sigma_i| - \frac{1}{2} (x_i - \mu_i)' \Sigma_i^{-1} (x_i - \mu_i), \quad (1)$$

where x_i is the vector of complete data for case i , μ_i contains the corresponding mean estimates derived from the entire sample, K_i is a constant that depends on the number of complete data points for case i , and Σ_i is the variance/covariance matrix for x_i . Like μ_i , the determinant and inverse of Σ_i are based only on the variance/covariance terms for which case i has complete data. The overall discrepancy function value is obtained by summing the n (ie number of observations) casewise likelihood functions as follows:

$$\log L(\mu, \Sigma) = \sum_{i=1}^N \log L_i. \quad (2)$$

Several points should be made about the full information maximum likelihood estimator. First, one of the advantages of the full information maximum likelihood is its applicability to both just-identified and over-identified models. In the latter case, the likelihood equation above is extended such that μ and Σ are expressed as functions of some parameter vector, γ .¹² As such, the method can be

applied to a wide variety of general linear model analyses. Secondly, when used in structural equation modelling applications, full information maximum likelihood yields a χ^2 of model fit. The chi-square statistic generated by full information maximum likelihood is not, however, expressed in the usual form $F(N-1)$, where F is the value of the fitting function, as there is no single value of N that is applicable to the entire sample. Also, unlike the usual structural equation modelling fitting functions, there is no minimum value associated with the FIML log-likelihood function — the value increases as model fit worsens, however. As such, the χ^2 statistic is calculated as the difference between log-likelihood values for the H_0 and H_1 models, with degrees of freedom equal to the difference in the number of estimated parameters between the two. Although many popular fit indices are available when using full information maximum likelihood, the specification of a means structure (see Equation 1) makes certain fit indices (eg GFI) undefined. Fourthly, indefinite covariance matrices are a potential by-product of the full information maximum likelihood approach, but Wothke¹³ suggests that these problems are not as prevalent as those associated with pairwise deletion. Finally, missing values are not imputed under full information maximum likelihood; model parameter estimates and standard errors are estimated directly using all available data.

Dempster, Laird and Rubin first discussed the concepts underlying expectation-maximisation.¹⁴ The expectation-maximisation estimator is currently available in the SPSS MVA (Missing Values Analysis), SAS, EQS, MPLUS, EMCOV¹⁵ and NORM.¹⁶ The expectation-maximisation estimator uses a two-step iterative procedure whereby missing observations are estimated and

unknown parameter estimates are obtained. To begin the iterative cycle, an initial estimate of the covariance matrix and mean vector must be obtained using listwise deletion, pairwise deletion or some imputation method.

In the *E* step, missing values are replaced with the conditional mean of the missing data given the observed data and the initial covariance matrix estimate. That is, missing values are replaced by predicted scores from a series of regression equations where each missing variable is regressed on the remaining observed variables for a case *i*. The subsequent *M* step is simply a complete-data maximum likelihood estimation problem, as maximum likelihood estimates of the mean vector and covariance matrix are obtained using the filled-in data from the *E* step. This updated covariance matrix is then used to derive regression equations for the next *E* step and the cycle begins again. The estimator repeatedly cycles through these steps until the difference between covariance matrices in subsequent *M* steps falls below some specified convergence criterion.¹⁷

Several points should be noted concerning expectation–maximisation. First, unlike the full information maximum likelihood approach, the expectation–maximisation estimator yields a mean vector and covariance matrix that can be used as input for subsequent general linear model analyses (eg regression). This covariance matrix can also be used to impute missing values following the final expectation–maximisation iteration. This is appealing due to the illusion of a complete data set, but there is a drawback associated with this practice. Although the imputed values are optimal statistical estimates of the missing observations, they lack the residual

variability present in the hypothetically complete data set; the values are imputed without a random error component. As a result, standard errors from subsequent analyses will be negatively biased, and bootstrap procedures must be employed to obtain correct estimates. A correction factor, however, is added to the conditional expectation of the missing data at each *E* step to correct for this negative bias in the output covariance matrix. As such, standard error bias should be less problematic when performing analyses using the expectation–maximisation covariance matrix rather than an EM-imputed data matrix, yet no studies have examined this issue. Alternatively, multiple imputation procedures designed to recover residual variability are available in EMCOV¹⁸ and NORM.¹⁹

MISSING DATA THEORY

Rubin was the first to describe the mechanisms that result in missing observations.²⁰ According to Rubin, observations are missing completely at random (MCAR) when the missing values on a variable *X* are independent of other observed variables as well as the values of *X* itself. This means that the missing *X* values are simply a random sample of the hypothetically complete data set. Both listwise deletion and pairwise deletion require this assumption in order to yield unbiased parameter estimates. While discarding incomplete observations will certainly decrease the efficiency (ie increase sampling variability) of parameter estimates and result in lower power, the distributional characteristics and covariance structure of the remaining complete cases will differ from the hypothetically complete data set only with respect to random sampling error. Thus, across repeated samples, the expected value of the parameter estimate

in question will not deviate from the population parameter as a result of using listwise deletion or pairwise deletion.

Maximum likelihood estimators should also yield unbiased parameter estimates when MCAR holds. They have, however, an important advantage. Maximum likelihood missing data estimators utilise covariation among variables to derive probable values for the missing data points during the estimation process. While this additional information does nothing to improve accuracy (ie bias) under MCAR, it does improve efficiency, resulting in decreased sampling error relative to listwise deletion and pairwise deletion. Thus, although maximum likelihood, listwise deletion and pairwise deletion should yield unbiased parameter estimates when MCAR holds, the parameter estimates from any single sample should, on average, be closer to the true estimate under maximum likelihood estimation. Several researchers have noted that the MCAR assumption is quite strict, and probably does not hold in many applied situations.²¹ If this contention is true, then the choice of missing data technique becomes an important data analytic decision.

A second missing data mechanism is the missing at random (MAR) condition.²² This is a misnomer, as missing values on a variable X are systematically related to other variables. Specifically, MAR holds when missing values on X are dependent on another observed variable but not on the values of X itself. Thus, the observed values of X are not a random sample of the hypothetically complete data set but are instead a random sample within subgroups defined by scores on another variable, Y .

To illustrate, suppose consumers are asked to complete two virtual reality shopping trips at time zero ($t = 0$) and

one year later ($t = 1$). After administering the second shopping simulation, it is found that those respondents whose perceived skill was low at navigating the $t = 0$ environment refused to participate in the second shopping simulation at $t = 1$. Missing values at $t = 1$ represent, however, a random sample within a subgroup of $t = 0$ cases sharing the same level of perceived skill. Thus, the propensity for missing data at $t = 1$ is dependent on one's perceived skill at the first simulation ($t = 0$) and is unrelated to the underlying level of performance that would have been attained at the second simulation ($t = 1$). A second example derived from the simulation study further illustrates the MAR mechanism.

Research on brand communities indicates that participation in a group is positively related to perceptions of sentiments about other members.²³ Consumers who rarely attend group events and activities may feel, however, that they have not had enough contact with the group to judge adequately their sentiments about other members. Consequently, those consumers may have higher nonresponse on measures that assess their attachments to other members. In this case the missing values on the sentiment measures are dependent on the level of participation but are unrelated to attachment itself for any group of respondents with identical levels of participation.

MISSING DATA LITERATURE

There is a growing body of literature supporting the use of maximum likelihood missing data estimators. Using bootstrapped samples from an actual dataset, Arbuckle found that full information maximum likelihood estimates of confirmatory factor analysis (CFA) model parameters were generally unbiased under MCAR and MAR

mechanisms.²⁴ Listwise deletion and pairwise deletion estimates were unbiased under MCAR but biased under MAR. Enders and Bandalos' Monte Carlo simulation suggested a similar pattern of results from both a CFA and full structural model,²⁵ as did Wothke using a latent growth curve model.²⁶ While these studies utilised multivariate normal data, Enders found that structural equation modelling parameter estimates were also unbiased under non-normal conditions, although standard error estimates were negatively biased — they were also biased for traditional missing data techniques, however.²⁷ Finally, Enders found that the full information maximum likelihood was superior to *ad hoc* methods in the context of a multiple regression model. In these studies full information maximum likelihood estimates were relatively unaffected by the amount of missing data, while the bias due to *ad hoc* methods increased with the missing data rate.²⁸

With respect to efficiency, research has demonstrated that full information maximum likelihood parameter estimates have substantially less sampling variability than listwise deletion under MCAR,^{29–31} and this efficiency advantage increases as the percentage of missing data increases. In contrast, efficiency comparisons between full information maximum likelihood and pairwise deletion have yielded mixed results. Arbuckle's³² results suggested that full information maximum likelihood parameter estimates have substantially less sampling variability than pairwise deletion, while the other researchers reported only a modest advantage.

Finally, maximum likelihood estimators appear to be superior with respect to model fit. Research has shown that the χ^2 fit statistic is inflated when using pairwise deletion due to violating Wishart distributional assumptions.^{33–35} In contrast,

maximum likelihood missing data estimators yield rejection rates near the nominal 5 per cent level when normality assumptions are met. For example, Enders and Bandalos found that full information maximum likelihood χ^2 rejection rates were close to the nominal 5 per cent level and were relatively uninfluenced by the missing data rate.³⁶ Like complete-data maximum likelihood estimators, however, χ^2 are inflated under non-normal conditions,³⁷ although *ad hoc* techniques provide no advantage either.

SIMULATION STUDY

Two small-scale simulation studies were conducted to illustrate the previous discussions. The simulations used as a population model a six-indicator structural model from an actual marketing study on brand communities (see Figure 1 for a path diagram) collected from 221 members of Winnebago-Itasca Travelers (WIT), a recreational vehicle group that is the nation's largest manufacturer-managed customer club. The model proposes that CONTACT with a customer community, measured by participation in group events and activities, is positively related to members' SENTIMENTS toward the community, as measured by member attachment and group attachment.³⁸ The SENTIMENTS latent variable is, in turn, positively related to BEHAVIOURAL INTENTIONS, which is defined by three indicator variables: positive referral, intention to repurchase and a willingness to pay more.³⁹ For identification purposes, the uniqueness term for the contact indicator was fixed using an internal consistency reliability estimate generated from the data. Likewise, a single factor loading from each of the two remaining latent variables was fixed at unity — this specification is not reflected in Figure 2,

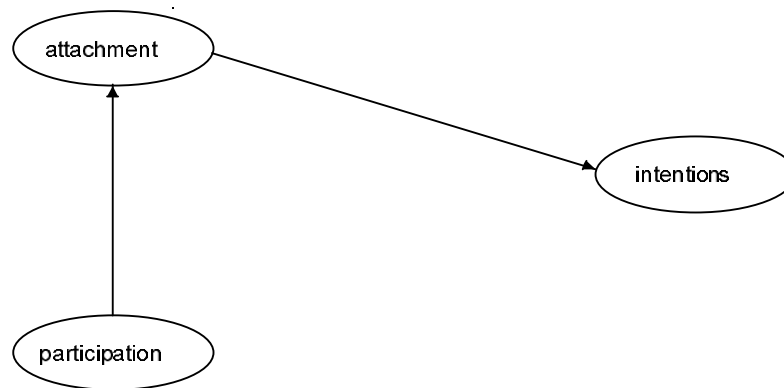


Figure 1 Path diagram of brand community

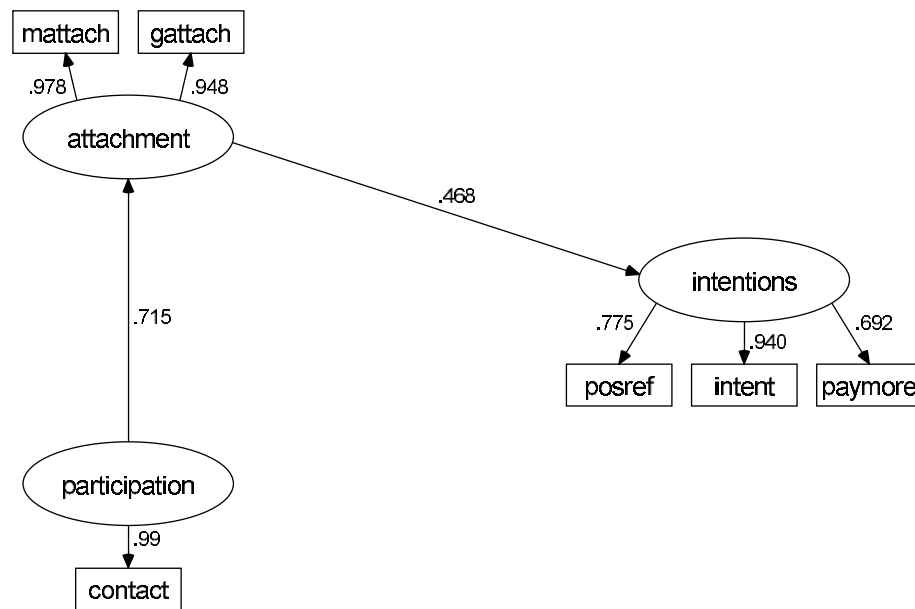


Figure 2 Standardised solution for the brand community structural model

which shows the population values for the standardised solution. See Peters and Grossbart for further details.⁴⁰

The reproduced covariance matrix from the model served as the population covariance matrix for the two simulations. As such, the parameter estimates shown in Figure 2 can be viewed as the ‘true’ population parameters. For both simulations, 200 sample data matrices were generated from the population model, missing

values were imposed, and five missing data estimators (full information maximum likelihood, expectation-maximisation, listwise deletion, pairwise deletion and mean imputation) were applied to the sample matrices. The goal of the simulations was to determine which missing data estimator/s best recovered the true population parameters.

The 200 sample data matrices (300 observations by six variables) were

generated using the RANNOR random number generator in the SAS IML procedure. Using the population covariance matrix generated from the model, these random normal variates were transformed to the desired covariance structure using Cholesky decomposition. Next, MCAR and MAR missing data patterns were imposed. To simulate MCAR data, 20 per cent of the cases were randomly deleted from the two attachment indicator variables. Random deletion was performed separately for both indicators, so observations could be missing on one or both of the variables. To simulate the MAR condition, missing values on the two attachment variables were dependent on the values of contact indicator variables. Specifically, the probability of missing data on the two attachment indicators was inversely related to an observation's rank order on the contact variable, such that observations with low contact scores had a higher probability of deletion. For example, an observation at the 10th percentile of the contact distribution would have a 90 per cent deletion probability for the two attachment indicators. Beginning with the lowest contact score, missing values were imposed in this manner until a 20 per cent missing data rate was reached on the attachment variables; in this simulation these two indicators were always deleted as a pair. This procedure simulates the scenario described earlier where group members having minimal contact with the community have a higher propensity to skip items measuring their group attachment — they lack the experience to answer these items.

Two maximum likelihood estimators, full information maximum likelihood and expectation-maximisation, were tested in the simulation. Full information maximum likelihood requires raw data,

so the 400 (200 MCAR + 200 MAR) raw data matrices were submitted directly to the AMOS 4.0 computer program for analysis. Expectation-maximisation yields a covariance matrix and mean vector that is subsequently used as input into a structural equation modelling program. The 400 expectation-maximisation covariance matrices were generated using a program written in SAS IML code. The results from this program were compared to output from the SPSS MVA program to verify programming accuracy. When using covariance matrices as input it is necessary to specify a single value of N . Not all elements of the expectation-maximisation covariance matrix, however, are estimated from a common sample size, and no research exists that provides insight into the appropriate sample size choice. Therefore, the minimum N per covariance term was used — this is the same value that is frequently input when using pairwise deletion. (The value of N is not a trivial issue as model fit statistics such as the χ^2 are impacted by this choice.) The remaining three missing data techniques (listwise deletion, pairwise deletion and mean imputation) were also implemented in SAS, and covariance matrices from these methods were used as input into the AMOS 4.0 computer program. Again, the minimum N per covariance term was used for the pairwise deletion matrices. The process of fitting the 2,000 models (400 samples \times five missing data estimators) was automated using a visual basic program, and parameter estimates and fit statistics from each sample were output to a file for analysis.

Three dependent variables were examined in the simulation: parameter estimate bias, model fit and parameter estimate efficiency (ie sampling variability). Parameter estimate bias was measured by the percentage of bias

relative to the true population parameter as follows:

$$\%BLAS = \frac{\hat{\theta}_j - \theta_j}{\theta_j} \star 100 \quad (3)$$

The parameter estimates and population values for the j th missing data estimator are given by $\hat{\theta}_j$ and θ_j , respectively. The mean values of this statistic were computed for each estimator. Model fit will be assessed using the χ^2 goodness-of-fit test. Because the 200 sample data matrices were generated from a population model with perfect fit, the percentage of model rejections at the 0.05 level of significance can be computed for each missing data estimator and compared to the nominal rate of 5 per cent. Finally, efficiency was measured using a relative efficiency ratio computed as

$$RE = \frac{\sigma_{\hat{\theta}_j}^2}{\sigma_{\theta_{FML}}^2}, \quad (4)$$

where $\sigma_{\hat{\theta}_j}^2$ is the empirical sampling variance for the j th missing data estimator computed across the 200 replications and $\sigma_{\theta_{FML}}^2$ is the empirical sampling variance of the full information maximum likelihood estimates. Because missing data theory predicts that maximum likelihood will yield the lowest sampling variability, it is reasonable to use full information maximum likelihood as a benchmark.

RESULTS

Not all of the 200 replications converged and gave admissible solutions. About 7 per cent of the 200 solutions from each estimator were discarded, although this value was higher for pairwise deletion under the MCAR condition (14.5 per cent) and lower for mean imputation under MCAR (0.5 per cent). Results are based only on those replications that

passed convergence and admissibility checks.

SIMULATION 1: MISSING COMPLETELY AT RANDOM

The first simulation modelled the situation where observations were randomly missing (ie MCAR) from the two indicators of the attachment latent variable. Table 1 gives the mean standardised parameter estimates and percentage of bias relative to the true population parameter for each missing data estimator. Structural and measurement model parameter estimates were close to the true population parameters, so the choice of missing data technique made virtually no difference, at least with respect to the long-run expectation of the parameter estimate. The two exceptions to this trend were seen with the mean imputation factor loadings for the attachment latent variable (the variable with missing values on its indicators). In this case, there was a substantial negative bias in these two loadings. This is not surprising, as mean imputation restricts the variance of the imputed variable, which negatively biases covariation with other variables.⁴¹ While judging the practical impact of bias is subjective, Muthén *et al.* suggests that bias levels less than 10 per cent are probably not problematic.⁴²

Table 1 also gives the percentage of model rejections based on χ^2 goodness-of-fit tests ($p < 0.05$). While most of the percentages were relatively close to the nominal rate, pairwise deletion rejection rates approach 20 per cent. Previous research has shown that the use of pairwise deletion violates structural equation modelling Wishart distribution assumptions, resulting in inflated χ^2 values.^{43,44} Also, the rejection rate for expectation-maximisation was less than that of full information

Table 1: Mean parameter estimates, percentage bias and chi-square rejection rates by algorithm

| Parameter | Pop | | FIML | | EM | | LD | | PD | | MI | |
|-----------------------------------|-------|-------|-------|-------|--------|--------|-------|--------|-------|--------|-------|--------|
| | Mean | Bias | Mean | Bias | Mean | Bias | Mean | Bias | Mean | Bias | Mean | Bias |
| MCAR | | | | | | | | | | | | |
| CONTACT→ATTACH | 0.715 | 0.00% | 0.714 | 0.0% | 0.712 | -0.4% | 0.713 | -0.2% | 0.714 | -0.2% | 0.714 | -0.1% |
| ATTACH→INTENTIONS | 0.468 | 0.0% | 0.468 | 0.0% | 0.464 | -0.7% | 0.468 | 0.0% | 0.467 | 0.0% | 0.467 | -0.2% |
| CONTACT | 0.990 | 0.0% | 0.990 | 0.0% | 0.990 | 0.0% | 0.990 | 0.0% | 0.990 | 0.0% | 0.990 | 0.0% |
| MATTACH | 0.978 | -0.1% | 0.977 | -0.1% | 0.977 | -0.1% | -0.4% | 0.874 | 0.874 | 0.874 | 0.874 | -10.6% |
| GATTACH | 0.948 | 0.1% | 0.948 | 0.0% | 0.948 | 0.0% | 0.950 | 0.3% | 0.844 | 0.3% | 0.844 | -10.9% |
| POSREF | 0.775 | 0.0% | 0.775 | 0.0% | 0.774 | -0.1% | 0.776 | 0.1% | 0.775 | 0.1% | 0.775 | 0.0% |
| INTENT | 0.940 | 0.0% | 0.939 | 0.0% | 0.939 | 0.0% | 0.940 | 0.0% | 0.939 | 0.0% | 0.939 | 0.0% |
| PAYMORE | 0.692 | 0.0% | 0.692 | 0.0% | 0.691 | 0.0% | 0.691 | -0.1% | 0.692 | -0.1% | 0.692 | 0.0% |
| MAR | | | | | | | | | | | | |
| CONTACT→ATTACH | 0.715 | -0.1% | 0.714 | -0.1% | 0.627 | -12.2% | 0.625 | -12.6% | 0.440 | -38.4% | 0.440 | -38.4% |
| ATTACH→INTENTIONS | 0.468 | -0.1% | 0.466 | -0.3% | 0.429 | -8.2% | 0.430 | -7.9% | 0.381 | -18.4% | 0.381 | -18.4% |
| CONTACT | 0.990 | 0.0% | 0.990 | 0.0% | 0.983 | -0.7% | 0.990 | 0.0% | 0.990 | 0.0% | 0.990 | 0.0% |
| MATTACH | 0.978 | -0.1% | 0.977 | -0.1% | 0.972 | -0.6% | 0.971 | -0.7% | 0.970 | -0.8% | 0.970 | -0.8% |
| GATTACH | 0.948 | 0.0% | 0.948 | 0.1% | 0.937 | -1.1% | 0.938 | -1.1% | 0.939 | -0.9% | 0.939 | -0.9% |
| POSREF | 0.775 | 0.0% | 0.775 | 0.0% | 0.770 | -0.6% | 0.774 | -0.1% | 0.775 | 0.0% | 0.775 | 0.0% |
| INTENT | 0.940 | 0.0% | 0.939 | -0.1% | 0.936 | -0.3% | 0.940 | 0.0% | 0.940 | 0.0% | 0.940 | 0.0% |
| PAYMORE | 0.693 | 0.2% | 0.693 | 0.2% | 0.685 | -0.9% | 0.692 | 0.1% | 0.692 | 0.1% | 0.692 | 0.1% |
| Chi-square rejection rates | | | | | | | | | | | | |
| MCAR | 6.70% | 7.49% | 3.74% | 7.57% | 7.61% | 18.71% | 6.03% | | | | | |
| MAR | 6.70% | 7.07% | 6.56% | 7.61% | 14.52% | 66.84% | | | | | | |

Table 2: Ratios by algorithm

| Parameter | EM/FIML | LD/FIML | PD/FIML | MI/FIML |
|---------------------------------|---------|---------|---------|---------|
| MCAR relative efficiency | | | | |
| CONTACT→ATTACH | 0.99 | 1.51 | 1.02 | 0.82 |
| ATTACH→INTENTIONS | 0.99 | 1.61 | 1.16 | 0.88 |
| CONTACT | 1.00 | 1.74 | 0.99 | 1.02 |
| MATTACH | 1.01 | 1.01 | 1.68 | 0.17 |
| GATTACH | 1.01 | 1.13 | 1.50 | 0.23 |
| POSREF | 1.00 | 1.48 | 1.00 | 1.04 |
| INTENT | 1.00 | 1.41 | 1.03 | 1.00 |
| PAYMORE | 1.00 | 1.68 | 1.00 | 1.02 |
| MAR MSE | | | | |
| CONTACT→ATTACH | 1.01 | 7.01 | 7.37 | 58.31 |
| ATTACH→INTENTIONS | 1.01 | 1.72 | 1.64 | 3.56 |
| CONTACT | 1.01 | 60.97 | 1.00 | 0.98 |
| MATTACH | 1.01 | 1.72 | 1.78 | 1.93 |
| GATTACH | 1.01 | 2.22 | 1.98 | 1.79 |
| POSREF | 1.01 | 1.33 | 1.02 | 0.97 |
| INTENT | 1.01 | 1.50 | 1.13 | 1.15 |
| PAYMORE | 1.01 | 1.37 | 0.99 | 0.99 |

maximum likelihood. This is probably due to the relatively conservative sample size used in this simulation — the minimum N per covariance term. The use of a less conservative value would increase the χ^2 , although there is no research to guide the sample size choice.

The most important aspect of the MCAR demonstration is seen in Table 2, which gives the relative efficiency ratios for each missing data estimator relative to that of full information maximum likelihood. Values larger than unity reflect situations where the sampling variability of a particular estimator is greater than that of full information maximum likelihood. The efficiency of expectation-maximisation estimates was virtually identical to that of full information maximum likelihood. Listwise deletion yielded substantially larger sampling variability than the maximum likelihood estimators. The listwise deletion sampling variance was, on average, approximately 50 per cent higher than the maximum likelihood estimators. Because sampling variance is inversely related to sample size, this suggests that a 50 per cent increase in

sample size would be required to yield the same level of efficiency as the maximum likelihood estimators. This suggests that, while listwise deletion is unbiased over the long run, parameter estimates from any single sample would be closer, on average, to the true population parameter when using maximum likelihood estimators. The differences between the maximum likelihood estimators, pairwise deletion, and mean imputation are generally less extreme. In the case of the two attachment loadings, mean imputation yielded much more efficient parameter estimates than maximum likelihood. As argued by Wothke, imputing the mean value results in precise, albeit biased, estimates.⁴⁵

SIMULATION 2: MISSING AT RANDOM

The MAR simulation modelled the situation where missing values on the attachment indicators were dependent on one's contact with the group; individuals with little contact were less likely to convey sentiments about the group. For

the MAR simulation, Table 1 shows that maximum likelihood estimators do yield less bias than traditional methods. In addition, the percentage of chi-square rejections are consistent with the first simulation, except that mean imputation χ^2 values were dramatically inflated, leading to rejection rates far above the nominal 5 per cent level. Both maximum likelihood estimators yielded mean parameter estimates that were virtually identical to the population values; bias did not exceed 0.3 per cent across all model parameters. In contrast, the standardised structural path coefficients for the three *ad hoc* methods showed substantial bias and were consistently lower in magnitude than the population values. Results from a repeated measures ANOVA indicated that the bias differences between the maximum likelihood and *ad hoc* methods were statistically significant, although this is not surprising given the power level achieved with 1,000 observations (200 replications \times five estimators). Consistent with previous research, bias was confined to the structural paths and did not impact the loadings to any great extent. This is problematic as the structural paths are typically the primary interest in structural equation modelling analyses.⁴⁶ Notice also that the level of bias was not trivial, and was either close to or exceeded the problematic criterion suggested by Muthén *et al.*⁴⁷

In some cases, bias may be tolerable if an estimator yields highly efficient parameter estimates. The efficiency of the estimator may compensate for bias, resulting in parameter estimates that are, on average, closer to the true population value compared to an unbiased but inefficient estimator. Thus, due to the bias noted above, it is appropriate to examine sampling variability around the population values rather than the mean of the empirical sampling distribution, as

done with the MCAR results. As such, the mean squared error (MSE) was calculated for each estimator as

$$MSE = \frac{\sum(\hat{\theta}_j - \theta_j)^2}{i}$$

where $\hat{\theta}_j$ and θ_j respectively represent a parameter estimate and the corresponding population value for the *j*th missing data estimator, and *i* is the number of iterations with admissible solutions. This statistic is essentially the sampling variance of the estimator in cases where parameter estimates are unbiased.

Table 2 gives the ratio of the MSE for each estimator relative to that of full information maximum likelihood. Consistent with the MCAR simulation, expectation-maximisation and full information maximum likelihood yielded nearly identical results. In virtually all cases the MSE values for the traditional methods were substantially larger than those of the maximum likelihood estimators. In a few cases MSE ratios close to unity were observed for certain factor loadings, which indicated that the traditional method yielded precision that was equivalent to the maximum likelihood estimators. Thus, it appears that the bias observed in the structural paths for the traditional methods is in no way compensated for by efficiency. As was the case in the MCAR simulation, the parameter estimates from any single sample were closer, on average, to the true population parameter when using maximum likelihood estimators; the difference between maximum likelihood and traditional estimators is substantial under MAR.

DISCUSSION

Missing data are inevitable, especially in large-scale, commercial marketing research surveys in which respondents

often choose not to answer due to fatigue, confusion or lack of knowledge or experience. The present results suggest that marketing researchers should pay greater attention to missing data assumptions as well as their choice of missing data correction mechanism.

Missing data estimators rely on different assumptions regarding the underlying cause of the missing data, and marketers should assess these assumptions when analysing incomplete data. While researchers routinely attend to statistical assumptions such as homogeneity of variance, missing data assumptions are seldom, if ever, addressed. Theory suggests that it is important that marketers identify whether the mean differences on Y_k are observed between those respondents who do and those who do not have missing values on a variable, X . If mean differences are observed, this suggests that the MCAR assumption may not be tenable. (For a detailed exposition of this type of diagnostic procedure, see Hair *et al.*⁴⁸ The SPSS MVA procedure provides a χ^2 test of the MCAR assumption, but no procedures are available for testing the MAR assumption.) Recall that if the MCAR assumption does not hold, listwise deletion and pairwise deletion are likely to yield biased parameter estimates.

Missing data mechanisms can substantially affect the accuracy of parameter estimates. While maximum likelihood missing data estimators may not necessarily yield unbiased estimation across all scenarios, this study suggests that these estimators are superior to traditional *ad hoc* methods. Even under strict MCAR conditions where *ad hoc* methods do not bias parameter estimates, maximum likelihood estimators are more efficient. Under maximum likelihood estimation, the parameter estimates from any single sample should, on average, be closer to the population estimate.

Moreover, these results suggest that the efficiency gain is not trivial — a 50 per cent increase in the listwise deletion sample size was required to yield sampling variability equivalent to that of maximum likelihood. In addition, data in commercial marketing research studies are often not completely missing at random (ie MCAR) due to structural skip patterns generally reflecting respondents' unfamiliarity with a block of questions. Since the MCAR assumption may not hold in many applied situations,^{49,50} most marketing researchers are facing MAR conditions and the choice of missing data technique becomes an important data analytic decision. Recall that under MAR, maximum likelihood estimators yield substantially less bias in parameter estimates than more traditional methods. For marketing practitioners, parameter estimate bias could substantially affect the interpretation of results. Important strategic decisions could be made on the interpretation of parameter estimates that are not truly representative of the population because of the choice of missing data correction mechanism.

While a large body of existing research has supported the use of maximum likelihood missing data estimators, there is still a lot of uncertainty associated with these techniques. For example, it was previously noted that maximum likelihood estimators are built on the assumption of multivariate normality. Even though Enders suggests that the estimator is superior to *ad hoc* methods under extreme non-normal conditions,⁵¹ more research needs to be done. In addition, much of the contemporary research has focused on full information maximum likelihood, leaving several methodological issues related to expectation-maximisation unresolved. For example, the expectation-maximisation

estimator can be used to impute missing values after the process has converged to a solution. Analyses of the imputed data will yield different results relative to the expectation-maximisation covariance matrix, but it is unclear to what extent, if any, the imputation process will bias results.

This research could be extended beyond the field of marketing research. For example, database marketing and customer relationship management systems with large sample sizes are often plagued with missing data due to systematic error in the capture, rejection and matching of data — rather than customer behaviour itself. Moreover, some researchers suggest that the MCAR and MAR assumptions may not hold in these situations.^{52–54} Thus, potential exists for the application and extension of this research to other fields of inquiry within the business discipline.

While many avenues for future research exist, there is little evidence to support the continued use of *ad hoc* missing data methods such as listwise deletion and pairwise deletion. Thus, it is recommended that marketing researchers be more aware of the potential impact of their missing data assumptions and decrease their reliance on *ad hoc* methods in favour of maximum likelihood estimators.

References

- Schafer, J. and Olsen, M. (1998) 'Multiple imputation for multivariate missing-data problems: a data analyst's perspective', *Multivariate Behavioral Research*, Vol. 33, No. 4, pp. 545–571.
- Malhotra, N. (1987) 'Analyzing marketing research data with incomplete information on the dependent variable', *Journal of Marketing Research*, Vol. 24, No. 1, pp. 74–84.
- Hair, J., Anderson, R., Tatham, R. and Black, W. (1998) 'Multivariate data analysis', 5th ed., Prentice Hall, Upper Saddle River, New Jersey.
- Kline, R. (1998) 'Principles and practice of structural equation modelling', Guilford, New York.
- Hartley, H. and Hocking, R. (1971) 'The analysis of incomplete data', *Biometrics*, Vol. 27, pp. 783–823.
- Muthén, B., Kaplan, D. and Hollis, M. (1987) 'On structural equation modelling with data that are not missing completely at random', *Psychometrika*, Vol. 52, pp. 431–462.
- Arbuckle, J. (1996) 'Full information estimation in the presence of incomplete data', in Marcoulides, G. and Schumacker, R. (eds) 'Advances in structural equation modelling', Lawrence Erlbaum Publishers, Mahwah, New Jersey, pp. 243–277.
- Finkbeiner, C. (1979) 'Estimation for the multiple factor model when data are missing', *Psychometrika*, Vol. 44, No. 4, pp. 409–420.
- Graham, J., Hofer, S. and MacKinnon, D. (1996) 'Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures', *Multivariate Behavioral Research*, Vol. 31, No. 2, pp. 197–218.
- Arbuckle, J. (1995) 'Amos user's guide', (computer software), Smallwaters, Chicago.
- Neale, M. (1995) 'Mx: statistical modeling', (computer software), 3rd edition, Department of Psychiatry, Medical College of Virginia, Virginia Commonwealth University, Richmond, Virginia.
- Arbuckle (1996) *op. cit.*
- Wothke, W. (2000) 'Longitudinal and multi-group modelling with missing data', in Little, T., Schnabel, K. and Baumert, J. (eds) 'Modeling longitudinal and multi-group data: Practical issues, applied approaches and specific examples', Lawrence Erlbaum Publishers, Mahwah, New Jersey.
- Dempster, A., Laird, N. and Rubin, D. (1977) 'Maximum likelihood from incomplete data via the EM algorithm', *Journal of the Royal Statistical Society*, Vol. 39, pp. 1–38.
- Graham, J. and Hofer, S. (1993) 'EMCOV reference manual', (computer software), University of Southern California, Institute for Prevention Research, Los Angeles.
- Schafer, J. (1998) 'NORM: multiple imputation of incomplete multivariate data under a normal model, version 2', (computer software), Windows 95/98/NT, <http://www.stat.psu.edu/~jls/misoftwa.html>.
- Little, R. and Rubin, D. (1987) 'Statistical analysis with missing data', Wiley, New York.
- Graham and Hofer (1993) *op. cit.*
- Schafer (1998) *op. cit.*
- Rubin, D. (1976) 'Inference and missing data', *Biometrika*, Vol. 63, pp. 581–592.
- Graham, Hofer and MacKinnon (1996) *op. cit.*
- Rubin (1976) *op. cit.*
- McAlexander, J. and Schouten, J. (1988) 'Brandfests: servicescapes for the cultivation of brand equity', in Sherry, J. (ed.) 'Servicescapes', AMA NTC Business Books, Chicago, pp. 377–401.
- Arbuckle (1996) *op. cit.*
- Enders, C. and Bandalos, D. (2001) 'The relative performance of full information maximum likelihood estimation for missing data in structural equation models', *Structural Equation Modeling*, Vol. 8, pp. 430–457.
- Wothke (2000) *op. cit.*

- 27 Enders, C. (2000a) 'The relative performance of missing data algorithms in structural equation models with nonnormal data', Paper presented at the American Educational Research Association Conference, April, New Orleans, Louisiana.
- 28 Enders, C. (2000b) 'The performance of maximum likelihood missing data algorithms for multiple regression', Paper presented at the American Educational Research Association Conference, April, New Orleans, Louisiana.
- 29 Arbuckle (1996) *op. cit.*
- 30 Enders and Bandalos (2001) *op. cit.*
- 31 Wothke (2000) *op. cit.*
- 32 Arbuckle (1996) *op. cit.*
- 33 Enders and Bandalos (2001) *op. cit.*
- 34 Kaplan, D. (1995) 'The impact of bib spiraling-induced missing data patterns on goodness-of-fit tests in factor analysis', *Journal of Educational and Behavioral Statistics*, Vol. 20, pp. 69–82.
- 35 Marsh, H. (1998) 'Pairwise deletion for missing data in structural equation models: Nonpositive definite matrices, parameter estimates, goodness of fit, and adjusted sample sizes', *Structural Equation Modeling*, Vol. 5, pp. 22–36.
- 36 Enders and Bandalos (2001) *op. cit.*
- 37 Enders (2000a) *op. cit.*
- 38 Prentice, D., Miller, D. and Lightdale, J. (1994) 'Asymmetries in attachments to groups and their members: Distinguishing between common identity and common bond groups', *Personality and Social Psychology Bulletin*, Vol. 20, pp. 484–493.
- 39 Zeithaml, V., Barry, L. and Parasuraman, A. (1995) 'The behavioral consequences of service quality', *Journal of Marketing*, Vol. 60, pp. 31–46.
- 40 Peters, C. and Grossbart, S. (2000) 'The good life: Communal factors and the expanded brand context in the Winnebago-Itasca travelers club', working paper, Department of Marketing, University of Georgia, Athens, GA 30602-6258.
- 41 Little and Rubin (1987) *op. cit.*
- 42 Muthén, Kaplan and Hollis (1987) *op. cit.*
- 43 Kaplan (1995) *op. cit.*
- 44 Enders and Bandalos (2001) *op. cit.*
- 45 Wothke (2000) *op. cit.*
- 46 Muthén, Kaplan and Hollis (1987) *op. cit.*
- 47 *Ibid.*
- 48 Hair *et al.* (1998) *op. cit.*
- 49 Graham, Hofer and MacKinnon (1996) *op. cit.*
- 50 Muthén, Kaplan and Hollis (1987) *op. cit.*
- 51 Enders (2000a) *op. cit.*
- 52 Bennett, G., Platts, G. and Crossley, J. (1996) 'Inferring the inferred?', *IMA Journal of Mathematics Applied in Business and Industry*, Vol. 7, No. 4, pp. 327–338.
- 53 Feelders, A. (1999) 'Credit scoring and reject inference with mixture models', *International Journal of Intelligent Systems in Accounting, Finance & Management*, Vol. 8, No. 4, pp. 271–279.
- 54 Hand, D. and Henley, W. (1993/4) 'Can reject inference ever work?', *IMA Journal of Mathematics Applied in Business and Industry*, Vol. 5, pp. 45–55.