THE ACCURACY OF PARAMETER ESTIMATES AND COVERAGE PROBABILITY OF POPULATION VALUES IN REGRESSION MODELS UPON DIFFERENT TREATMENTS OF SYSTEMATICALLY MISSING DATA

by

LUCAS ONYANGO A. OTHUON

B. Ed. (Hons.), The University of Nairobi, 1980
M. A., The University of British Columbia, 1993

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

Department of Educational Psychology & Special Education

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

Sep 18, 1998

© Lucas Onyango A. Othuon, 1998

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Educational Psych. & Special Education

The University of British Columbia
Vancouver, Canada

Date Sep 23 / 1998

ABSTRACT

Several methods are available for the treatment of missing data. Most of the methods are based on the assumption that data are missing completely at random (MCAR). However, data sets that are MCAR are rare in psycho-educational research. This gives rise to the need for investigating the performance of missing data treatments (MDTs) with non-randomly or systematically missing data, an area that has not received much attention by researchers in the past.

In the current simulation study, the performance of four MDTs, namely, mean substitution (MS), pairwise deletion (PW), expectation-maximization method (EM), and regression imputation (RS), was investigated in a linear multiple regression context. Four investigations were conducted involving four predictors under low and high multiple R^2, and nine predictors under low and high multiple R^2. In addition, each investigation was conducted under three different sample size conditions (94, 153, and 265). The design factors were missing pattern (2 levels), percent missing (3 levels) and non-normality (4 levels). This design gave rise to 72 treatment conditions. The sampling was replicated one thousand times in each condition. MDTs were evaluated based on accuracy of parameter estimates. In addition, the bias in parameter estimates, and coverage probability of regression coefficients, were computed.

The effect of missing pattern, percent missing, and non-normality on absolute error for $R^2_{estimate}$ was of practical significance. In the estimation of R^2, EM was the most accurate under the low R^2 condition, and PW was the most accurate under the high R^2 condition. No MDT was consistently least biased under low R^2 condition. However, with nine predictors under the high R^2 condition, PW was generally the least biased, with a tendency to overestimate population R^2. The mean absolute error (MAE) tended to increase with increasing non-normality and increasing percent missing. Also, the MAE in $R^2_{estimate}$ tended to be smaller under monotonic pattern than under non-monotonic pattern. MDTs were most differentiated at the highest level of percent missing (20%), and under non-monotonic missing pattern.

In the estimation of regression coefficients, RS generally outperformed the other MDTs with respect to accuracy of regression coefficients as measured by MAE. However, EM was competitive under the four predictors, low R^2 condition. MDTs were most differentiated only in the estimation of β_1, the coefficient of the variable with no missing values. MDTs were undifferentiated in their performance in the estimation for $b_2,\ldots,b_p$, $p = 4$ or 9, although the MAE remained fairly the same across all the regression coefficients. The MAE increased with increasing non-normality and percent missing, but decreased with increasing sample size. The MAE was generally greater under non-monotonic pattern than under monotonic pattern. With four predictors, the least bias was under RS regardless of the magnitude of population R^2. Under nine predictors, the least bias was under PW regardless of population R^2.

The results for coverage probabilities were generally similar to those under estimation of regression coefficients, with coverage probabilities closest to nominal alpha under RS. As expected, coverage probabilities decreased with increasing non-normality for each MDT, with values being closest to nominal value for normal data. MDTs were most differentiated with respect to coverage probabilities under non-monotonic pattern than under monotonic pattern.

Important implications of the results to researchers are numerous. First, the choice of MDT was found to depend on the magnitude of population R^2, number of predictors, as well as on the parameter estimate of interest. With the estimation of R^2 as the goal of analysis, use of EM is recommended if the anticipated R^2 is low (about .2). However, if the anticipated R^2 is high (about .6), use of PW is recommended. With the estimation of regression coefficients as the goal of analysis, the choice of MDT was found to be most crucial for the variable with no missing data. The RS method is most recommended with respect to estimation accuracy of regression coefficients, although greater bias was recorded under RS than under PW or MS when the number of predictors was large (i.e., nine predictors). Second, the choice of MDT seems to be of little concern if the proportion of missing data is 10 percent, and also if the missing pattern is monotonic rather than non-monotonic. Third, the proportion of missing data seems to have less impact on the accuracy of parameter estimates under monotonic missing pattern than under non-monotonic missing pattern. Fourth, it is recommended for researchers that in the control of Type I error rates under low R^2 condition, the EM method should be used as it produced coverage probability of regression coefficients closest to nominal value at .05 level. However, in the control of Type I error rates under high R^2 condition, the RS method is recommended. Considering that simulated data were used in the present study, it is suggested that future research should attempt to validate the findings of the present study using real field data. Also, a future investigator could modify the number of predictors as well as the confidence interval in the calculation of coverage probabilities to extend generalization of results.

TABLE OF CONTENTS

LIST OF TABLES

TABLE

LIST OF FIGURES

FIGURE

ACKNOWLEDGEMENTS

I would like to extend my sincere gratitude to members of my research committee, Dr. Nand Kishor (Research Supervisor), Dr. James Steiger, Dr. Don Allison, and Dr. Bob Conry, for their support in the development of this thesis. Without their commitment, nothing much could have been achieved.

My acknowledgements go to Maseno University College for according me study leave to pursue the program that has culminated in the writing of this dissertation, and to the Department of Educational Psychology and Special Education, University of British Columbia, for the academic and financial support during my stay in Canada. I must thank all my friends in the Department of Educational Psychology and Special Education, University of British Columbia, for their encouragement whenever the going was tough.

I would like to sincerely thank my wife, Akeyo, for taking good care of our lovely children, Akinyi, Ochieng', and Omondi, during the four years I was away. Nobody else could have done it better! And lastly, I dedicate this dissertation to my parents, Ayado and Wamwayi; although not formally schooled by default, they had a vision beyond measure.

CHAPTER 1

INTRODUCTION

Researchers employing multivariate data analysis techniques (e.g., multiple regression, discriminant analysis, and canonical correlation) commonly encounter the problem of dealing with missing data on one or more variables (Gleason & Staelin, 1975). In particular, the problem of missing data is most acute in field experiments (Lepkowski, Landis, & Stehouer, 1987) and surveys (Kim & Curry, 1977). With missing data, it may be necessary to apply a suitable treatment during the statistical analysis. The basic aim of treating missing data is to improve the accuracy of the parameter estimates of interest. This may reduce the likelihood of making a wrong statistical inference.

Missing data may be planned by a researcher or unplanned. The present study is primarily concerned with unplanned missing data. Common causes of unplanned missing data include omission and attrition. Omissions occur when some respondents fail to answer some items on a questionnaire. Such may occur in the middle of a survey or at the end. Omissions that occur in the middle of a survey instrument may be due to a respondent simply not seeing a question, or forgetting to go back to skipped questions. Attrition occurs when subjects drop out prematurely. Examples of studies where attrition is common include panel surveys or cohort studies with designs that involve repeated measurements. The resulting data are unbalanced, sometimes with unequal numbers of measures for each respondent.

The easiest method for handling missing data is to discard all cases with missing values. However, Raymond and Roberts (1987) noted a serious limitation of this approach is that it encourages researchers to overlook available data. Further, as the number of variables increases, increasing amounts of data are ignored, even if the total number of missing values remains constant.

Several techniques, referred to as missing data treatments (MDTs), have been proposed for handling missing data. One popular method is the pairwise deletion or piecemeal method. With this method, all available data are used to compute means and variances, while all available pairs of values are employed in the computation of covariances. Another common procedure is mean substitution in which missing values are replaced by the mean of the variable. In addition, numerous regression imputation procedures have been proposed. These methods make use of information present in the covariates to estimate the missing values for a variable of interest. The more complex procedures include the iterated linear regression methods, one form of which is known as expectation-maximization method (EM).

Listwise deletion, pairwise deletion, and mean substitution are already included in various statistical computer softwares. In the present study, I investigated the performance of commonly used MDTs that are popular in computer software, and also EM which is currently much talked about in electronic forums, SEMNET and AERA-D*. Important research issues discussed below guided the investigation.

Research Issues

There are a number of issues pertaining to current knowledge about MDTs that require attention. Firstly, whereas MDTs have been around for over six decades, the number of studies on some MDTs has remained inadequate. In particular, Roth (1994) noted that although the number of studies that have assessed the bias of simple MDTs (e.g., mean substitution and pairwise deletion) is substantial, research on more complex MDTs like EM is limited. The small number of studies on MDTs did confine comparisons of the EM method to listwise and pairwise deletion methods, and paid little or no attention to other treatments.

Few studies have systematically incorporated the various patterns of missing data, and most researchers have made use of randomly deleted data. Although most MDTs were developed for use with randomly missing data, such data sets are rare in psycho-educational research where missingness in one variable is likely to be related to measurements in another variable within the same data set. Only two studies that made use of non-randomly missing data were found in the

* SEMNET is an electronic forum for academic exchanges on topics related to structural equations modeling. AERA-D is an electronic forum for exchanges on methodological issues among members of the American Educational Research Association.

literature. Azen, Van Guilder, and Hill (1989) compared the performance of three approaches (complete cases, pairwise deletion, and EM algorithm) in estimating regression parameters and missing values under varying design conditions that included pattern of missing data. However, they did not investigate how the MDTs might affect the test of hypotheses about the regression coefficients. In addition, their simulation consisted of only 50 replications under design conditions whose selection criteria were arbitrary.

Kromrey and Hines (1994) compared the performance of MDTs by investigating the effects of systematically missing data on the parameter estimates of a multiple linear regression model. As in Azen et al., they did not investigate the relative performance of MDTs in hypothesis testing. Also, they used only a two-predictor regression model. Two predictor linear regression models are rarely found in applied research. There is need to compare different MDTs within or across data sets with different non-random patterns of missing data and under simulation conditions that are more representative of real data. The common procedure in most MDT studies is that samples with known statistical properties are randomly selected from a population, and portions of the samples are then randomly deleted and treated with various MDTs. For each MDT, the bias or accuracy of parameter estimates is assessed with respect to the population parameters.

Many of the probability statements yielded by methods of statistical inference are based on the assumption of normality. In particular, the OLS estimation of regression parameters is one of the statistical procedures most commonly understood to be affected by distribution assumptions (Mooney & Duval, 1993). To make inferential statements using OLS, we need to assume that the random residual in the model is normally distributed (Draper & Smith, 1981, p.23). This assumption is required because the sampling distribution of the OLS estimator is based on the random residual of the model. If we assume normality when non-normality actually holds, our hypothesis tests could lead to false rejection of the null hypothesis, that is, our confidence intervals and hypothesis tests could have a greater than nominal probability of Type I error. Alternatively, such hypothesis tests could lead to Type I error rates that are smaller than the nominal value, and this can be problematic if they come with a corresponding reduction in statistical power.

Whereas researchers have investigated the effectiveness of various MDTs under different sample sizes, proportion of missing data, size of correlation, and number of variables, very few studies have explicitly investigated the relative performance of MDTs under violation of multivariate normality. As observed by Fleishman (1978), many of the psychological variables found in practice are skewed and/or kurtotic to various degrees. In a study based on over 500 score distributions which were reasonably representative samples of psychometric, achievement, and ability measures, Micceri (1989) found that only 6.8% of the distributions exhibited both tail weight and symmetry approximating that expected of normal distributions. In addition, 100% of the distributions were significantly non-normal at an alpha level of .01. This shows how prevalent non-normal distributions are in psycho-educational research and the importance of investigating how MDTs might function under non-normality of distributions.

Current Study

Previous researchers have investigated the accuracy of parameter estimates after different treatments of missing data. However, most of these studies have tended to focus more on normally distributed data sets in which data were missing completely at random (MCAR). Little is known about the bias and accuracy of parameter estimates when different MDTs are used to treat systematically missing data. Also, little is known about the performance of MDTs under violation of normality. In addition, researchers have tended to ignore investigating how hypothesis testing might be affected under various design conditions. In the present study, systematically missing data were used, and non-normality was incorporated as a factor in the design.

The present study was based on multiple linear regression models in which data were missing only in the predictors (X's). The related problem of missing values in the outcome variable (Y) is less interesting in the sense that if the predictors are complete, then the missing cases contribute no information to the regression equation of Y on the set of predictors under conditions of MCAR (cf. Little, 1992). Multiple regression models have been used in a majority of previous studies. Raymond and Roberts (1987), Beale and Little (1975), and Little (1992) compared MDTs in a linear regression context. With respect to non-randomly missing data, Azen, Guilder, and Hill (1989) and Kromrey and Hines (1994) also investigated the effectiveness of MDTs in a multiple linear regression context. Further, as noted by Little (1992), the focus on regression models is justified by the fact that many studies in social and behavioral sciences often make use of these models. Considering that regression imputation and the EM method have not received much attention, I investigated the accuracy and bias of parameter estimates, as well as the coverage probability for population regression coefficients under expectation-maximization method (EM), mean substitution (MS) method, pairwise deletion (PW) method, and simple regression imputation (RS) method.

Shchigolev (1960/1965) noted that the two most useful measures in the evaluation of errors are the mean absolute error of estimation (MAE) and the mean square error of estimation (MSE). Note that the MAE is not the same as the mean absolute deviation of the parameter estimate, and the MSE is not the same as the standard deviation of the parameter estimate. The difference arises from the fact that for mean absolute deviation and standard deviation, the deviations are from the mean. However, for MAE and MSE, the deviations are from the true population value. Both the MAE and the MSE provide information on the spread (closeness) of the parameter estimates around the true population value, the only advantage of MSE over MAE being that MSE is more mathematically tractable, an advantage that was not of necessity in the present study. Therefore, the MAE was used to evaluate the accuracy of parameter estimates in the present study. The MAE was obtained by first computing the absolute error of estimation, defined as the absolute deviation of each parameter estimate from the model value, then calculating the mean of the absolute values. In addition, as suggested by Judge, Hill, Griffiths, Lutkepöhl, and Lee (1988), an optimal performance was obtained when both MAE and bias of parameter estimates were smallest. Therefore, as in Anderson, Stone-Rumero, and Tisak (1996), the bias of parameter estimates, defined as the difference between the true parameter (that derived from the population model) and the mean of the estimators was also computed. The following research questions guided the conduct of the present study:

1. What are the effects of missing pattern, percent missing, and non-normality on absolute error of parameter estimates of a multiple regression model (MRM) when missing data are treated with selected MDTs?
2. What is the relative performance of MDTs with respect to accuracy and bias of parameter estimates under various conditions?
3. What is the relative performance of MDTs with respect to coverage probabilities of regression coefficients at the conventional alpha level of .05?

In real world settings, missing data are the rule rather than the exception (Kolb & Dayton, 1995). The common practice of deleting cases with missing observations is of questionable validity since, typically, there is little evidence supporting that the data are missing randomly. Unfortunately, researchers have continued to give greater attention to evaluating MDTs under MCAR condition, simply because commonly used MDTs are based on the MCAR assumption. Whether data sets violate the MCAR assumption or not, researchers will continue using MDTs. It is therefore important to evaluate MDTs using non-MCAR data, as this is what researchers commonly encounter in the field. Consequently, the researchers would be able to make informed judgment concerning which MDT to apply for various types of data. This may improve the accuracy of decision-making in research, with the possibility of influencing practice in psycho-educational settings.

The EM method of treating missing data is based on normal theory. However, most field data are non-normal (Micceri, 1989). By employing simulation methodology, it would be possible to provide evidence of conditions under which normal theory based EM method breaks down. Such findings would be of importance to practitioners in understanding when not to apply the EM method.

CHAPTER 2

LITERATURE REVIEW

This chapter begins with a review of types of missing data. This is followed by a description of methods of treating missing data, and an analysis of previous studies involving MDTs that were identified through manual and electronic searches. The databases searched included ERIC: Educational Resources, Education Index, and PsychInfo. The search began by using the title words "missing data" and "incomplete data." This led to a number of books and articles with the following descriptors: statistical data, statistical regression, multiple regression, models, maximum likelihood, mathematical modeling, multivariate analysis, and algorithms. These descriptors were used for a more in-depth search.

Types of Missing Data

According to Cochran (1983), missing data may be of two types, namely, unit non-response or item non-response. This section describes these two types of missing data.

Unit Non-response

Sometimes, none of the variables of interest is measured for a unit or subunit in a survey. For example, the mailed questionnaire may not be returned, or the interviewer may not find anyone at home. Cochran (1983) called this type of non-response "unit non-response." Unit non-response occurs if a unit is selected for the sample and is eligible for the survey, but no response is obtained for the unit or the obtained response is unusable. Causes of unit non-response include language problems, total refusal to respond, or the responses given are classified as unusable.

The problem of how to handle unit non-response goes back to the 1930s, the same period when many of the standard techniques and results in probability sampling were developed (Cochran, 1983). Quota sampling is an early example of a group of methods whose aim was to cut down or eliminate non-response when data are being collected in the field (King, 1983). This method is based on full probability sampling of a target population. The method was developed to cut down on field costs by taking stratified samples for public opinion surveys. It enables one to obtain a desired number of completed cases in a relatively short period of field time. Sudman (1967) noted that the method involves drawing sampling locations down to the block level with exactly the same technique of multistage selection with probabilities proportional to size as in full probability sampling.

Missing data are encountered in randomized experiments as well, and the development of methods for making estimates of the effects of the treatment from experiments with missing data also began in the 1930s. According to Cochran (1983), the technique most widely used in field experiments applies the standard least squares model for the analysis of the complete results of field experiments to the missing data that are present. This is an approach that makes use of "filling in" or "imputing" values for all the missing observations by formulae so that the standard analysis of the completed data gives the same estimates as the least squares analysis of the missing data. This approach does not require that the units (e.g., respondents) for whom the data are missing be a random subsample of the units in the experiment (Cochran, 1983).

Item Non-response

In a survey, it sometimes occurs that for certain questions, either no answer is given or the answer is judged to contain a gross error and is deleted during editing. Usually, such questions are sensitive. Examples of such questions may concern income, or when the respondent does not have the information. This type of missing data is called an item non-response (Cochran, 1983).

Cochran (1983) provided a more detailed explanation of circumstances that may lead to item non-response. He noted that item non-response occurs if questions that should be answered are not answered, or if the answers are classified as unusable. Sometimes, item non-response occurs because response to an instrument or question is broken off after being partly completed, but the partial response is not classified as a unit non-response. Item non-response may occur because the respondent does not have the information needed for one or more questions, because the respondent refuses to answer specific questions, or because the interviewer or respondent skips the question. Item non-response may occur for blocks of questions: an interviewer may miss a branch point in an interview, or a respondent may refuse to answer all questions on a specific subject, say, income or the final questions in an interview. Such data sets tend to display non-random patterns because the cause of missingness is related to the measurement process and instrument design.

According to Cochran (1983), the earliest paper on item non-response was that of Wilks in 1932. In that paper, a random sample from a bivariate normal distribution was considered. Wilks (1932) considered the following situation under MCAR condition. Suppose in n_{12} observations, X_1 and X_2 were both measured. Also, suppose that in n_{10} observations, X_1 is measured but X_2 is missing. Further, suppose that in n_{02} observations, X_2 is measured but X_1 is missing. Wilks found maximum likelihood estimates of the parameters μ_1, μ_2, σ_1, σ_2, and ρ, and gave large sample formulas for the variances and covariance of $\hat{\mu}_1$ and $\hat{\mu}_2$. However, he noted that his estimates were not particularly simple to compute. Therefore, he proposed mean substitution as a simpler method for treating missing data. In this case, if the n_{02} units in which X_1 was missing were a random sample, the mean of all the $n_{12} + n_{10}$ measured values of X_1 is used as an estimate of μ_1.

Methods of Treating Missing Data

Existing techniques of handling unplanned missing data may be classified into two broad categories: deletion methods and imputation (filling in) methods.

Deletion Methods

Listwise Deletion. This method, also referred to as complete case analysis, is the most obvious method of handling missing data. The method involves discarding cases with any missing values. In this approach, entire data records with missing data are deleted. Advantages of listwise deletion are ease of implementation and the fact that valid inference is obtained when missingness depends on the regressors (Little, 1992). Other advantages are that it always generates consistent covariance and correlation matrices, and test statistics used with complete data can be used without modification (Kim & Curry, 1977).

Although listwise deletion is commonly used (Gilley & Leone, 1991), a serious limitation of the approach is that it encourages researchers to overlook available data. Also, it sacrifices huge amounts of data (Reinecke & Schimdt, 1996; Malhotra, 1987; Stumpf, 1978; Little, 1992), with the possible result of a decrease in power (Gilley & Leone, 1991) and an introduction of bias in parameter estimation (Donner, 1982; Little & Rubin, 1987). If the number of predictors is large, then even a sparse pattern of missing data in the predictors can result in a substantial number of missing cases (Little, 1992). Researchers should therefore avoid the use of listwise deletion as a treatment method for missing data as it involves the complete loss of individual cases. If the MCAR (missing completely at random) assumption is not fulfilled, the values of the covariance matrix based on listwise deletion procedure are underestimated (Reinecke & Schimdt, 1996).

Pairwise Deletion. In this method, all possible cases with valid data are used to calculate means and variances, while all available pairs of values are used in the computation of covariances (Little & Rubin, 1987). As noted by Reinecke and Schimdt (1996), this method provides more benefit than the listwise deletion because it uses all available information. Like in listwise deletion, pairwise deletion assumes that missing data are MCAR. When the sample size is small and the number of cases with missing information is high, or when the predictor variables are highly correlated, the resulting covariance matrix of the predictors generated by pairwise deletion of missing data may not be positive-definite. A symmetric matrix **A** and its associated quadratic form are positive definite if $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ for all nonnull **x**. If a covariance matrix is not positive-definite, regression coefficients based on it are indeterminate (Kim & Curry, 1977; Little, 1992), because all required computations are not possible.

In contrast to listwise deletion, the values of the covariance matrix based on pairwise deletion method are overestimated. Comparisons as to which of the matrices give more consistent and efficient results are lacking in the literature (Allison, 1987; Brown, 1994).

Imputation Methods

Mean Substitution. One of the simplest techniques used in the imputation of missing data is the mean substitution method, also known as unconditional mean substitution. This method, proposed by Wilks (1932), simply substitutes missing values by the mean of the non-missing values for that variable.

Although mean substitution preserves data and is easy to use, it also tends to attenuate variance estimates (Roth, 1994). Computed variance estimates decrease as more means are added to calculations. For example, a researcher may have 40 subjects, but 5 have missing data. Mean substitution might suggest that we add 5 means to the 35 scores. This would increase the N in the calculation of variance, but would not increase the deviations around the mean added by the 5 additional cases. Subsequently, covariance estimates are also attenuated (Little, 1988; Malhotra, 1987). Researchers may also believe they have more degrees of freedom than is warranted, because substituted means are not independent from other observations in the data.

Regression Imputation. Regression imputation is yet another method for treating missing data. This method, proposed for the first time by Buck (1960), begins with the estimation of the mean and covariance matrix from the sample mean and covariance matrix based on complete cases. These estimates are then used to regress the variable with missing data on a variable with complete data. Using the resulting linear regression equation, missing data are imputed case by case. Put differently, this method computes predicted estimates for any missing entry by using a linear regression equation.

Expectation-Maximization (EM) Method. The EM method for estimating missing data is one of the various maximum likelihood (ML) imputation approaches. ML methods for imputing missing data generally assume that the observed data are a sample drawn from a multivariate normal distribution and MCAR (DeSarbo et al., 1986). Little and Rubin (1987) described one such method for general patterns of missing data with ignorable non-response. In this method, it is assumed that the data are MAR and the objective is to maximize the likelihood function. If the likelihood is differentiable and unimodal, maximum likelihood estimates can be found. However, if this is not possible, then iterative methods like the Newton-Raphson algorithm or the method of scoring can be applied. Both these methods involve calculating the matrix of second derivatives of the loglikelihood. For complex patterns of incomplete data, the entries in this matrix tend to be complicated functions. As a result, to be practicable the methods can require careful algebraic manipulations and efficient programming.

Little and Rubin (1987) noted that an alternative computing strategy for incomplete-data problems, which does not require second derivatives to be calculated or approximated, is the Expectation-Maximization (EM) method. This method is remarkably simple, both conceptually and computationally. Graham, Hofer, and Piccinin (1994) stressed the importance of the fact that different EM methods are required for different kinds of analysis. However, they noted that in many analyses involving the general linear model, the iterative EM method could be extensively useful because the covariance matrix can be used as input. The iterative EM method replaces missing values by estimated values, estimates parameters, reestimates the missing values assuming the new parameter estimates are correct, reestimates parameters, and so forth, iterating until convergence (i.e., until there is no improvement in the estimated parameters). This procedure is available in BMDP AM (Dixon, 1988). An important difference between the scoring algorithm and the EM method is that the former requires inversion of the information matrices of μ and Σ at each iteration.

The EM method consists of two steps: an E step (estimation step) and an M step (maximization step). According to Little and Rubin (1987), the E step is used to calculate the conditional expectations of the "missing data" given the observed data and current estimated parameters. The calculated conditional expectations are then substituted for the "missing data." The quotations around "missing data" imply that EM does not necessarily substitute the missing values. The M step of the EM algorithm performs maximum likelihood estimation of θ just as if there were no missing data, that is, as if they had been substituted with estimates at the E step. The mathematical treatment of iterative EM algorithm as used in the present study, and discussed in greater detail in Little (1983) and Little and Rubin (1987), is as follows:

Suppose that the hypothetical complete data $y = (y_{ij})$ is a random sample from the K-variate normal distribution, with mean μ and covariance matrix Σ. This belongs to the regular exponential family

$$f(y_s, \theta) = \frac{B(y_s)\exp\{\theta^T t(y_s)\}}{A(\theta)}$$

where A and B are scalar functions, θ is an r by 1 vector of parameters, and $t(y_s)$ is an r by 1 vector of complete data sufficient statistics. In this case, the sufficient statistics $t(y_s)$ is given by the sample means

$$\bar{y}_j = \frac{1}{N_s}\sum_{i=1}^{N_s} y_{ij}, \qquad j = 1, \ldots, K,$$

and the uncorrected sample sum of squares and cross-products matrix d, with (j,k) th element

$$d_{jk} = \sum_{i=1}^{N_s} y_{ij} y_{ik}, \qquad j,k = 1,\ldots,K.$$

The complete data maximum likelihood estimates are given by

$$\hat{\mu}_j = \bar{y}_j, \qquad j = 1,\ldots,K,$$

$$\hat{\sigma}_{jk} = \frac{d_{jk}}{N_s} - \bar{y}_j \bar{y}_k, \qquad j,k = 1,\ldots,K.$$

Now suppose we have incomplete data p. The E step of the EM algorithm consists in finding expected values of the sufficient statistics:

$$\bar{y}_{ja} = E(\bar{y}_j | p; \mu_a, \Sigma_a)$$

$$d_{jka} = E(d_{jk} | p; \mu_a, \Sigma_a)$$

where μ_a and Σ_a are current estimates of μ and Σ. The M step consists in forming new estimates μ_b, Σ_b by substituting $\bar{y}_{ja}$ and d_{jka} for $\bar{y}_j$ and d_{jk}, respectively. It remains to describe the calculation of $\bar{y}_{ja}$ and d_{jka}. This clearly depends on the pattern of missing data. Note that

$$\bar{y}_{ja} = \frac{1}{N_s} \sum_{i=1}^{N_s} E(y_{ij} | p; \mu_a, \Sigma_a)$$

and

$$d_{jka} = \sum_{i=1}^{N_s} E(y_{ij} y_{ik} | p; \mu_a, \Sigma_a).$$

To calculate the terms in these sums, consider unit i with present items p_i and missing items m_i. With independent observations, given (μ_a, Σ_a),

$$E(y_{ij}|p;\mu_a,\Sigma_a) = E(y_{ij}|p_i,\mu_a,\Sigma_a) = \hat{y}_{ija}$$

and

$$E(y_{ij}y_{ik}|p;\mu_a,\Sigma_a) = E(y_{ij}y_{ik}|p_i;\mu_a,\Sigma_a)$$

which is equal to

$$\hat{y}_{ija}\hat{y}_{ika} + Cov(y_{ij}, y_{ik}|p_i;\mu_a,\Sigma_a).$$

If y_{ij} is present, that is, belongs to p_i, then y_{ija} is simply the observed value y_{ij}, and $Cov(y_{ij,}y_{ik})|p_i;\mu_a,\Sigma_a)$ is zero. If y_{ij} is missing, then y_{ija} is the predicted value of y_{ij} from a linear regression of y_j on the variables present in unit *i*, with intercept and regression coefficients calculated from μ_a and Σ_a. If y_{ij} and y_{ik} are both missing, $Cov(y_{ij}, y_{ik}|p_i;\mu_a,\Sigma_a)$ is the residual covariance (or if $j = k$, variance) from the regression of y_j and y_k on the variables present in unit *i*, calculated from Σ_a. Summing over units,

$$\bar{y}_{ja} = \frac{1}{N_s}\sum_{i=1}^{N_s}\hat{y}_{ija},$$

and

$$d_{jka} = \frac{1}{N_s}\{\hat{y}_{ija}\hat{y}_{ika} + Cov(y_{ij}, y_{ik}|p_i;\mu_a,\Sigma_a)\}.$$

Hence, the complete EM algorithm is as follows:

(i) Calculate initial estimates μ_a and Σ_a.

(ii) Complete the data matrix by calculating

$$\hat{y}_{ija} = \begin{cases} y_{ij} & if \quad y_{ij} \quad is \quad present \\ E(y_{ij} | p_i ; \mu_a , \Sigma_a) & if \quad y_{ij} \quad is \quad not \quad present \end{cases}$$

where $E(y_{ij} | p_i ; \mu_a , \Sigma_a)$ is obtained by the linear regression of y_j on variables present in unit i, with coefficients calculated from μ_a and Σ_a.

(iii) Form means and sums of squares cross product matrix of the completed data matrix; whenever y_{ij} and y_{ik} are missing add to the (j, k)th element of the sums of squares cross product matrix the residual covariance (variance if $j = k$) of y_i and y_k given the variables present in unit i, calculated from Σ_a.

(iv) Calculate new estimates μ_b and Σ_b from the means and adjusted cross-products matrix, and iterate until convergence.

Little and Rubin (1987) suggested four options of obtaining initial values of the parameters. The options are: (1) to use listwise deletion; (2) to use pairwise deletion (all-value); (3) to form the sample mean and covariance matrix of the data filled in by one of the imputation methods (e.g., linear regression); and (4) to form means and variances from observed values of each variable and set all correlations equal to zero. Option 1 provides consistent estimates of the parameters if the data are MCAR and there are at least K + 1 complete observations. Option 2 makes use of all the available data but can yield an estimated covariance matrix that is not positive definite, leading to problems in the first iteration. Options 3 and 4 generally yield inconsistent estimates that are positive semidefinite and hence usually workable as initial estimates. Based on the characteristics of the four options, option 3 (linear regression) was chosen for computing the initial value for use in iterative EM method.

Previous Research on MDTs

A number of factors were identified in the literature that may influence the performance of MDTs. These factors include pattern of missing data, sample size, proportion of missing data, magnitude of correlations among dependent and independent variables, number of predictors, and non-normality. However, the effect of these factors on absolute error in parameter estimates and control of Type I error rates when various MDTs are used in the treatment of missing data, has not received adequate attention. In the following section, important findings from previous studies on the effects of some of the factors are presented. Strengths and weaknesses in these studies, if notable, are also highlighted.

Proportion of Missing Data

Researchers should consider a number of factors when choosing a missing data treatment. One of the most obvious factors is the amount of missing data. Generally, the choice of MDT is not critical if the amount of missing data is small (Frane, 1976; Gilley & Leone, 1991). Monte Carlo studies suggest there is little difference in the parameter estimates and answers to research questions when less than 10% of the data are missing in random patterns (Raymond & Roberts, 1987) or systematic patterns (Malhotra, 1987).

The choice of MDTs seems to become more important as the amount of missing data approaches 15 - 20% of the data set (Raymond & Roberts, 1987) and most important as missing data approaches 30-40% (Malhotra, 1987). However, when missing data approaches 30 - 40% of the data set, one might question the wisdom of conducting any analyses (Roth, 1994). Therefore, in this study, use was made of data sets in which 10 - 20% was missing.

Violation of Normality

A regression model often used in research settings takes the form

$$Y = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_p X_{pi} + \varepsilon_i$$

where Y is the dependent variable, β_0 is an intercept, $\beta_1,...,\beta_p$ are unknown regression coefficients for $X_{1i},...,X_{pi}$, respectively, and ε_i is a random residual for the i^{th} observation, i = 1, 2,...,n. Dillon and Goldstein (1984) suggested that if we are only concerned with obtaining "good" estimators of regression coefficients, then the following four assumptions are necessary:

1. The expected value of the residual vector ε_i is zero, i.e., $E(\varepsilon) = \mathbf{0}$.
2. There is no correlation between the i^{th} and j^{th} residual terms, and the residuals exhibit constant variance, i.e., $E(\varepsilon\varepsilon') = \sigma^2\mathbf{I}$. In scalar notation, this corresponds to the following two assumptions: $E(\varepsilon_i\varepsilon_j) = 0$ for $i \neq j$ and $E(\varepsilon_i\varepsilon_j) = E(\varepsilon_i^2) = \sigma^2$ for $i = j$. These two assumptions are better known as the assumptions of *no serial correlation* and *homoscedasticity*.
3. The covariance between the X's and the residual terms ε_i is zero, an assumption that is automatically fulfilled if the X variables are nonprobabilistic, so that the $n \times p$ data matrix $\mathbf{X}$ consists of fixed numbers.
4. The rank of the data matrix $\mathbf{X}$ is p, the number of columns in $\mathbf{X}$, and less than n, the number of observations, i.e., r($\mathbf{X}$) = p, where $p < n$. This assumption means that there are no exact linear relationships among the X variables. It is better known as the assumption of *no multicollinearity*.

However, typically we want to go beyond mere point estimation and make statements or inferences about population regression coefficients on the basis of sample coefficients; or, more generally, we want to use the sample-based regression model to draw inferences about the population model. Therefore, we need to impose distribution assumptions on the residual terms.

Because the ordinary least squares (OLS) estimators are linear functions of ε_i (since Y_i is a linear function of ε_i), the probability distributions of the OLS estimators will depend on the specification of the probability distribution of ε_i. The classical multiple linear regression model assumes that each ε_i is normally distributed with mean $\mathbf{0}$, common variance $\sigma^2\mathbf{I}$, and uncorrelated residuals.

Given normality, each element of **b**, the OLS estimator of β, is normally distributed with mean equal to the corresponding element of the true β and variance given by σ^2 times the appropriate diagonal element of the inverse matrix $(\mathbf{X'X})^{-1}$. Although it is assumed that the matrix **X** is fixed, yielding different values in repeated sampling, Dunteman (1984) noted that this assumption can be relaxed to provide for predictors that are randomly selected (called stochastic regressors) for each sample replication and the classical model based upon the assumption of fixed regressors will still apply assuming that **X** is multivariate normal. That is, the parameters are estimated in exactly the same way, and hypothesis testing of these parameters proceeds identically.

Few studies have specifically evaluated the effectiveness of MDTs when the assumption of normality is violated. Non-normality may affect the performance of MDTs, especially with large proportions of missing data. In their study, Graham, Hofer, and MacKinnon (1995) found that estimates obtained for skewed data were somewhat less accurate than for normal data, and that the expectation-maximization (EM) estimates were satisfactory when skew was in the interval $\pm$ 3.

Missing Pattern

Simonoff (1988) and Little and Rubin (1987) stated that heuristic methods of treating missing data require the assumption that data are missing completely at random (MCAR). However, most field data do not miss randomly, and very little is known about the effectiveness of MDTs for such data sets (Little, 1992; Roth, 1994). Graham, Hofer, and Piccinin (1994) noted that non-randomly missing data occur when some variables are correlated with some important attribute of the individuals tested with those measures. Examples of non-randomly missing data include situations when respondents have trouble understanding the meaning of a question and skip it, and omissions due to failure to complete the survey. Such omissions normally form clusters at the middle or end of a questionnaire, with the possibility of forming a non-monotonic pattern of missing data. Another example of non-randomly missing data pattern encountered in the analysis of social science data is the monotonic pattern. Such a pattern is normally a result of attrition in panel survey. In such a pattern, for $j = 1,..., K-1$, X_j is observed whenever X_{j+1} is observed. Further elaboration on monotonic and non-monotonic patterns is on page 48.

Kromrey and Hines (1994) and Brockmeier, Kromrey, and Hines (1996) compared the effectiveness of five and eight MDTs, respectively, for non-randomly missing data within the context of two-predictor regression models in which missing data occurred on only one of the predictors. They found that existing MDTs that are intended for use with randomly missing data might be used with non-randomly missing data as well. However, they cautioned researchers to be careful in the selection of MDTs as there were substantial differences in their effectiveness across different types of data. For example, it is not clear how MDTs perform under various conditions of non-normality. Further, the researchers' results were limited to the two-predictor regression model in which systematically missing data occurred on only one of the predictors. They recommended that extensions to regression models with more predictors and to data sets with missing data on more than one predictor needed additional research. The present study was an attempt to address these issues.

The pattern of missing data is closely related to what Little and Rubin (1987) called the mechanism of missing data. The key issue in understanding the mechanism of missing data is to establish whether missingness is related to the data values (Little, 1992). According to Little and

Rubin (1987), knowledge (or absence of knowledge) of the mechanism that led to certain values being missing is a key element in choosing an appropriate analysis and in interpreting the results. For example, when the probability that X_1 is missing for a case is independent of data values, then the mechanism is said to be both missing completely at random (MCAR) and missing at random (MAR). However, when the probability that X_1 is missing for a case depends on the value of $X_2,..., X_p$ for that case, then the mechanism is said to be MAR and not MCAR. Further, when the probability that X_1 is missing depends on the value of X_1 for that case, then the mechanism is not MAR because X_1 is not fully observed. Engel and Meyer (1996) noted that a major problem in any attempt to handle missing data in survey research is to cope with nonignorable non-response. Nonignorable non-response arises whenever the missing data pattern is neither missing completely at random (MCAR) nor missing at random (MAR).

Little (1992) provided a more formal definition of missing data mechanism in the following manner. Let Z denote the *n* by (*p* + 1) data matrix, including observed and missing values, and let Z_{obs} denote the set of observed values of Z, and let Z_{mis} denote the set of missing values. By introducing a missing-data indicator matrix R, with (*i, j*)th element $R_{ij}=1$ if X_{ij} is observed and $R_{ij}=0$ if X_{ij} is missing, the notion of a missing-data mechanism is formalized in terms of a model for the conditional distribution $p(R|Z,\varphi)$ of R given Z, indexed by unknown parameters φ. Data are missing at random (MAR) if the distribution depends on the data Z only through the observed values Z_{obs}; that is, $p(R|Z,\varphi) = p(R|Z_{obs},\varphi)$ for all Z. Data are missing completely at random (MCAR) if the distribution of R does not depend on the observed or missing values of Z; that is, $p(R|Z,\varphi) = p(R|\varphi)$ for all Z.

According to Graham, Hoffer, and Piccinin (1994), most mechanisms of missing data may be classified as either "accessible" or "inaccessible." A missing data mechanism is accessible when the cause of missingness has been measured and is available for use in the analysis. If a researcher includes a measure of causes of missingness during data collection, and draws a random sample of the cases with missing data after data collection, then the researcher can use this information to account for missingness. In this case, the mechanism of missingness is accessible. Graham et al. suggested that an instrument for measuring the mechanism of missingness should include items on the reading speed of the participants, the participant's lack of motivation to complete survey, the participant's rebelliousness, and whether the participant refuses to participate because of the nature of scores on the dependent variable. However, researchers have not followed this approach.

Graham et al. gave two possible ways in which inaccessible missing data mechanisms can arise. Firstly, such mechanisms can arise when the variable containing the missing data itself is the cause of missingness. For example, the mechanism would be inaccessible if the people who drop out of a drug use prevention study do so because they currently are high-level drug users. Secondly, inaccessible mechanisms also can arise if another unmeasured variable is the cause of missingness and that variable is correlated with the one containing the missing data.

On the other hand, Little and Rubin (1987) referred to "accessible" missing data mechanisms as "nonignorable." However, according to Graham et al., the term "accessible" refers strictly to the mechanism and the term "ignorable" refers to a combination of the mechanism and the analysis used. Knowledge of the mechanism of missing data is important because there may be a need for the inclusion of that mechanism of missingness in the statistical model by including a distribution for response indicator variables that take the value of 1 if an item is recorded and the value of 0 otherwise (Graham et al.).

Little and Rubin (1987) described three examples of planned mechanisms of missing data in survey research. These are the processes of sample selection, double sampling, and censoring. Graham et al. described two unplanned mechanisms of missing data in survey research which are of interest in this study; omission and attrition.

Omission is a mechanism of unplanned missing data. The occurrence of such omissions may be somewhere in the middle of the survey instrument or at the end. Omissions that occur in the middle of a survey instrument may be due to a respondent simply not seeing a question, or forgetting to go back to skipped questions. A respondent may have trouble understanding the meaning of the question and may skip it. A participant may fail to answer a particular question because s/he is afraid of possible negative consequences of answering it, or because the question evokes negative feelings he or she does not want to experience. An omission may also be due to failure to complete the survey. Graham et al. (1994) noted that if time is not a limiting factor, then two main reasons for an incomplete survey are lack of ability and lack of motivation. A participant may lack the ability to finish because he or she is a slow reader or because of language problems.

Attrition is yet another mechanism of unplanned missing data. It occurs when a respondent is measured in an attribute for at least one wave of measurement, but is absent entirely for additional waves of measurement. Attrition may be due to a random or non-random process. An example of a non-random cause of attrition is when a subject is ill for the measurement session. In this case, the cause of attrition is completely independent of the measurement. Underlying values not observed in a given data set may determine whether information is missing or not. For example, in an attitudinal survey, a respondent with low cognitive skills may refuse or be unable to give answers to many questions. In this case, the pattern of missing data will exhibit clustering. Such a cause of attrition is not completely independent of the measurement (Kim & Curry, 1977).

There is also the possibility of having a non-random cause of attrition related to the dependent variable itself. For example, students with disability may be more likely to drop out of a study than are students who are not disabled. The present study involved non-randomly missing data structures in which missingness may have been through omission or attrition.

Studies with Randomly Missing Data

Limited number of studies using randomly missing data suggests that EM method produces less biased estimates than listwise deletion (Graham & Donaldson, 1993). Haitovsky (1968) conducted a simulation study with highly correlated data and found the method of pairwise deletion to be markedly superior to listwise deletion. Kim and Curry (1977) found a similar outcome for weakly correlated data as well.

Timm (1970) used simulation method to assess the ability of three methods in predicting correlation and variance-covariance matrices for observations missing at random. Instead of randomly generating numerous matrices, numbers were selected for use from the literature. With $\Sigma = \rho$, Kaiser's (1968) measure of average intercorrelation $\tau = \dfrac{\alpha_1 - 1}{p - 1}$, guided the selection, where α_1 is the largest eigenvalue of ρ and p is the number of variables. Values of α range from zero to unity. Three matrices for α = 0.2, 0.5, and 0.8 were chosen with p = 2, 5, and 10.

Having obtained the variance-covariance matrices from the literature, Kaiser and Dickman (1962) procedure was used to generate complete data matrices of size N = 50, 100, and 200 from a multivariate normal population with $\mu = 0$ and $\Sigma = \rho$. From these complete data matrices, two incomplete data matrices were obtained each with 1%, 10%, and 20% of the data deleted at random. Using these data with two replicates, the effect of sample size, number of variables, and average intercorrelation were examined when three MDTs were used. The three methods investigated were mean substitution, regression technique, and a modified form of Dear's (1959) technique in which the matrix Y is decomposed into its known and unknown components afterwhich a principal components solution is used to estimate the missing values. A major weakness of the study is that only two replications were used! This makes it difficult to have a reasonable degree of confidence on the results of the simulation study because of lack of accuracy.

Timm (1970) found no uniformly best technique for the estimation of variance-covariance matrices when observations were missing at random. However, the Dear and regression methods were generally superior to the mean substitution method for high and intermediately correlated variables, but less satisfactorily under low intercorrelation of variables.

Timm (1970) observed that in the comparison of techniques for the estimation of Σ, the Euclidean norm $\left\|\Sigma - \hat{\Sigma}\right\|^2$, an additive function of eigenvalues, provides a convenient measure of "closeness" and is defined as $\left\|\Sigma - \hat{\Sigma}\right\|^2 = Tr\left[(\Sigma - \hat{\Sigma})(\Sigma - \hat{\Sigma})'\right]$. Because the distribution of the Euclidean norm measure is unknown, Timm (1970) formulated the following rule to compare the MDTs: "Choose that technique which minimizes $E\left\|\Sigma - \hat{\Sigma}_i\right\|^2$,where i denotes the various techniques used in the estimation of Σ." As a measure of the overall comparative efficiency of any pair of procedures, the ratio of the Euclidean norms was used. For example, the efficiency index for two MDTs (A and B) in estimating Σ is given by

$$eff(\Sigma_A / \Sigma_B) = \frac{\left\|\Sigma - \Sigma_A\right\|^2}{\left\|\Sigma - \Sigma_B\right\|^2}$$

However, Timm (1970) found that any missing data treatment given preference in the estimation of Σ by employing the Euclidean criterion or the average relative efficiency index may not necessarily function best in minimizing the largest eigenvalue mean square error.

Gleason and Staelin (1975) provided an approach similar to Timm's method for determining which technique produces from incomplete data a correlation matrix most similar to the correct matrix (i.e., the one calculated from complete data). This dissimilarity measure is defined as follows. Let $\|A\|$ denote the Euclidean norm of the matrix A, let R denote the correlation matrix for complete data, p the number of variables, and R_α the correlation matrix estimated from incomplete data using method α. Then

$$D_\alpha = \sqrt{\left(\frac{\|R_\alpha - R\|^2}{p(p-1)}\right)}$$

represents root-mean-square deviation of predicted versus actual values for the off-diagonal correlations. The formula can be used to determine which of the MDTs yields the better estimates. To compare different MDTs for predicting missing data, Gleason and Staelin (1975) constructed the following measure of the quality of various estimates. Let x_{ij} be the value in a complete data matrix that corresponds to a missing entry in an incomplete matrix; let $\sigma_j^{\;2}$ be the variance, in the complete matrix, of the variable in which the entry lies; let $x_{ij}^{\;(\alpha)}$ be the estimate of x_{ij} derived by method α; and let K denote the set of entries missing from the matrix. Then a measure of the dissimilarity between true and estimated values for method α is given by

$$Q_\alpha = \sqrt{\left\{\frac{\sum\left(x_{ij}^{\;(\alpha)} - x_{ij}\right)}{\sigma_i^{\;2} mp\pi}\right\}}$$

where i, j are elements of K, π is the proportion of missing entries, m is the sample size, and p is the number of variables. This number is the root-mean-square standardized residual.

Beale and Little (1975) compared six methods for handling missing data:

(1) Ordinary least squares (OLS) using complete observations only

(2) Buck's (1960) method

(3) Iterated Buck, or corrected maximum likelihood

(4) OLS on observations with y present, after fitting missing values of the independent variables only,

(5) Method 4, but with missing observations given fractional weights and

(6) Method 5, but using a covariance matrix for all variables, found by method 3, to find the fitted values and estimate the weights.

Data were generated from a multivariate normal population with one variable identified as the dependent variable, and between two and four independent variables. Samples of size 50, 100 or 200 were used. Deletion of 5, 10, 20 or 40 percent of the observed values was conducted randomly. Their criterion for judging the effectiveness of each MDT was the residual sum of squares of deviations of the observed and fitted values of the dependent variables when the deleted values were restored. This may be written as

$$S = \sum \left(y_i - b_0 - \sum b_j x_{ij}\right)^2,$$

where b_0 and b_j are the regression coefficients estimated from missing data by one of the six MDTs, and x_{ij} and y_i are the true values of all variables without deletions. A small value of S represents a successful method. Findings showed that Buck and iterated Buck outperformed OLS using complete observations only.

Finkbeiner (1979) discussed a maximum likelihood method of estimating the parameters of the multiple factor model when data were missing at random from the sample. He used a Monte Carlo study to compare the ML method with five other heuristic methods of dealing with missing data. The five methods were complete data, mean replacement, pairwise deletion, regression replacement, and principal components. Two proportions of missing data were chosen: one to have only a little data missing and the other a lot, both for samples of size 64. For each sample, random, independent normal numbers were generated by the Box and Muller (1958) method with 50 replications for each of the two patterns. All the six MDTs were used to analyze each of the 100 samples. Sampling distributions for each method on each missingness pattern were developed.

Three pieces of information were used in examining estimation effectiveness: sampling means and dispersions of parameter estimates, mean squared error of parameter estimates, and an index of common factor recovery. For the first, the least biased, minimum dispersion estimate was considered optimal. For the second, the best method produced the smallest mean squared error most consistently, regardless of bias. For the last, the average of the squared multiple correlation coefficients of congruence was used. This index is referred to as a "congruence coefficient." It varies in value from zero to one, with zero indicating no recovery of the population common factor space and 1 indicating perfect recovery. The three best methods were maximum likelihood, mean replacement and complete pair only.

Tirri and Silander (1998) proposed an imputation procedure for the treatment of missing data called stochastic complexity. They recommended the method for multivariate categorical data. The method is related to multiple imputation, which uses independent realizations of the posterior predictive distribution of the missing data under some complete-data model and prior. Intuitively, the approach involves modeling the set of data records as a matrix of incomplete discrete data. The missing data are estimated by assuming a functional form for the probability distribution of the cases (i.e., a model class). Based on the given model class assumption, an information-theoretic criterion can be derived to select between the different complete data matrices for the more likely one. Stochastic complexity represents the shortest code length needed for coding the complete data matrix relative to the model class chosen. A weakness of the method is that the exact criteria are very hard to compute for many interesting models, but it can be approximated by the Bayesian marginal likelihood computed by integrating over all the possible models in the chosen model class.

In their validation study, Tirri and Silander (1998) produced synthetic missing data problems from real data sets by randomly deleting known portions of the data. Sample size was controlled for. They performed a classification analysis to study the practical implications of different procedures for handling missing data. For each data set, they created three different sub-samples of sizes 10%, 25% and 50% of the original data set. Each time 50% of the data was reserved for the subsequent out-of-sample classification analysis. In each sample, 5%, 10%, 20%, 35% and 50% of the elements were deleted, thus creating artificial missing data problems satisfying MCAR assumption. Mean substitution and stochastic complexity were used to create complete data matrices by imputing the missing values. Results showed that the stochastic complexity based approach performed well in recovering the original missing data.

Studies with Non-randomly Missing Data

Most missing-data treatment techniques require the assumption that data are missing at random. However, data sets in which values are randomly missing are rare in the fields of psychology and education. Research on the extent to which MDTs developed for randomly missing data may be useful for non-randomly missing data is therefore important. This area of research has received little attention.

Azen et al. (1989) carried out a simulation study to compare the performance of three methods (listwise deletion, pairwise deletion, and expectation maximization) in estimating regression coefficients and missing values, for situations with varying proportions of missing data (5% and 25%) and varying magnitudes of correlations (R^2 = .5 and .9). Two evaluation criteria were used. The first criterion, C_1, was a measure of how well the MDT estimated regression coefficients:

$$C_1 = \frac{\Sigma(b_i - \beta_i)}{[2\sigma^2(2 + 2\rho - \rho^2]/[n(1 + 2\rho)(1 - \rho)]}$$

The denominator of C_1 is a normalizing factor equal to the expected value of the numerator conditioned on no missing data. The second criterion, C_2, was a measure of how well the MDT imputed the data in the sense that it measured how well each MDT estimated the mean vector and covariance matrix:

$$C_2 = (3nf)^{-1}\Sigma\Sigma d^2_{ij}/\min d^2_{ij}$$

where the summation is over all missing values, and

$$d^2_{ij} = \min d^2_{ij} + (X^*_{ij} - X^\#_{ij})^2$$

$$\min d^2_{ij} = V(X_j) - C_{XW}C^{-1}_{WW}C_{WX}$$

$$X^*_{ij} = E(X_{ij} \mid W_i, \Lambda)$$

$$X^\#_{ij} = E(X_{ij} \mid W_i, \Lambda^\#)$$

and $\Lambda^\#$ is the estimated covariance matrix, min d^2_{ij} is that part of d^2_{ij} which results if Λ was estimated perfectly, and is a constant over all MDTs.

For estimating regression coefficients at 5% missing, analyses revealed that all three methods performed well for $R^2 = 0.5$ condition. However, pairwise deletion was generally inferior for $R^2 = 0.9$ condition. EM and listwise deletion performed equally well regardless of the magnitude of correlations, when the percentage of missing data was only 5 per cent. When the percentage missing increased to 25, EM was generally the best. In addition, PW was competitive only for situations with weak correlations and/or little missing data. With a pattern of censored missing data, the PW method performed better than the EM method.

For estimating regression coefficients at 25% missing, analyses of variance showed that the EM method was generally the preferred method. Pairwise deletion was competitive for weak correlations. In addition, an analysis of variance showed that EM method performed statistically better for the random and related patterns of missing data than it performed with censored data.

A limitation of the study by Azen et al. (1989) is that they made use of only 50 replications. Efron (1980) suggested that although as few as 100 replications could provide a stable estimate of various statistics, a tenfold increase in the number of replications would provide desirable accurate results. Also, Azen et al. (1989) did not consider two important imputation methods; mean substitution and regression imputation. There is need for studying the effectiveness of these methods in comparison to pairwise deletion and EM algorithm using a large number of replications (say, at least 1,000) in order to offer more reliable guidance to field researchers.

Kromrey and Hines (1994) investigated the effects of non-randomly missing data in two predictor regression analyses. They examined differences in the effectiveness of mean substitution, simple regression, multiple regression, listwise deletion, and pairwise deletion on estimates of R^2 and regression coefficients. They used bootstrap samples drawn from three large sets of actual field data, representing Likert-type rating data, achievement test data, and psychological trait data. The study design was a $3 \times 3 \times 6 \times 5$, with three between-subject factors (parent population, sample size, proportion of data missing) and one within-subjects factor (MDT). For comparison, they computed the effect sizes obtained from the MDT conditions relative to the complete sample condition. Kromrey and Hines found that the differences among the MDTs increased as the proportion of missing data increased. They also found that multiple regression imputation consistently yielded overestimates of R^2. Conversely, the use of mean imputation consistently yielded underestimates of R^2. The simple regression imputation procedure overestimated R^2 only in the psychological trait data, where the overestimation was consistent. Listwise deletion underestimated R^2, except in the psychological trait data, where an inconsistency in the direction of the effect was realized. Although they found that no MDT yielded consistently best estimates of R^2 across the data sets, sample sizes, and proportion of missing data, pairwise deletion and the simple regression imputation appeared to perform better in most situations than the use of mean substitution and multiple regression imputation.

In the estimation of regression coefficients, Kromrey and Hines (1994) found that listwise deletion provided the best overall performance. The imputation techniques (multiple regression, simple regression, and mean substitution) did not perform well in estimating population R^2 as well as the regression coefficients when applied to non-randomly missing data. Whereas Kromrey and Hines provided support for the use of certain MDTs in situations where data are missing non-randomly, they stressed the importance of careful selection of missing data treatment because of the substantial differences in the effectiveness of the MDTs. The differences in the effectiveness of the treatments across different types of data highlighted the need for further research to identify the types of data matrices that may be more amenable to analysis by those methods.

Whereas Kromrey and Hines (1994) offered important guidelines on the effectiveness of MDTs, there is need to extend their findings by including EM in the comparisons. A key limitation in the study by Kromrey and Hines (1994) is the extent to which the results may be generalized. The study was limited to missing data occurring on only one predictor in a 2-predictor model. Such a model is rare in psycho-educational research. There is need for further research to examine the use of MDTs when missing data occur systematically on more than one predictor.

In a more recent study, Brockmeier, Kromrey, and Hines (1996) investigated within the context of a two-predictor multiple regression analysis with non-randomly missing data, the effectiveness of eight missing data treatments on multiple R^2 and each standardized regression coefficient. They recommended the use of stochastic regression in estimating population R^2 except when the proportion of missing data is high (60%). They also recommended the use of pairwise deletion and listwise deletion in estimating the sample estimate when percent missing is low (10%). When examining the standardized regression coefficients, none of the MDTs were found effective when missing data was 60%. They recommended the use of deletion and stochastic regression procedures to generate unbiased parameter estimates when percent missing is low. A limitation of the study is that they used a two-predictor model only.

Fraser and Halperin (1998) compared listwise deletion, pairwise deletion, and maximum likelihood method in the estimation of the mean vector and covariance matrix of a random sample from a multivariate normal distribution. A monotonic missing pattern was used, formed by three time periods in a longitudinal study. The parameters of interest were the mean, variance, and covariance. Specifically, they wanted to discover whether maximum likelihood would outperform the quick methods that were simpler and easier to obtain. There were four between-group factors in their design. These were number of variables in each time period (2, 3, 2) and (3, 2, 1), correlation among the variables (R^2 =.32 and .71), number of observations for variables in the first block (50 and 70), and attrition rate (20% and 40%). The within-factor was the missing data treatment (listwise deletion, pairwise deletion, and maximum likelihood).

A split-plot factorial analysis was used for which only the interactions, which included the missing data methods, were discussed. Findings showed that attrition rate discriminated among the missing data methods for the mean and variance. For the mean vector, the interaction between estimation methods and the attrition rate was significant with an effect size in the medium to high range according to Cohen's (1988) guidelines. The other design factors did not differentiate among the methods. For the mean, pairwise deletion was recommended for 20% attrition, but maximum likelihood was recommended for 40% attrition. For the variance, pairwise deletion was recommended at both 20% and 40% attrition rates. The only significant interaction was that of estimation methods and attrition rate, with an effect size in the middle to higher range. For the covariance, none of the design factors were able to discriminate among the missing data methods.

Studies in Structural Equations Modeling

A number of researchers have conducted studies involving MDTs in the context of structural equation modeling. A structural model is specified in part by

$$\eta = \mathbf{B}\eta + \Gamma\xi + \zeta$$

where η is a vector of endogenous variables, ξ is a vector of exogenous variables, ζ is a vector of unobserved, exogenous disturbances, and **B** and Γ are matrices of coefficients. ζ and ξ are assumed to be uncorrelated. For convenience, all variables are assumed to have means of zero, thus eliminating the need for intercept terms in the model. To achieve identification, one must usually impose restrictions on **B**, Γ, and var(ζ). Important special cases of the above equations include multivariate regression (**B** = **0**), multiple regression (**B** = **0** and η is a scalar) and recursive systems (**B** is subdiagonal and var(ζ) is diagonal).

Allison (1987) proposed a maximum likelihood method which he claimed to be both consistent and efficient in estimating parameters for the general linear structural relations model when data are missing. The method capitalizes on the ability of statistical software for structural equations modeling to estimate simultaneously the same model for two or more samples. For missing data problems, the sample is divided into subsamples, each having a different set of variables present. The model is then estimated simultaneously for all subsamples, constraining corresponding parameters to be equal across subsamples.

To facilitate estimation of the above model with missing data, Allison (1987) modified the model to the equivalent form

$$\eta^* = \mathbf{B}^*\eta^* + \zeta^*,$$

where $\eta^*=(\eta, \xi)'$, $\zeta^*=(\zeta, \xi)'$, and $\mathbf{B}^*=\begin{pmatrix} B & \Gamma \\ 0 & 0 \end{pmatrix}$. This class of models is greatly expanded by allowing for the possibility that η may not be directly measured. This is accomplished by adding equations that specify the effects of the latent variables on a set of observed indicators:

$$\mathbf{y} = \Lambda_y\eta + \varepsilon.$$

Here, **y** is a vector of observed variables, and ε is a vector representing random measurement error. Λ_y is a matrix of coefficients that must usually be restricted in some way to achieve identification.

To allow for missing variables, Allison (1987) suggested that any variable that is missing from any of the subsamples must be an indicator of a latent variable. Any missing covariances are set to zero, missing means to zero, and missing variance to one. The coefficients in the Λ_y matrix are then used to "switch" the variable "on" or "off" depending on whether it is present or absent in a particular subsample. Specifically, in subsamples with data missing for a particular variable y_i, all elements in row i of Λ_y must be fixed at zero. One must also fix var(ε_i) = 1 and cov(ε_i, ε_j) = 0 for $i \neq j$. These constraints ensure that the pseudo-values of zero and one in the sample covariance matrix for that subsample will be fitted exactly.

In subsamples with data present for that variable y_i, the treatment depends on whether or not the model allows for random error in the measurement of that variable. If the model does not allow for random error, one of the λ_{ij} coefficients should be fixed at zero and var(ε_i) should be fixed at zero. If random error is allowed, no special constraints are needed; the λ coefficient and the error variance should be left as free parameters to be estimated. Finally, all other parameters are constrained to be equal across subsamples. With the LISREL computer program, Allison (1987) demonstrated the superiority of maximum likelihood estimation when data are missing at random but not observed at random.

Muthen, Kaplan, and Hollis (1987) gave a general latent variable model that includes the specification of a missing data mechanism. In their model, MCAR was not a prerequisite for unbiased estimation in large samples, as when using the traditional listwise or pairwise deletion approaches. Using artificial data, they found that likelihood estimation was superior to traditional estimation methods in situations involving data that were not missing completely at random.

Arminger and Sobel (1990) constructed a nonlinear method, called pseudo-maximum likelihood estimation (PML), which can be used with missing data. They illustrated how maximum likelihood estimation of parameters using missing data (see Little and Rubin, 1987) can be extended directly to the PML estimation of the same parameters.

Brown (1994) assessed the relative efficacy of five indirect methods for dealing with missing data in structural equation models using simulated data. The five methods studied were listwise deletion, pairwise deletion, mean substitution, hot-deck imputation and similar response pattern imputation. However, Brown's study was restricted to a cross-sectional design, and he did not consider other estimation methods, like the EM algorithm in his comparison although he recommended this for future research.

Reinecke and Schimdt (1996) compared listwise deletion, pairwise deletion, and similar response pattern imputation with a non-iterative maximum likelihood procedure developed for nested missing values in panel data. The similar pattern method is used to impute the item non-responses in each wave while the non-iterative maximum likelihood procedure is used to consider unit non-response because of panel attrition. They found that listwise deletion method led to an unacceptable waste of information across and within panel waves. However, pairwise deletion method led to a reasonable amount of explained variances with the best model fit. The non-iterative maximum likelihood procedure produced reasonable estimates of the covariance matrix taking into account unit non-response, and imputing the item non-responses in each wave did not contribute to the explained variances and the overall model fit.

Employing simulation methodology, Marsh (1998) addressed the use of sample covariance matrices constructed with pairwise deletion for data missing completely at random. In the study, 3 levels of sample size (N = 200, 500, 1,000) and 5 levels of percent missing (0, 1, 10, 25, and 50) were used. A population covariance matrix was generated from a simple population model in which three latent factors were defined by three measured variables such that each of the nine measured variables had a nonzero loading on one and only one latent factor. Population values were .7 for all factor loadings, .51 for all measured variable uniqueness, and .4 for all correlations among the three latent factors. Five sets of 100,000 cases were simulated from this population covariance matrix by GENRAW, a program for data generation within LISREL 8 by Jöreskog and Sörbom (1993). The cases differed only in respect to the five levels of missing values to allow for comparisons. For each set, the probability that a particular data point was missing was specified to be 0%, 1%, 10%, 25%, or 50%. A random variable was paired with each data point using the random variable routine in SPSS, and this random variable was used to select values that were missing.

The investigation emphasized the effects of missing values both on the χ^2 test statistic, subjective indices of fit, and parameter estimates and on how these effects interact with sample size. Three indices of fit, namely, relative noncentrality index (RNI), nonnormed fit index (NNFI) and root mean square error of approximation (RMSEA) were considered. Evaluation of parameter estimates was based on their means and standard deviations. To assess the relative size of the various effects and to provide a nominal test of statistical significance, two-way analyses of variance were conducted in which the effects of the three sample sizes and five levels of missing data were considered. One potentially serious limitation of the pairwise deletion approach is the occurrence of a non-positive definite covariance matrix. Of the 4,000 covariance matrices generated in the simulation study, only 27 non-positive definite covariance matrices were observed. This occurred only in the cell with N = 200 and percent missing of 50.

Surprisingly, contrary to the findings of Brown (1994), mean estimates of factor loadings, of factor correlations, and of uniqueness were similar across all the cells of the design. In sum, estimates based on pairwise deletion were unbiased even when percent missing was large and the sample size was small.

Marsh (1998) also found that consistent with expectations, the sizes of standard deviations varied inversely with sample size and directly with percent missing. Sample size, percent missing, and their interaction substantially influenced the magnitudes of observed standard deviations. With respect to the fit indices (RNI, NNFI, RMSEA), the pattern of the effects was similar to that observed for the χ^2 test statistic. Results also showed that the χ^2 test statistics were largely successful in eliminating the effect of missing values, although there were still small effects associated with sample size and the interaction of sample size and percent missing.

Marsh (1998) recognized the potential limitation of generalizing the results from his study. Whereas a wide range of sample sizes and percent missing were considered, the results were based on only one population generating model, and the variance-covariance matrix constructed under pairwise deletion contained no misspecification error (i.e., only "true' models that should be able to explain the data were considered). Hence there is need to explore a wider range of conditions in further simulation research. One other limitation of the study was the use of MCAR, an assumption that is unreasonable in most applications.

Summary

Missing data can be of two types, namely unit non-response and item non-response. The distinction between unit non-response and item non-response was discussed. For example, unit non-response may occur when a respondent totally refuses to respond in a questionnaire. Item non-response may be a result of an instrument or question being broken off after being partly completed.

Several methods of treating missing data were presented. These methods may generally be classified into two categories: deletion methods and imputation methods. Deletion methods include listwise deletion and pairwise deletion. Imputation methods include mean substitution, regression imputation and expectation maximization procedure.

Factors that may influence the performance of MDTs were discussed. These included missing pattern, percent missing, non-normality, number of predictors and magnitude of correlation. In particular, it was noted that the effects of missing pattern and non-normality on parameter estimates after different treatments of missing data had not received enough attention by researchers.

MDTs were essentially developed for use under the assumption of MCAR. Therefore, previous researchers tended to evaluate MDTs using data sets under the MCAR condition, with little regard of the fact that the MCAR condition is difficult to obtain in real research settings. Two common systematic missing patterns were identified: monotonic and non-monotonic patterns.

A number of limitations of previous studies were noted. Of importance were the small number of replications, and failure by researchers to use regression models with more than two predictors.

It was also noted that an area that needed further research is the effectiveness of regression imputation and EM method when compared with MS and PW methods. In doing so, it is important to consider the role of other factors like pattern of missing data, size of R^2, number of predictors, sample size and percent missing, the levels of which must be based on a sound rationale.

Various criteria for the evaluation of MDTs were presented, indicating that different researchers used different criteria, the choice depending mainly on the parameter estimate of interest. Such criteria included effect size, Euclidean norm, RMSE, MSE, sampling means and standard deviations, as well as the average of the squared multiple correlation coefficients of congruence.

Notable findings in previous research include that of Azen et al. (1989) who found that in the estimation of regression coefficients, with 5 percent missing, both EM and PW performed well when multiple $R^2 = 0.5$. However, PW was inferior to EM when multiple $R^2 = 0.9$. At 25 percent missing, EM performed better than PW, although PW was competitive when $R^2 = 0.5$. On the other hand, with a pattern of censored data, the EM method did not perform well in the estimation of parameters as it was biased. The bias was possibly due to the non-linearity of the regression relationship among variables. Kromrey and Hines (1994) found that RS and MS methods did not perform well in estimating population R^2 and regression coefficients when applied to non-randomly missing data. Fraser and Halperin (1998), using a monotonic missing pattern, found that the PW method was better than maximum likelihood method for variance estimation at both 20% and 40% missing.

CHAPTER 3

METHOD

To answer research questions presented earlier, MDTs were evaluated under various conditions using simulated data. In the present chapter, the advantages of using simulation methodology, issues in data generation, factors influencing the effectiveness of MDTs, the study design, the procedure, and analysis and evaluation criteria are discussed.

Advantages of Simulation Methodology

Simulation methodology may be used to study the properties of statistical techniques under diverse settings. The main goal of conducting a simulation study is to collect evidence for evaluating specific procedures. As noted by Efron (1980), computer intensive studies can both advance statistical theory, and provide improved methods of solution where no good theory exists. Examples of areas where such methodology can be applied include examination of robustness properties, assessment of small sample versus asymptotic agreement, or comparison of a statistical method with its competitors (Johnson, 1987). Rubin (1991), like Efron (1980), also noted that simulation methodology is an inherently statistical idea in that it capitalizes on the proposition that the fine detail available in an analytic solution is often not necessary. In practice, inferences about a parameter θ can be as beneficially drawn from several hundred equally likely values of θ, as from formulas giving the mean and variance of an infinite number of such draws. In fact, Rubin (1984) made an argument that several hundred random draws often may yield better practical inferences than traditional analytic solutions based on asymptotic approximations. First, asymptotic approximations can provide a less accurate picture of inferential uncertainty than simulated draws when the likelihood function is not nearly normal. Second, implemented simulation techniques allows the investigator to explore a variety of underlying models and thereby avoid fixation with possibly inappropriate models that have neat analytic or asymptotic solutions.

Issues in Data Generation

A simulation study requires the consideration of how to generate the observations to be used in the study. In this study, multivariate random numbers from populations with specified skew, kurtosis, and intercorrelations were used. Although several random number generating methods are available, many of the generators currently in use are seriously flawed (L'Ecuyer, 1988, 1990, 1994). Therefore, it was important to select a random number generator with extreme caution.

L'Ecuyer (1994) provided important requirements for a good random number generator. He argued that a good general-purpose generator must have specified statistical properties, namely, adequate period length, ease of implementation, efficiency, portability, and reproducibility. He recommended the use of higher-order congruential generators instead of multiplicative linear congruential generators. EQS (Bentler, 1989), a statistical program that uses a higher-order congruential generator, was used to generate data.

Factors Influencing MDT Performance

A number of factors may influence the performance of MDTs when used to treat missing data in research situations. In this study, the factors that were considered include Non-normality (NM), Sample Size (SZ), Proportion of Missing Data (PM), and Pattern of Missing Data (PT). However, there is a possibility that in a multiple regression context, the Number of Predictors (p), and the magnitude of multiple R^2 may also have an impact on the effectiveness of MDTs. For this reason, four studies were conducted. In the first study, there were four predictors under low R^2 condition; in the second study, there were four predictors under a high R^2 condition; in the third study, there were nine predictors under a low R^2 condition; and the fourth study involved the use of nine predictors under high R^2 condition.

To be closer to what researchers have encountered, the values used at different levels of the manipulated factors were largely based on the findings from a comprehensive review of papers in four journals, namely, American Educational Research Journal, Journal of Applied Psychology, Journal of Educational Psychology, and Journal of Educational Research. In the review, I found that a total of 177 studies in the 1990 to 1995 issues of these journals [excluding articles in the 1993 Journal of Applied Psychology, 78(4) to 78(6)] used multiple linear regression analysis. Of the 177 studies, 9.04% (16) were in the American Educational Research Journal, 46.33% (82) were in the Journal of Applied Psychology, 29.94% (53) were in the Journal of Educational Psychology, and 14.69% (26) were in the Journal of Educational Research. These studies were reviewed to help identify which values of the design variables are commonly found in psycho-educational research so that representative levels could be selected for use in this study. Table 1 is a breakdown of the number of studies in each journal that used multiple regression analysis.

Table 1: Number of studies using multiple regression in four journals

	YEAR						
	1990	1991	1992	1993	1994	1995	TOTAL
AERJ	2	5	3	1	1	4	16
JAP	8	22	13	8	10	21	82
JEP	18	7	8	12	2	6	53
JER	6	5	4	6	3	2	26
TOTAL	34	39	28	27	16	33	177

Note:
AERJ = American Educational Research Journal
JAP = Journal of Applied Psychology
JEP = Journal of Educational Psychology
JER = Journal of Educational Research

The following is a description of my findings based on the review of the four journals regarding proportion of missing data, sample size, non-normality, number of predictors, and magnitude of R^2 that can be considered typical in psycho-educational research.

Proportion of Missing Data

A review of the four journals mentioned above showed that most studies reported response rates at the unit level, but not at the item level. Of the 177 studies that used OLS regression analysis, 29.38% (52) reported response rates. I derived non-response rates based on these response rates. The mean non-response rate was 28.8 percent, with a standard deviation of 21.03 ($N = 52$). The distribution was multimodal, with the smallest mode being 22.86. The lower quartile was 12.83, the median 22.86, and the upper quartile 41.38. Considering that these values included both unit non-response and item non-response, it is most likely that the second quartile is the upper bound and the first quartile is the lower bound for item non-response. Based on these findings, the levels for percent missing in the present study were 10%, 15%, and 20%. These values were similar to what other researchers had used. For example, Raymond and Roberts (1987) had three levels: 2%, 6%, and 10%, and Azen et al. had two levels: 5% and 25%. The number of data points deleted to create missing data patterns in the present study was computed as percent missing multiplied by the total number of data points in the predictors.

Sample Size

Of the 177 studies using multiple regression, 19 used data sets with more than 1,000 cases (e.g., High School and Beyond data with $N = 14,825$). Such data sets were found to severely skew the distribution of sample sizes in a positive direction. I eliminated sample sizes greater than 1,000 because most surveys would use values below this. However, this did not totally eliminate skew as the resulting distribution was still positively skewed, but not as severely skewed as when sample sizes greater than 1,000 were included. It was also difficult to ascertain the sample size used in a number of studies. In all, 148 sample sizes below 1,000 were reported. The mean sample size was 218 with a standard deviation of 192 ($N = 148$). The minimum sample size was 20 and the maximum, after eliminating those above 1,000, was 983, giving a range of 963. The first quartile was 94, the median 153, and the third quartile 265. In order to have typical sample sizes in this study, samples of size 94, 153, and 265 were used.

Non-normality

Whereas the normality assumption is a prerequisite in multiple regression analysis, Micceri (1989) and Sawilosky and Blair (1992) observed that many data sets in real situations are skewed and/or kurtotic, thus violating the normality assumption. When investigating the effects of non-normality, it is extremely important to consider non-normality in the degree found in the literature. Before discussing the levels of non-normality used in this study, it is perhaps important to present an observation on how often researchers report violation of assumptions in their studies. Of the 177 articles that were reviewed, reports on regression diagnostics were contained in only five studies! This shows how researchers have often failed to report important issues before doing multiple regression analysis. Two of the five studies contained reports on violation of normality assumption, one had reports on multicollinearity diagnostics, one had reports on violation of homogeneity of variance assumption, and one had reports on outlier diagnostics. The study in which outlier diagnostics were reported also reported deletion of missing data. The review did not help in choosing appropriate levels of non-normality. So, I relied on previous reviews to choose levels of non-normality.

After observing various empirical distributions, Pearson and Please (1975) found that typical non-normality had skew less than 0.8 and kurtosis between –0.6 and 0.6. However, Sawilosky and Blair (1992) who based their report on the findings of Micceri (1989) later provided eight real-world distributions different from what was reported by Pearson and Please (1975). Three types of psychometric distributions were identified as typical; discrete mass at zero with gap, extreme asymmetry, and extreme bimodality, with (skew, kurtosis) = (1.65, 0.98), (1.64, 1.52), and (–.08, –1.70), respectively; the values for kurtosis having been such that normal kurtosis is 0.0. Whereas these were considered typical distributions, Micceri (1989) noted that typical kurtosis estimates ranged from –1.70 to 37.37. In their simulation study, Graham, Hofer, and MacKinnon (1996) used distributions with skew –.68 to 3.30 and kurtosis –.04 to 13.11.

Based on previous findings, it is highly probable that typical distributions in psychometric scores would have skew hardly exceeding 3.0. However, typical distributions of psychometric scores seem to have kurtoses that vary widely, from –2.0 to approximately 40.0. Therefore these range of values were used as a guide in this study. The distributions used in this study were a normal distribution (skew = 0 and kurtosis = 0), a low level of non-normality (skew = 1 and kurtosis = 3), a medium level of non-normality (skew = 1.8 and kurtosis = 6), and a high level of non-normality (skew = 3, kurtosis = 25).

Number of Predictors

In the reviews of four journals described earlier, of the 177 studies that used multiple regression analysis, 174 reported number of predictors. The mean number of predictors was 7.5 with a standard deviation of 5.94 (N = 174). The minimum value was 2 and the maximum was 40, giving a range of 38. The first quartile was 4, the median 5 and the third quartile 9. The distribution was unimodal, with a mode of 4. Based on these outcomes, this study included 4 and 9 predictor regression models.

Size of Multiple R^2

Researchers usually want to explain maximum variance in the criterion variable. This is more particularly so when using multiple regression analysis. With this in mind, and using the 177 studies that employed multiple regression analysis, I recorded the maximum value of R^2 in each study. The value of R^2 was reported in 148 studies. The mean R^2 was .43 with a standard deviation of .22 (N = 148). The distribution of R^2 was bimodal. In other words, the studies that were reviewed could be categorized into two groups, those that had low R^2 from 0.01 to 0.30, and those that had moderate to high R^2 from 0.40 to 0.90. Seaman, Algina and Olejnik (1985) noted that correlations encountered in behavioral science research often occur in the interval from .2 to .7, values that are consistent with those found in my review. Based on these findings, the two levels of R^2 used in this study were LOW (with R^2 approximately 0.2) and HIGH (with R^2 approximately 0.6).

Pattern of Missing Data

Two patterns of missing data considered in this study were monotonic pattern and non-monotonic pattern. These patterns were selected because they represent two major categories of systematically missing data discussed in Little and Rubin (1987) and Little (1992). The monotonic pattern of missing data arises when the variables can be arranged so that for $j = 1,\ldots,K-1$, X_j is observed whenever X_{j+1} is observed. Little and Rubin (1987) noted that attrition from a panel survey leads to data of this form. To illustrate the mechanism of non-monotonic pattern of missing data, assume that Y is a measure of pupils' ability in Figure 2. If X_p represents father's income, then pupils of high ability may be too sensitive to respond, and pupils of low ability may not know their father's income. On the other hand, pupils with moderate ability may be in a position to respond to X_p. Figure 1 shows the monotonic pattern, and Figure 2 shows the non-monotonic pattern used in the present study. The criterion variable (Y) and p predictors, $p = 4$ or 9, were all continuous. In both figures, rectangles stand for available values. To create dependency of missing values in the set of predictors on the magnitude of Y-values, the data were sorted using Y before deletion, with the lowest score being at the top and the highest score being at the bottom. Deleting data using templates with configurations given in Table 2 and 3 respectively created the patterns in Figure 1 and 2. The same numbers of data points were deleted for both monotonic and non-monotonic patterns for corresponding X's. In other words, for each variable across pattern, the same number of data points was deleted.

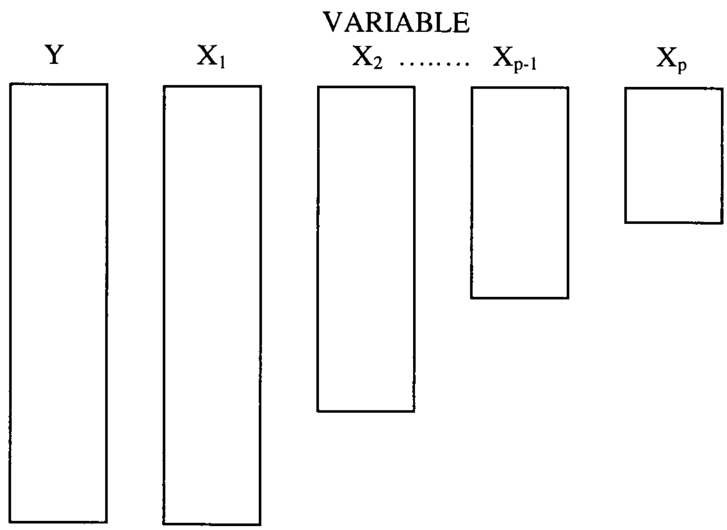


Figure 1: Monotonic pattern of missing data

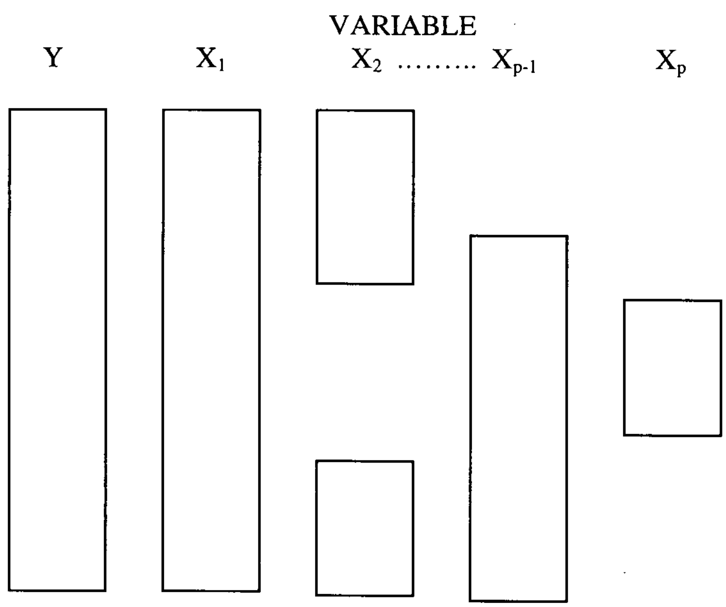


Figure 2: Non-monotonic pattern of missing data

Table 2
Number of data points deleted to create missing patterns in the 4-predictor model

Sample Size	Percent Missing	Number of Data Points Deleted				Total
		X_1	X_2	X_3	X_4	
94	10	0	8	15	15	38
	15	0	12	22	22	56
	20	0	15	30	30	75
153	10	0	13	24	24	61
	15	0	18	37	37	92
	20	0	24	49	49	122
265	10	0	22	42	42	106
	15	0	31	64	64	159
	20	0	42	85	85	212

Table 3
Number of data points deleted to create missing patterns in the 9-predictor model

Sample Size	Percent Missing	Number of Data Points Deleted									Total
		X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	
94	10	0	2	6	8	10	12	14	16	18	86
	15	0	8	10	12	14	16	20	22	26	128
	20	0	10	12	16	18	24	26	30	34	170
153	10	0	8	10	12	14	18	20	26	30	138
	15	0	12	14	18	22	26	32	38	44	206
	20	0	18	20	24	28	34	44	50	58	276
265	10	0	14	18	20	24	30	36	44	52	238
	15	0	22	26	32	38	46	54	64	76	358
	20	0	30	34	42	50	60	72	86	102	476

Study Design

Four separate studies were conducted in this investigation because the design included 2 levels of population R^2 (approximately 0.2 and 0.6) and 2 levels of number of predictors (4 and 9). Study 1 was with four predictors under low R^2 condition, Study 2 was with four predictors under high R^2 condition, Study 3 was with nine predictors under low R^2 condition, and Study 4 was with nine predictors under high R^2 condition. As shown in Table 4, there were four levels of non-normality, three levels of sample size, three levels of percent missing, and two levels of missing pattern. Each study was a 4 (non-normality) × 3 (percent missing) × 2 (missing pattern) fully crossed factorial design, across 3 levels of sample sizes, giving rise to 72 conditions. The within-factor was MDT including mean substitution, pairwise deletion, regression imputation, and EM method of treating missing data.

Table 4
Design Factors and Levels for the Study

Design Factors					
Distribution	R^2	Predictors	Sample Size	Pattern	Percent Missing
(Sk., Kurt.)					
(0, 0)	Low (.2±.05)	4	94	Monotonic	10
(1, 3)	High (.6±.05)	9	153	Non-monotonic	15
(1.8, 6)			265		20
(3, 25)					

Procedure

Population correlation matrices were used to generate multivariate normal random variables using data generation algorithms built in EQS (Bentler, 1989), a widely used structural equations modeling program. Another reason for choosing EQS is that it has a built-in procedure developed by Fleishman (1978), whose equations were extended to cater for multivariate data by Vale and Maurelli (1983). The procedure makes it possible to generate data with specified skew and kurtosis, while keeping the correlations invariant. This is necessary if comparisons are to be made across different data sets. The correlations used in the study represent typical values found in educational and psychological research as identified in the review of studies.

The two levels of number of predictors 4 and 9, gave rise to 5 and 10 variables, respectively, if Y is included. The use of 5 variables implies a 5 by 5 zero order correlation matrix and the use of 10 variables implies a 10 by 10 zero order correlation matrix. Considering that these correlation matrices are symmetric, one needs 10 bivariate correlations to construct a zero order correlation matrix for 5 variables, and 45 bivariate correlations to construct a zero order correlation matrix for 10 variables. In order to construct correlation matrices for each level of R^2, a set of sixty correlations from .01 to .60 in increasing steps of .01 was used, from which 10 and 45 correlations, respectively, were randomly selected using a table of random digits. Only positive values were considered because the direction of relationship would not affect generalizability. The interval .01 to .60 was selected because in the review of 177 articles in four journals discussed earlier, the distribution of minimum correlations across all studies had a median of .01 and the distribution of maximum correlations had a median of .60. The randomly selected correlations were then rearranged from top to bottom of a lower triangular matrix, with the first column representing correlations between Y and X's and the remaining columns representing inter-correlations among the X's. Using matrix data input in SPSS computer software (SPSS Base System Reference Guide, Release 6.0, 1993), the variance in Y explained by the X's (i.e., multiple R^2) was obtained. If the resulting value of R^2 was not as desired (i.e., $0.20 \pm .05$ for LOW R^2 or $0.60 \pm .05$ for HIGH R^2), then another correlation was randomly chosen from the set of 60 to replace a single entry in the correlation matrix. The replaced element was dictated by whether the value of obtained R^2 was larger or smaller than desired. This procedure was repeated until the desired R^2 values were obtained within $R^2 \pm .05$. Once the matrices were generated, they were examined for positive definiteness. The steps used in constructing correlation matrices are summarized as follows:

1. Start with a set of sixty correlations from .01 to .60 in increasing steps of .01.
2. Randomly select correlations in (1) to get the elements in the lower triangle of a 5×5 or a 10×10 correlation matrix.
3. For the correlation matrix generated in (2), determine the value of R^2 using matrix data input in SPSS.
4. Check whether the R^2 in (3) is as desired (within $R^2 \pm .05$). If R^2 is as desired, then use the correlation matrix for data generation. If not, then randomly select another correlation from (1) to replace one value in the correlation matrix. The correlation to be replaced was selected on the basis of whether the value of R^2 in (4) was more or less than the desired value.
5. Repeat (4) until the desired R^2 is obtained.

The population correlation matrix for the model, $Y = .140X_1 + .206X_2 + .244X_3 + .137X_4$, having $R^2 = 0.19264$, was as follows:

$$\begin{array}{c} \begin{array}{ccccc} Y & X_1 & X_2 & X_3 & X_4 \end{array} \\ \begin{bmatrix} 1 & & & & \\ .26 & 1 & & & \\ .25 & .14 & 1 & & \\ .30 & .26 & .01 & 1 & \\ .23 & .20 & .16 & .13 & 1 \end{bmatrix} \end{array}$$

For the four predictor model, $Y = .321X_1 + .253X_2 + .333X_3 + .220X_4$, having $R^2 = 0.59437$, the correlation matrix was as follows:

$$\begin{array}{c} \begin{array}{ccccc} Y & X_1 & X_2 & X_3 & X_4 \end{array} \\ \begin{bmatrix} 1 & & & & \\ .52 & 1 & & & \\ .51 & .34 & 1 & & \\ .60 & .32 & .26 & 1 & \\ .45 & .03 & .28 & .45 & 1 \end{bmatrix} \end{array}$$

The population correlation matrix for $Y = .113X_1 + .060X_2 + .140X_3 + .056X_4 + .097X_5 + .128X_6 + .144X_7 + .168X_8 + .130X_9$, having $R^2 = .21267$ was as follows:

$$\begin{array}{c}
\begin{matrix} Y & X_1 & X_2 & X_3 & X_4 & X_5 & X_6 & X_7 & X_8 & X_9 \end{matrix} \\
\begin{bmatrix}
1 & & & & & & & & & \\
.21 & 1 & & & & & & & & \\
.15 & .04 & 1 & & & & & & & \\
.21 & .16 & .01 & 1 & & & & & & \\
.20 & .19 & .14 & .13 & 1 & & & & & \\
.18 & .10 & .09 & .02 & .17 & 1 & & & & \\
.19 & .11 & .13 & .07 & .08 & .03 & 1 & & & \\
.22 & .18 & .15 & .09 & .16 & .07 & .06 & 1 & & \\
.23 & .05 & .08 & .01 & .20 & .13 & .04 & .05 & 1 & \\
.21 & .03 & .12 & .14 & .10 & .14 & .07 & .02 & .11 & 1
\end{bmatrix}
\end{array}$$

For the nine predictor model, $Y = .119X_1 + .304X_2 + .053X_3 + .264X_4 + .053X_5 + .336X_6 - .284X_7 + .260X_8 + .024X_9$ having $R^2 = 0.58590$, the correlation matrix was as follows:

$$\begin{array}{c}
\begin{matrix} Y & X_1 & X_2 & X_3 & X_4 & X_5 & X_6 & X_7 & X_8 & X_9 \end{matrix} \\
\begin{bmatrix}
1 & & & & & & & & & \\
.21 & 1 & & & & & & & & \\
.45 & .08 & 1 & & & & & & & \\
.35 & .26 & .32 & 1 & & & & & & \\
.46 & .21 & .12 & .31 & 1 & & & & & \\
.21 & .18 & .14 & .16 & .09 & 1 & & & & \\
.55 & .10 & .39 & .32 & .37 & .13 & 1 & & & \\
.13 & .28 & .41 & .42 & .23 & .39 & .30 & 1 & & \\
.45 & .11 & .22 & .34 & .20 & .48 & .16 & .18 & 1 & \\
.34 & .20 & .36 & .07 & .10 & .12 & .24 & .19 & .46 & 1
\end{bmatrix}
\end{array}$$

With the resulting correlation matrices, EQS (Bentler, 1989) was used to generate 1000 random samples with (skew, kurtosis) = (0, 0), (1, 3), (1.8, 6), (3, 25) with three sample sizes of 94, 153, and 265. The seed for data generation was 524697.

Simulation Program

A simulation program was written that created the pattern of missing data, performed a simultaneous linear multiple regression analysis upon implementing a MDT, and appended the parameter estimates and standard errors of regression coefficients in an output file. The routine for creating missing pattern was verified manually by examining hard copies of a set of data files. For MS, the mean of the variable was substituted for all the missing cases in that variable. For PW, the mean and variance for each variable were estimated using all available observations for that variable, and the covariances between pairs of variables were estimated using only cases that were complete in both variables. For RS, the missing values were predicted from a linear regression equation between X_i and the predictor with no missing values, X_1. For EM, the initial means, variances, and covariances were based on the values under RS. Figure 3 shows a flowchart of the EM method for imputing the missing values in X_i as implemented in the simulation program. The accuracy of MS and PW algorithms were checked using SPSS statistical program (SPSS Inc., 1993). The accuracy of RS was checked using BMDP AM (Dixon, 1988) with the command

```
/ESTIMATE
    METHOD = REGR
    TYPE = COMPLETE
```

Similarly, the accuracy of the EM algorithm was checked using BMDP AM (Dixon, 1988) with the command

```
/ESTIMATE
    METHOD = REGR
    TYPE = ML
```

Considering that the matrix **X′X** may not be positive definite under pairwise deletion method, the simulation program was tailored to flag such samples. The Gauss-Jordan decomposition approach, adapted from Numerical Recipes (Press, Teukolsky, Vetterling, & Flannery; 1992), was used for matrix inversion. As described in Marascuilo and Levin (1983), the matrix of predictors **X** was not augmented by a column of ones in the first column because the estimation of β_0 was not required. By partitioning the variance-covariance matrix of the original data set into $\hat{\Sigma}_{xx}$ and $\hat{\Sigma}_{xY}$, the vector of regression coefficients was calculated using the equation $\mathbf{b} = \hat{\Sigma}^{-1}_{xx} \hat{\Sigma}_{xY}$. The p standard errors of $b_1, b_2, \ldots, b_p$ were calculated using

$$\sqrt{\frac{MS_{res} \hat{\Sigma}^{-1}_{xx}}{N-1}}$$

where MS_{res} denotes the estimate of the residual variance about the estimated regression equation for p predictor variables and N is the sample size. The parameter estimates from the simulation program were congruent to those from BMDPAM (Dixon, 1988) to six decimal places.

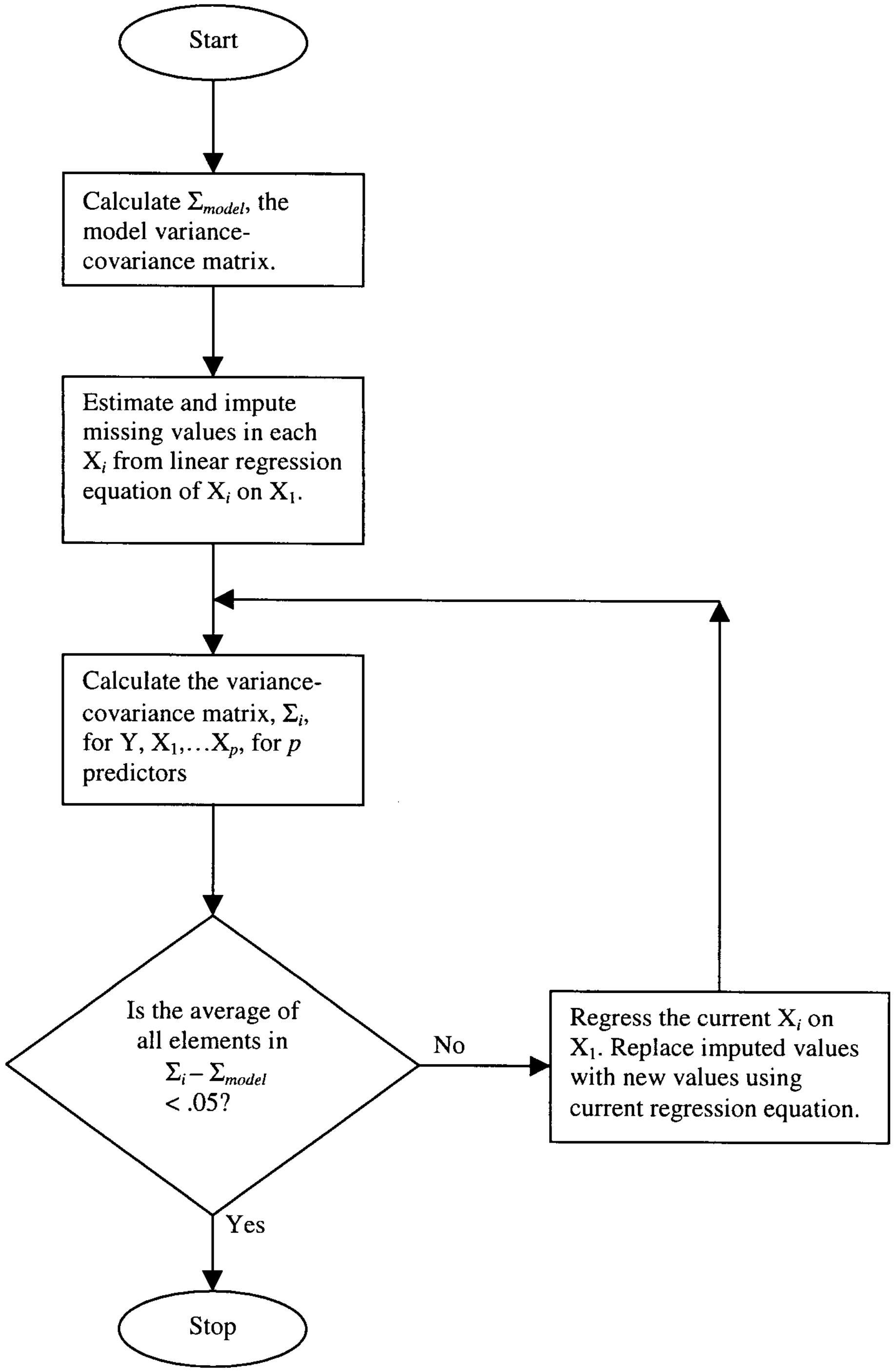


Figure 3: Flowchart for EM method used in the study

Evaluation Criteria

According to Larsen and Marx (1981), four qualities of a good estimator are unbiasedness, efficiency, consistency, and sufficiency. Unbiasedness is one way to decide whether or not an estimator's distribution is suitably centered. In other words, if θ is a parameter to be estimated, then its estimator W should be somewhat "centered" with respect to θ. If it is not, the estimator will tend either to overestimate or underestimate θ, a condition that is not desirable. An estimator is said to be unbiased if "on the average", it will yield the true parameter value; that is, if the underlying experiment is repeated with infinitely many samples of size T, the average value of the estimates from all those samples will be equal to the true value. More formally, let $Y_1, Y_2, \ldots, Y_n$ be a random sample from $f_Y(y; \theta)$. An estimator $W = h(Y_1, Y_2, \ldots, Y_n)$ is said to be unbiased if $E(W) = \theta$, for all θ. The term bias should be used to describe a situation in which there is a tendency for an estimator to consistently overestimate or underestimate parameters across a large number of samples.

Unbiased estimators are not unique (Larsen & Marx, 1981). This implies that an unbiased estimator may not necessarily be the best. In order to select the best estimator, we need to compare their dispersions and choose the estimator with the smallest dispersion, so that the probability of its being close to the true θ will be large (Larsen & Marx, 1981). This property of a good estimator is called efficiency. Put differently, suppose W_1 and W_2 are two unbiased estimators of θ. Then, W_1 is said to be more efficient than W_2 if the variability of W_1 is less than that of W_2.

Although the concepts of unbiasedness and efficiency lead to the most basic characterizations of point estimates, consistency is another important property of a good estimator. Consistency is attractive because it says that as the sample size increases indefinitely, the distribution of the estimator becomes entirely concentrated at the parameter value (Goldberger, 1991). Roughly speaking, an estimator is consistent if W_n, the n^{th} member of an infinite sequence of estimators, W_1, W_2,..., W_n,..., lies arbitrarily close to the parameter being estimated, as n becomes large (Larsen & Marx, 1981). This has two immediate implications: (1) W_n is asymptotically unbiased and (2) the variance of W_n converges to 0. In mathematical terms, an estimator $W_n = h(Y_1, Y_2, \ldots, Y_n)$ is said to be consistent (for θ) if it converges stochastically to θ, that is, if for all $\varepsilon > 0$ and $\delta > 0$, there exists an $n(\varepsilon, \delta)$ such that $P(|W_n - \theta| < \varepsilon) > 1 - \delta$, for $n > (\varepsilon, \delta)$.

Sufficiency is the fourth property of a good estimator. It is related to the amount of "information" a given estimator contains. If $Y_1, Y_2, \ldots, Y_n$ is a random sample from $f_Y(y; \theta)$, then the statistic $W = h(Y_1, Y_2, \ldots, Y_n)$ is said to be sufficient for θ if, for all θ and all possible sample points, the conditional probability density function of $Y_1, Y_2, \ldots, Y_n$ given w does not depend on θ, either in the function itself or in the function's domain.

Whereas the present study was not concerned with the evaluation of estimators, there exists some correspondence between the properties of a good estimator and the evaluation criteria applied in the study. The two criteria used in the evaluation of MDTs in the present study were accuracy (precision) and bias of parameter estimates. As noted by Roth (1994), the accuracy of parameter estimates is usually the criterion for the evaluation of MDTs. The concept of accuracy of parameter estimates is not equivalent to that of efficiency of an estimator, although the two concepts are related because they both make use of dispersion procedures. For example, it is possible to investigate the relative accuracy of two estimators, the mean and the median. But when only one estimator is used as in the present study, it is the spread of the parameter estimates around a population value that is of interest, and the magnitude of the spread should depend on which MDT was used. Raymond and Roberts (1987) defined accuracy of parameter estimates as the amount of dispersion found around a population parameter.

Judge, Hill, Griffiths, Lütkepohl, and Lee (1988) stressed the importance of considering the bias of estimators. Note that the bias of estimators is similar to the bias of parameter estimates, but the two are not equivalent. The term bias in parameter estimates may be used to describe a situation in which there was a tendency for an MDT to consistently overestimate or underestimate parameters across a large number of samples following treatment of missing data by that MDT (Roth, 1994). In other words, analysis of bias in parameter estimates involves the examination of any consistent errors due to MDTs. Judge et al. (1988) noted that the estimator with the smallest dispersion (i.e., the most accurate), should be preferred even if it is biased. In a similar manner, the MDT under which the parameter estimate of interest has the smallest dispersion around the true value should be preferred even if it is biased. Although Finkbeiner (1979) used both accuracy and bias of parameter estimates for assessing the optimal performance of MDTs, the best MDT was that with the smallest dispersion regardless of bias.

There is no consistency among researchers on which measure of dispersion is the most preferred in the evaluation of MDTs. Based on my literature review, previous researchers have used Euclidean distance, root mean square error (RMSE), mean square error (MSE), and residual sum of squares of deviations, among others. According to Judge et al. (1988) and Barford (1985), two important methods for establishing estimation accuracy are mean square error (MSE) and mean absolute error (MAE), the choice of which one to use being arbitrary. Either way we obtain a positive quantity. The smaller either quantity is, the greater the accuracy of the parameter estimate. Intuitively, both MSE and MAE represent the average loss or risk incurred in the estimation of the parameter. Lyon (1970) noted that MSE and MAE are linearly related, with MSE being about 1.25 times larger than MAE when sample size is greater than 4. The advantage of MSE over MAE is that MSE is more mathematically tractable, a characteristic not required in the present study. Shchigolev (1960/1965) noted that MAE offers an excellent option as a criterion for evaluating the accuracy of parameter estimates if the quality of the measurements is poor, a condition that is quite likely with missing data. For this reason, MAE was used as the criterion for MDT evaluation in the present study. The MAE was defined as

$$\frac{1}{B}\sum_{b=1}^{B}\left|\hat{\theta}_b - \theta\right|$$

where B is the number of parameter estimates (1,000 in this study), $\hat{\theta}_b$ is the b^{th} parameter estimate, and θ is the population parameter. Bias was defined as follows:

$$\text{Bias} = E(\hat{\theta}) - \theta$$

where $E(\hat{\theta})$ is the mean of the 1,000 parameter estimates per cell, and θ is the true parameter value.

In the third and last part of the analysis, the performance of MDTs was assessed with respect to actual coverage of confidence intervals when nominal alpha is .05. This was achieved by first constructing confidence intervals using each sample regression coefficient and its corresponding standard error. The formula used in the construction of the confidence intervals was $[b_i - 1.96SE(b_i)] \le \beta_i \le [b_i + 1.96SE(b_i)]$, where b_i is the sample estimate of β_i, the population regression coefficient for predictor variable X_i, and $SE(b_i)$ is the standard error of b_i. All confidence intervals spanning the actual population regression coefficient used in data generation were counted and converted into percentages. Good coverage probabilities were those that were closest to 95%. The best MDT was selected on the basis of magnitude of mean absolute error of estimation. However, if the corresponding bias was the least, then the performance of that MDT was considered optimal.

Analysis

There were four analyses. First, as in Milligan (1980, 1981, 1989a) and Donoghue (1995), a fully crossed factorial analysis of variance was conducted in order to determine the effects of pattern, percent missing and non-normality on absolute error of estimation across different sample sizes. The independent variables were missing pattern, percent missing, and non-normality. Second, MAE for parameter estimates in each cell was calculated and graphed.

Third, bias in parameter estimates in each cell was calculated and graphed. Fourth, mean coverage probability was calculated and graphed. The graphs were used as an aid for evaluating the performance of MDTs.

Given the very large amount of data (72,000 regression analyses for each MDT), many of the main effects and some two-way interactions were statistically significant ($p < .001$). The purpose of the ANOVA was to summarize the data and help to highlight important effects. Therefore, a measure of effect size was adopted in place of traditional significance testing. Usually, η^2 would be used in this context, where

$$\eta^2 = \frac{SS_{effect}}{SS_{Tot}}.$$

As noted by Donoghue (1995), this index has a disadvantage in large designs in that the denominator contains not only error variance and systematic variance of interest, but also irrelevant systematic variance of other factors in the design. The larger the design becomes, the more apparent this effect becomes. This defect was worsened by the fact that some factors had very large effects, thus obscuring the effects of other factors. For this reason, as recommended by Donoghue, the following alternate equation for effect size was used:

$$\eta_{alt}^2 = \frac{SS_{effect}}{SS_{effect} + SS_{error}}.$$

Considering that statistically significant effects may not necessarily be of practical significance, a practical criterion of $\eta^2_{\text{alt}} \geq .03$ was selected. Although this criterion is somewhat arbitrary, it was chosen because computer intensive studies in other areas have used similar criteria (e.g., $\omega^2 \geq .03$ was used by Anderson & Gerbing, 1984, and Gerbing & Anderson, 1985; $\eta^2_{\text{alt}} \geq .03$ was used by Donoghue, 1995).

Summary

In the present chapter, the advantages of simulation methodology were highlighted. The greatest advantage of simulation methodology is that it allows the investigator to manipulate design conditions that would otherwise be impossible to include in a study.

Issues in data generation were also presented. It was noted that the choice of random number generator is crucial in any simulation study of this nature. Higher-order congruential generators are generally preferred to multiplicative linear congruential generators.

Several factors were identified that may influence the performance of MDTs in a multiple linear regression context. These include sample size, missing pattern, percent missing, non-normality, size of multiple R^2, and number of predictors. It was also noted that the selection of typical levels of design factors is necessary to enhance generalizability of findings. In the present study, the choice of levels of design factors was based on a review of four psycho-educational research journals. However, the review did not help in the choice of levels of non-normality. Previous research on typical levels of non-normality was used in this case.

The design of the present study was discussed. Four separate studies were conducted in the investigation (2 levels of number of predictors by 2 levels of multiple R^2). In each study, the independent variables were missing pattern (2 levels), percent missing (3 levels), and non-normality (4 levels). The within-factor was MDT (mean substitution, pairwise deletion, regression imputation, and EM method).

Upon generation of data using EQS (Bentler, 1993) at 3 levels of sample size (N = 94, 153, 265), a customized simulation program was used to delete part of the data to create monotonic and non-monotonic patterns. The program was then used to impute the deleted data with different MDTs. The simulation program ran a simultaneous multiple linear regression analysis. The standard multiple regression output (multiple $R^2_{estimate}$, regression coefficients, and standard errors of the regression coefficients) was saved and imported into SPSS for analysis.

A 2 × 3 × 4 fully crossed factorial ANOVA was conducted using absolute error of estimation for $R^2_{estimate}$ and regression coefficients as the dependent variables. The design factors were missing pattern, percent missing, and non-normality. Analysis was at each level of sample size. The effect size (η^2) was chosen as a criterion for practical significance. Accuracy, as measured by mean absolute error of estimation for $R^2_{estimate}$ and mean absolute error of estimation for regression coefficients, was used to evaluate the MDTs. Bias in $R^2_{estimate}$ and regression coefficients was used to determine which MDTs yielded estimates "close" to the true parameter. Lastly, coverage probability for regression coefficients was used to determine the Type I error rates under each MDT. Results are presented in the next chapter.

CHAPTER 4

RESULTS

In this chapter, results from the four studies are presented in the following order:

1. Study 1: Four predictors under low R^2 condition
2. Study 2: Four predictors under high R^2 condition
3. Study 3: Nine predictors under low R^2 condition
4. Study 4: Nine predictors under high R^2 condition.

Tables and graphs for mean absolute error are presented for monotonic and non-monotonic missing patterns, arranged in increasing order of sample size, percent missing, and non-normality. For ease of comparison, results in tables and graphs are consistent across studies, and graphical plots at each level of sample size have the same scale. Similarly, tables and graphs for bias in parameter estimates are provided in Appendix A and B, respectively.

Study 1: Four Predictors under Low R^2 Condition

Effects on Absolute Error for $R^2_{estimate}$

The effects of design variables on absolute error for $R^2_{estimate}$ are listed in Table 5. Using $\eta^2_{alt} \geq .03$ as the criterion, the table shows that none of the effects was of practical significance under all MDTs with samples of size 94 and 153. Under sample size 265, the effect of percent missing on error of estimation was of practical significance under EM and RS, the largest effect being under EM.

Relative Performance of MDTs in the Estimation of Population R^2

Table 6 contains means and standard deviations of absolute error for $R^2_{estimate}$. The influence of non-normality on the performance of MDTs at sample size 94 is revealed in Figure 4. The performance of all MDTs as measured by the mean absolute error of estimation deteriorated with increasing non-normality. Under monotonic missing pattern, PW treatment of missing data was the worst and EM treatment the best. However, there was little differentiation in the performance of EM, RS and MS, especially when data were normal (skew = 0 and kurtosis = 0). This differentiation increased with increasing non-normality. Under non-monotonic pattern, it was again EM that outperformed other MDTs, followed by RS, MS and PW, in that order. This finding was similar to that for monotonic pattern, except that MDTs differed more in performance under non-monotonic pattern than under monotonic pattern. MDTs differed the least when data were normal, and differed most at the highest level of non-normality (skew = 3 and kurtosis = 25). EM and RS methods performed almost equally well.

Table 5: Effects on absolute error for $R^2_{estimate}$ fulfilling the $\eta^2_{alt} \geq .03$ criterion

Treatment	Sample Size	Effect	df	SS_{effect}	SS_{error}	η^2_{alt}
EM	94	-	-	-	-	-
	153	-	-	-	-	-
	265	Percent Missing	2	1.77	48.67	.035
MS	94	-	-	-	-	-
	153	-	-	-	-	-
	265	-	-	-	-	-
PW	94	-	-	-	-	-
	153	-	-	-	-	-
	265	-	-	-	-	-
RS	94	-	-	-	-	-
	153	-	-	-	-	-
	265	Percent Missing	2	1.77	50.24	.034

Note: EM = Expectation-maximization method, MS = Mean substitution, PW = Pairwise deletion, RS = Regression imputation,
- = No effect of practical significance.

The influence of non-normality on the performance of MDTs at sample size 153 is revealed in Figure 5. The findings were similar to those under sample size 94, except that mean absolute error of estimation under sample size of 153 was much smaller. Figure 6 shows the influence of non-normality on the performance of MDTs at sample size 265. In this case the mean absolute error of estimation was much smaller than under sample size 94 and 153.

Table A.1 in Appendix A contains bias in $R^2_{estimate}$ under low R^2 condition with four predictors. The bias in $R^2_{estimate}$ across all sample sizes varied from –.04 to .06, a range of .10. Figure B.1 in Appendix B is a set of typical graphical plots for bias in $R^2_{estimate}$ (N = 94) showing that all the MDTs had positive bias at 10% missing. At this level, the smallest bias was under EM. At 15% missing, the smallest bias was under EM for monotonic pattern.

Table 6
Means and standard deviations* of absolute error for $R^2_{estimate}$

			Monotonic Missing Pattern				Non-monotonic Missing Pattern			
N	Percent Missing	Normality (sk., kurt.)	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation
94	10	(0,0)	0589(045)	0598(046)	0635(050)	0594(046)	0532(041)	0585(046)	0616(047)	0548(042)
		(1,3)	0628(050)	0639(051)	0684(053)	0634(050)	0550(041)	0622(047)	0657(049)	0568(042)
		(1.8,6)	0701(057)	0714(058)	0761(060)	0708(058)	0607(045)	0691(053)	0726(056)	0624(047)
		(3,25)	0885(080)	0912(082)	0965(087)	0898(081)	0706(059)	0852(072)	0894(077)	0719(060)
	15	(0,0)	0570(043)	0583(044)	0630(049)	0577(043)	0526(038)	0575(044)	0603(046)	0523(039)
		(1,3)	0612(046)	0628(048)	0672(052)	0621(047)	0570(043)	0609(046)	0642(049)	0540(039)
		(1.8,6)	0678(054)	0697(056)	0745(059)	0688(055)	0681(056)	0665(051)	0712(055)	0584(044)
		(3,25)	0850(076)	0886(080)	0947(085)	0865(078)	0432(031)	0821(069)	0884(078)	0689(057)
	20	(0,0)	0557(041)	0576(042)	0620(048)	0565(041)	0496(036)	0561(042)	0599(045)	0506(037)
		(1,3)	0602(044)	0623(046)	0664(050)	0612(045)	0515(037)	0595(044)	0636(048)	0524(038)
		(1.8,6)	0666(052)	0691(055)	0735(057)	0678(053)	0554(042)	0647(050)	0709(054)	0563(042)
		(3,25)	0824(072)	0869(077)	0931(085)	0842(074)	0667(054)	0792(067)	0910)108)	0672(055)
153	10	(0,0)	0457(034)	0462(035)	0487(038)	0460(034)	0421(032)	0453(034)	0471(036)	0431(032)
		(1,3)	0474(036)	0480(037)	0505(040)	0477(037)	0413(032)	0457(036)	0483(038)	0423(033)
		(1.8,6)	0526(042)	0534(042)	0564(044)	0530(042)	0444(035)	0496(040)	0535(042)	0455(036)
		(3,25)	0683(058)	0700(060)	0723(061)	0690(059)	0521(042)	0619(052)	0669(060)	0528(043)
	15	(0,0)	0448(033)	0453(033)	0483(037)	0451(033)	0432(031)	0460(034)	0461(036)	0439(032)
		(1,3)	0468(034(	0475(035)	0500(039)	0472(035)	0418(032)	0455(035)	0472(037)	0423(033)
		(1.8,6)	0514(039)	0525(040)	0555(044)	0520(040)	0442(034)	0487(039)	0523(041)	0447(035)
		(3,25)	0662(056)	0685(058)	0710(060)	0671(057)	0528(040)	0602(050)	0661(060)	0531(041)
	20	(0,0)	0461(033)	0466(034)	0481(037)	0464(033)	0456(032)	0469(034)	0454(035)	0456(032)
		(1,3)	0480(034)	0489(035)	0495(039)	0485(035)	0455(032)	0464(035)	0469(036)	0453(033)
		(1.8,6)	0519(038)	0534(040)	0547(043)	0526(039)	0476(034)	0499(038)	0519(047)	0475(034)
		(3,25)	0656(054)	0685(057)	0698(059)	0666(055)	0558(040)	0604(048)	0669(062)	0558(040)
265	10	(0,0)	0346(027)	0348(027)	0359(027)	0347(027)	0341(025)	0349(026)	0346(027)	0344(026)
		(1,3)	0351(027)	0353(027)	0372(029)	0352(027)	0338(025)	0352(027)	0357(027)	0341(026)
		(1.8,6)	0383(030)	0386(030)	0413(032)	0385(030)	0359(027)	0377(028)	0397(030)	0363(027)
		(3,25)	0483(038)	0491(039)	0521(043)	0487(039)	0415(031)	0453(036)	0488(039)	0417(032)
	15	(0,0)	0361(028)	0370(028)	0369(029)	0370(028)	0372(027)	0366(026)	0340(026)	0370(026)
		(1,3)	0368(028)	0370(028)	0369(029)	0370(028)	0377(027)	0368(027)	0353(027)	0374(027)
		(1.8,6)	0397(030)	0400(030)	0409(032)	0399(030)	0394(028)	0391(029)	0394(029)	0392(028)
		(3,25)	0484(037)	0495(038)	0513(042)	0489(038)	0453(031)	0459(034)	0485(039)	0451(031)
	20	(0,0)	0393(029)	0390(029)	0355(027)	0392(029)	0422(028)	0388(027)	0337(026)	0415(028)
		(1,3)	0400(029)	0398(029)	0367(028)	0400(029)	0427(028)	0389(027)	0353(027)	0420(028)
		(1.8,6)	0424(030)	0425(031)	0405(031)	0425(031)	0443(029)	0411(028)	0394(029)	0437(029)
		(3,25)	0496(036)	0508(037)	0504(041)	0502(037)	0508(032)	0482(034)	0490(039)	0506(032)

* All values in the table are preceded by an omitted decimal point. Standard deviations are in parentheses. Standard errors of the means can be obtained by dividing each standard deviation by $1000^{½}$.

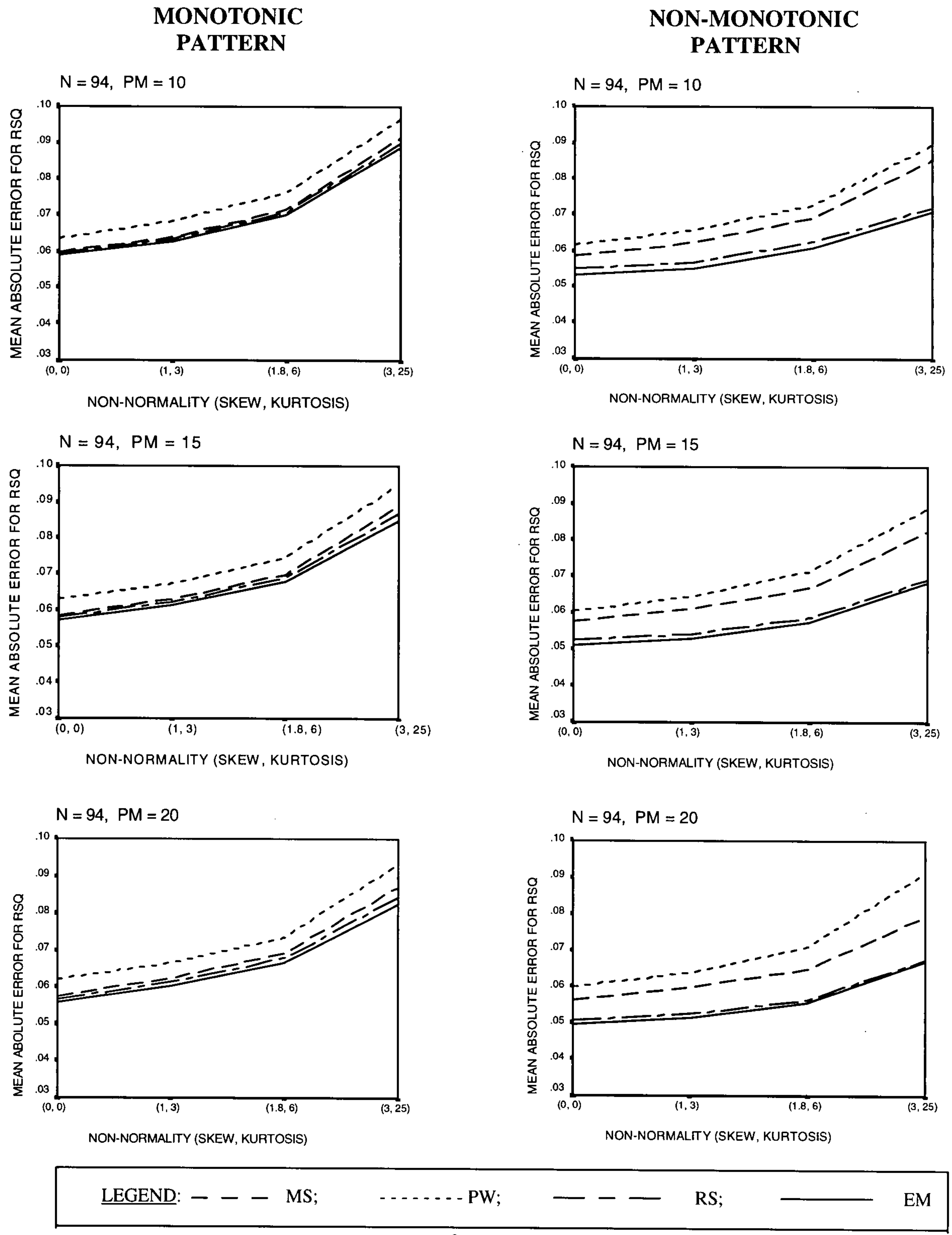


Figure 4: Mean absolute error for $R^2_{estimate}$ across levels of non-normality (N = 94)
Note: N = Sample size, PM = Percent missing, RSQ = $R^2_{estimate}$

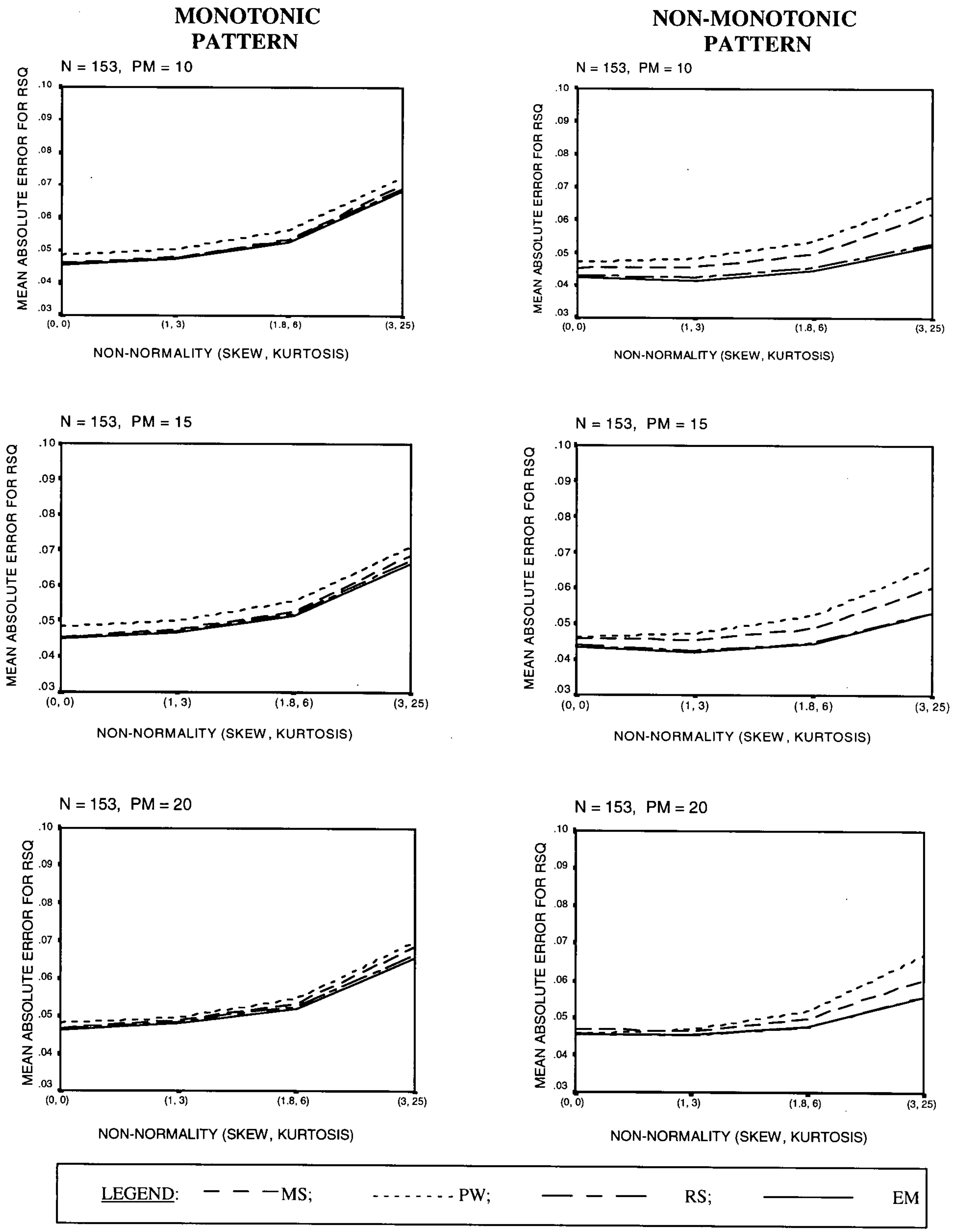


Figure 5: Mean absolute error for $R^2_{estimate}$ across levels of non-normality (N = 153)
Note: N = Sample size, PM = Percent missing, RSQ = $R^2_{estimate}$

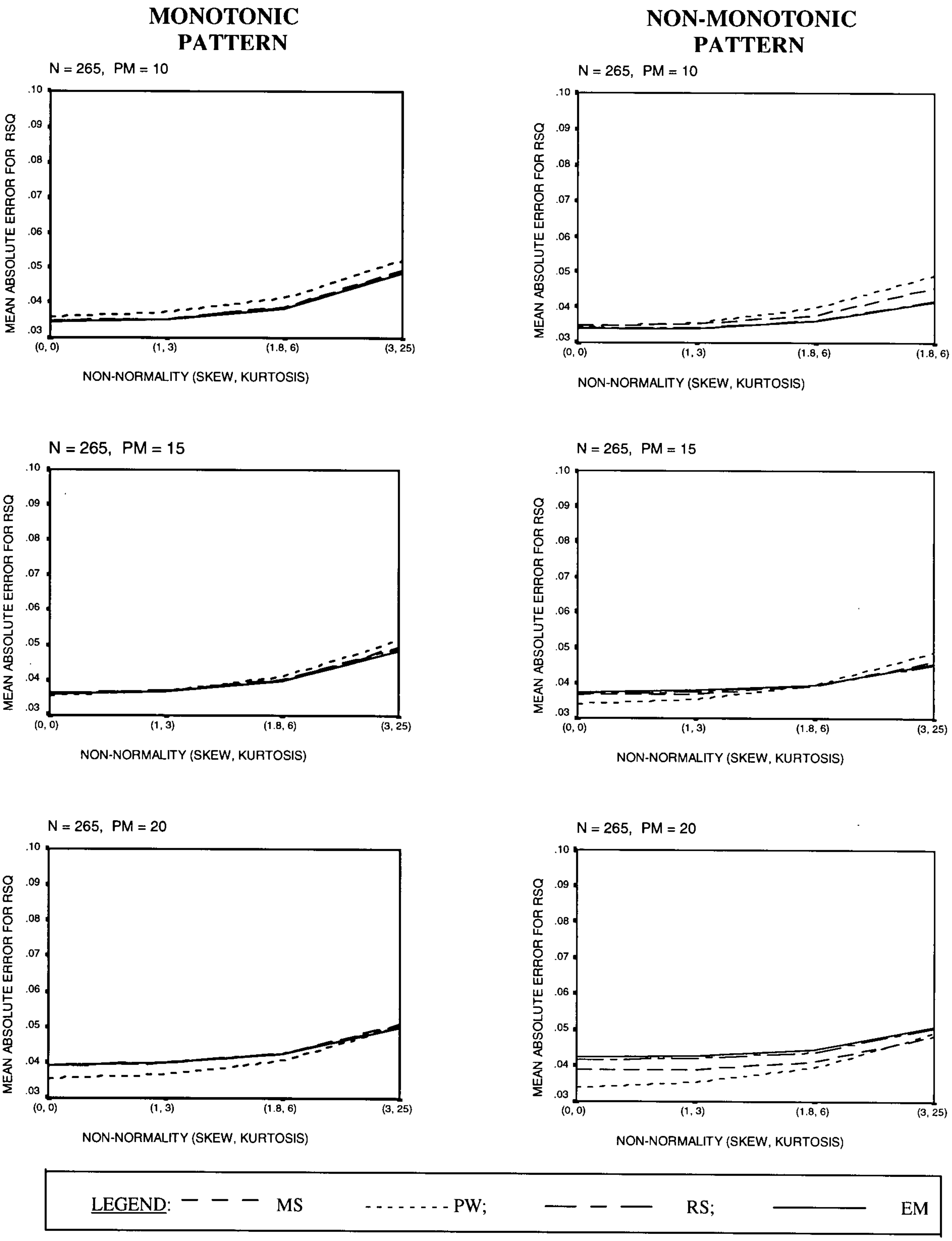


Figure 6: Mean absolute error for $R^2_{estimate}$ across levels of non-normality (N = 265)
Note: N = Sample size, PM = Percent missing, RSQ = $R^2_{estimate}$

Effects on Absolute Error for Regression Coefficients

The effects of design variables on absolute error for b_1 are listed in Table 7. For b_2 and b_4, no effect of practical significance was observed. For b_3, however, the effect of non-normality was of practical significance for all MDTs and at each level of sample size.

Using $\eta^2_{alt} \geq .03$ as the criterion, the table shows that for b_1, the effect of non-normality was of practical significance under EM and RS. The effect of pattern of missing data was of practical significance under MS and PW. Also, the effect of pattern of missing data was the strongest, with little differentiation between EM and RS where its effect was of practical significance.

Table 7
Effects on absolute error for regression coefficients fulfilling the criterion of $\eta^2_{alt} \geq .03$*

Sample size	MDT	b_1	b_2	b_3	b_4
94	EM	-	-	NL (033)	-
	MS	-	-	NL (034)	-
	PW	-	-	NL (034)	-
	RS	-	-	NL (034)	-
153	EM	NL (031)	-	NL (036)	-
	MS	PT (030)	-	NL (037)	-
	PW	PT (030)	-	NL (037)	-
	RS	NL (031)	-	NL (037)	-
265	EM	NL (038)	-	NL (041)	-
	MS	PT (080)	-	NL (043)	-
	PW	PT (079)	-	NL (043)	-
	RS	NL (041)	-	NL (042)	-

* Values of η^2_{alt} are in parentheses preceded by an omitted decimal point.
Note: EM = EM imputation, MS = Mean substitution, PW = Pairwise deletion, RS = Regression imputation, PT = Pattern, NL = Non-normality, - = No effect of practical significance.

Relative Performance of MDTs in the Estimation of Regression Coefficients

Table 8 contains mean absolute error and the corresponding standard deviation for b_1. MDTs provided similar results for b_2, b_3, and b_4, and for this reason, only results for b_1 are presented. The influence of non-normality on the performance of MDTs at sample size 94 is revealed in Figure 7. The performance of all MDTs deteriorated with increasing non-normality as measured by mean absolute error for b_1. This was regardless of missing pattern. At sample size 94, under monotonic pattern of missing data, MDTs overlapped in pairs; MS and PW formed one pair and EM and RS formed another pair. This outcome was observed at all levels of percent missing. Under monotonic pattern of missing data, EM and RS consistently outperformed MS and PW.

At sample size 94 with non-monotonic pattern, again pairing of MDTs similar to those under monotonic pattern was observed, with EM and RS outperforming MS and PW. Whereas the mean absolute error of estimation under EM and RS was smaller with non-monotonic pattern than with monotonic pattern, the mean absolute error of estimation under MS and PW was smaller under monotonic pattern than under non-monotonic pattern. The influence of non-normality on the performance of MDTs at sample size 153 and 265 are revealed in Figures 8 and 9, respectively. The outcome was similar to that under sample size 94.

In sum, the best performance was under EM and RS, regardless of missing pattern. However, at each level of percent missing, the mean absolute error of b_1 under EM and RS was smaller under non-monotonic pattern than under monotonic pattern. Also, at each level of percent missing, the mean absolute error of estimation under MS and PW was smaller with monotonic pattern than with non-monotonic pattern.

Table A.2 in Appendix A shows bias in b_1. Figure B.2 in Appendix B shows graphs for the bias in b_1 ($N = 94$). The bias across all conditions of the design variables had a range of .16. The smallest bias was under RS across all levels of the design variables. However, the bias under EM was very close to that under RS. Whereas EM and RS consistently underestimated β_1, MS and PW, which had equivalent estimators, consistently overestimated β_1. The bias under non-monotonic missing pattern was generally larger than that under monotonic missing pattern. The maximum range in bias in estimating β_2, β_3, and β_4 was .04, a relatively smaller value compared to that of β_1. This indicates that the largest bias was for β_1. Therefore, graphs for β_2, β_3, and β_4 were omitted.

Table 8
Means and standard deviations* of absolute error of estimation for b_1

			Monotonic Missing Pattern				Non-monotonic Missing Pattern			
N	Percent Missing	Normality (sk., kurt.)	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation
94	10	(0,0)	0765(056)	0837(061)	0837(061)	0722(057)	0658(047)	0959(071)	0959(071)	0639(047)
		(1,3)	0828(062)	0899(068)	0897(068)	0834(063)	0701(049)	1024(081)	1023(080)	0674(049)
		(1.8,6)	0907(071)	0979(077)	0978(077)	0913(072)	0768(055)	1094(091)	1091(091)	0741(055)
		(3,25)	1030(087)	1109(101)	1108(101)	1035(089)	0919(064)	1295(128)	1292(128)	0895(065)
	15	(0,0)	0749(055)	0860(063)	0860(063)	0755(056)	0629(046)	1045(076)	1044(076)	0607(046)
		(1,3)	0816(061)	0920(070)	0921(069)	0820(062)	0680(048)	1103(086)	1101(086)	0656(048)
		(1.8,6)	0896(070)	0997(079)	0998(078)	0897(070)	0742(053)	1168(095)	1164(095)	0720(053)
		(3,25)	1021(088)	1128(106)	1128(105)	1019(089)	0913(067)	1380(134)	1372(133)	0912(070)
	20	(0,0)	0738(055)	0886(065)	0885(065)	0737(055)	0623(045)	1109(079)	1108(079)	0598(045)
		(1,3)	0803(061)	0941(072)	0941(072)	0802(061)	0673(049)	1160(089)	1158(089)	0648(049)
		(1.8,6)	0880(069)	1015(081)	1015(081)	0877070)	0747(055)	1220(099)	1214(099)	0726(054)
		(3,25)	1011(088)	1150(108)	1149(108)	1001(089)	0906(071)	1430(137)	1421(137)	0920(075)
153	10	(0,0)	0575(043)	0643(047)	0642(047)	0580(044)	0514(036)	0817(057)	0816(057)	0490(036)
		(1,3)	0620(046)	0681(051)	0681(051)	0621(046)	0558(037)	0861(066)	0860(066)	0526(037)
		(1.8,6)	0687(051)	0742(058)	0741(058)	0686(052)	0613(041)	0910(076)	0907(075)	0581(040)
		(3,25)	0820(071)	0873(078)	0873(078)	0814(071)	0731(051)	1120(105)	1114(105)	0703(050)
	15	(0,0)	0564(042)	0668(048)	0667(048)	0568(042)	0502(034)	0919(061)	0918(061)	0471(034)
		(1,3)	0604(045)	0707(053)	0705(053)	0606(045)	0548(036)	0962(070)	0961(070)	0513(035)
		(1.8,6)	0666(051)	0762(061)	0761(061)	0664(051)	0593(040)	1005(080)	1001(079)	0565(040)
		(3,25)	0800(070)	0895(082)	0893(083)	0790(070)	0709(051)	1215(109)	1207(109)	0695(052)
	20	(0,0)	0548(041)	0697(049)	0695(049)	0552(041)	0490(034)	0994(065)	0993(065)	0462(034)
		(1,3)	0590(044)	0732(055)	0730(055)	0591(044)	0525(036)	1034(073)	1033(073)	0502(035)
		(1.8,6)	0651(050)	0787(063)	0784(063)	0650(050)	0571(039)	1074(082)	1069(082)	0553(039)
		(3,25)	0783(068)	0911(085)	0910(085)	0771(067)	0691(051)	1272(112)	1264(111)	0690(053)
265	10	(0,0)	0439(033)	0491(036)	0491(036)	0441(033)	0390(028)	0710(046)	0710(046)	0359(027)
		(1,3)	0466(036)	0513(039)	0513(039)	0465(036)	0434(030)	0763(052)	0761(052)	0391(028)
		(1.8,6)	0517(039)	0560(044)	0560(044)	0515(040)	0471(032)	0796(058)	0793(058)	0431(031)
		(3,25)	0640(053)	0674(059)	0674(059)	0633(053)	0593(040)	0993(081)	0988(081)	0561(039)
	15	(0,0)	0428(032)	0517(037)	0517(037)	0430(032)	0385(027)	0826(050)	0825(050)	0346(026)
		(1,3)	0455(034)	0541(040)	0541(040)	0454(034)	0424(029)	0873(055)	0872(055)	0386(027)
		(1.8,6)	0508(037)	0580(046)	0580(046)	0502(038)	0460(031)	0903(061)	0899(061)	0427(030)
		(3,25)	0614(051)	0692(060)	0692(060)	0607(051)	0571(039)	1089(084)	1083(084)	0554(038)
	20	(0,0)	0415(031)	0545(039)	0545(039)	0412(031)	0385(027)	0915(052)	0915(052)	0350(026)
		(1,3)	0441(033)	0567(042)	0566(042)	0436(033)	0419(028)	0955(057)	0954(057)	0389(027)
		(1.8,6)	0488(036)	0604(047)	0604(047)	0480(036)	0448(031)	0982(063)	0977(063)	0423(030)
		(3,25)	0602(048)	0708(063)	0708(063)	0585(049)	0558(039)	1150(086)	1143(086)	0551(039)

* All values in the table are preceded by an omitted decimal point. Standard deviations are in parentheses. Standard errors of the means can be obtained by dividing each standard deviation by $1000^{½}$.

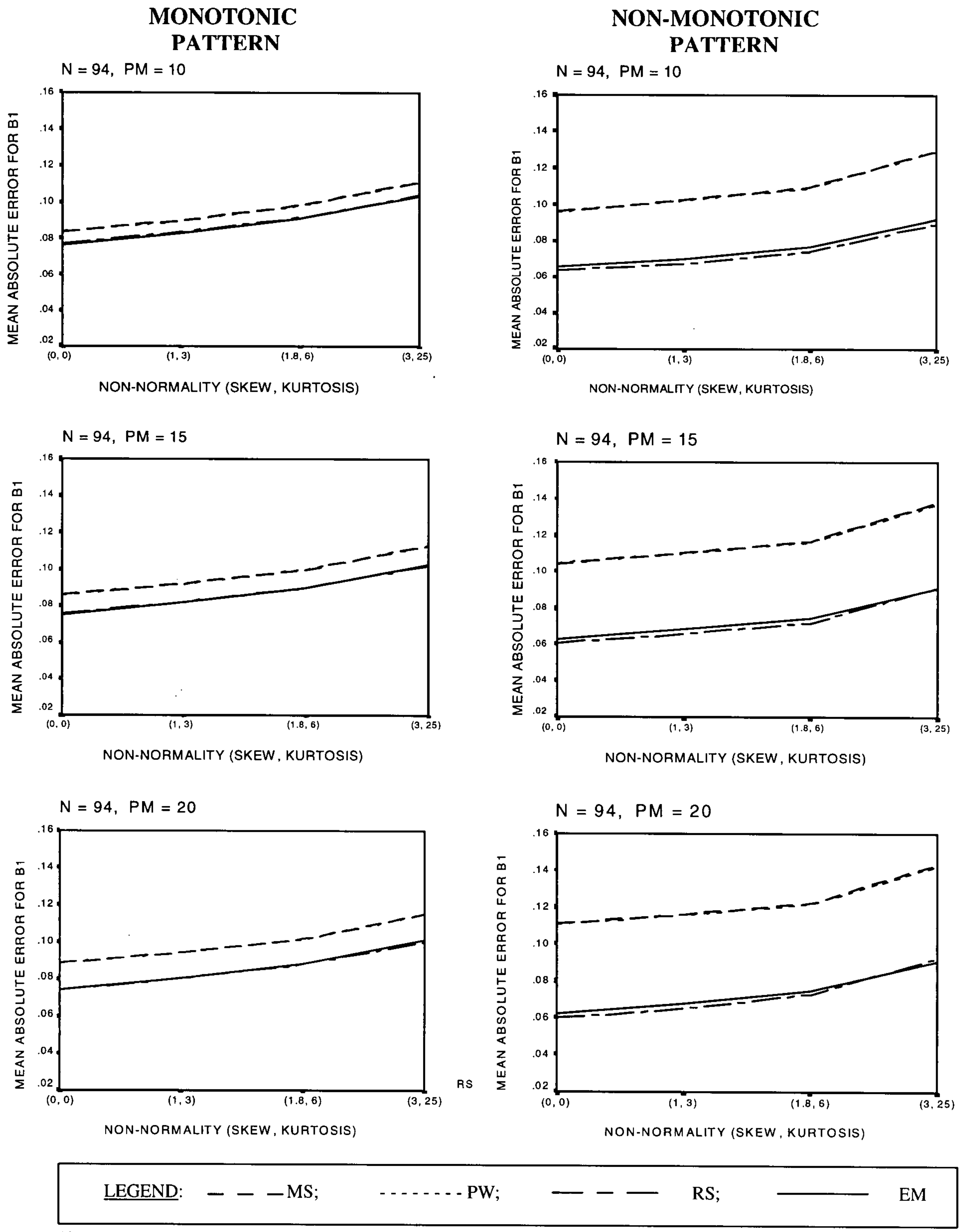


Figure 7: Mean absolute error for b_1 across levels of non-normality (N = 94)
Note: N = Sample size, PM = Percent missing.

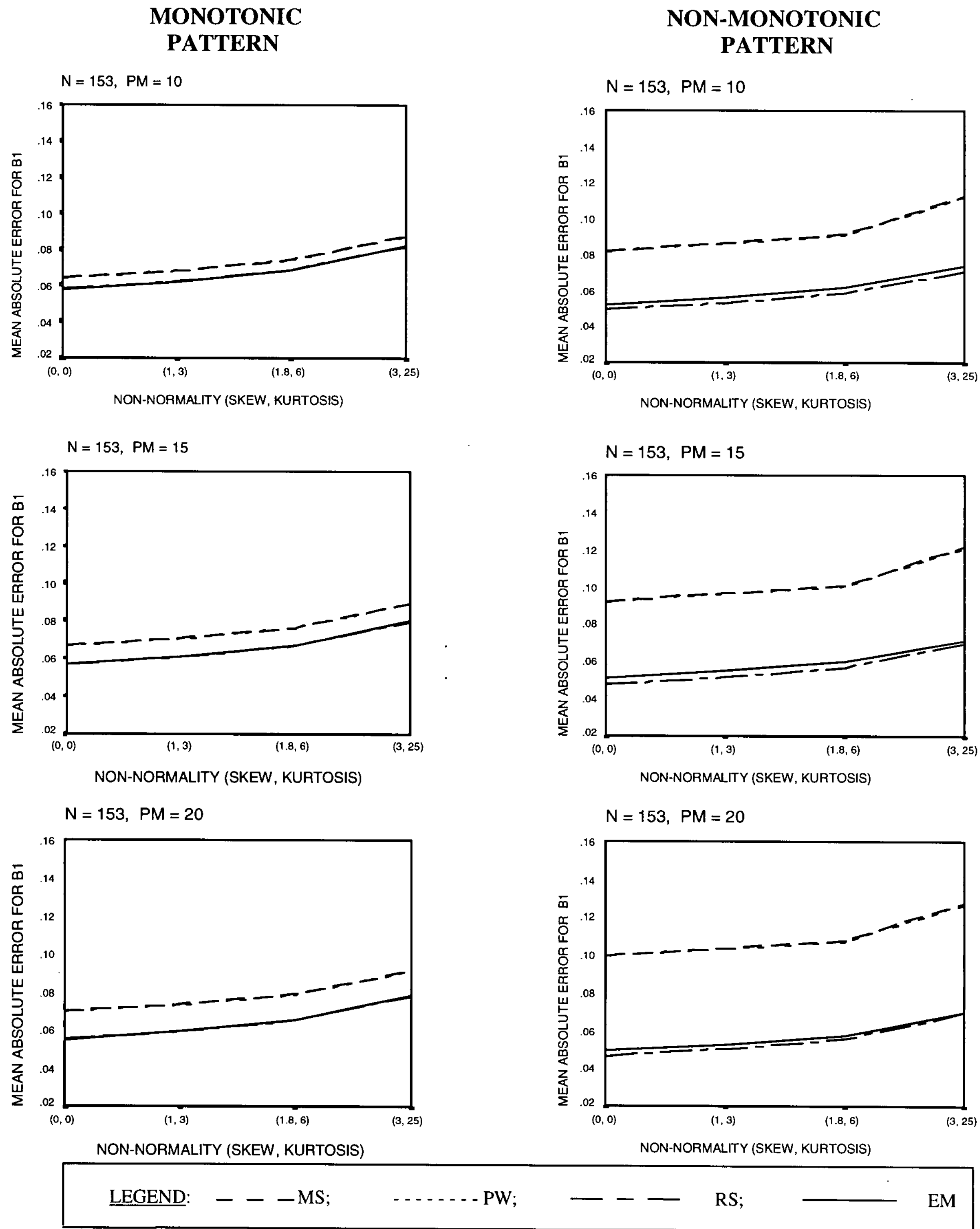


Figure 8: Mean absolute error for b_1 across levels of non-normality (N = 153)
Note: N = Sample size, PM = Percent missing

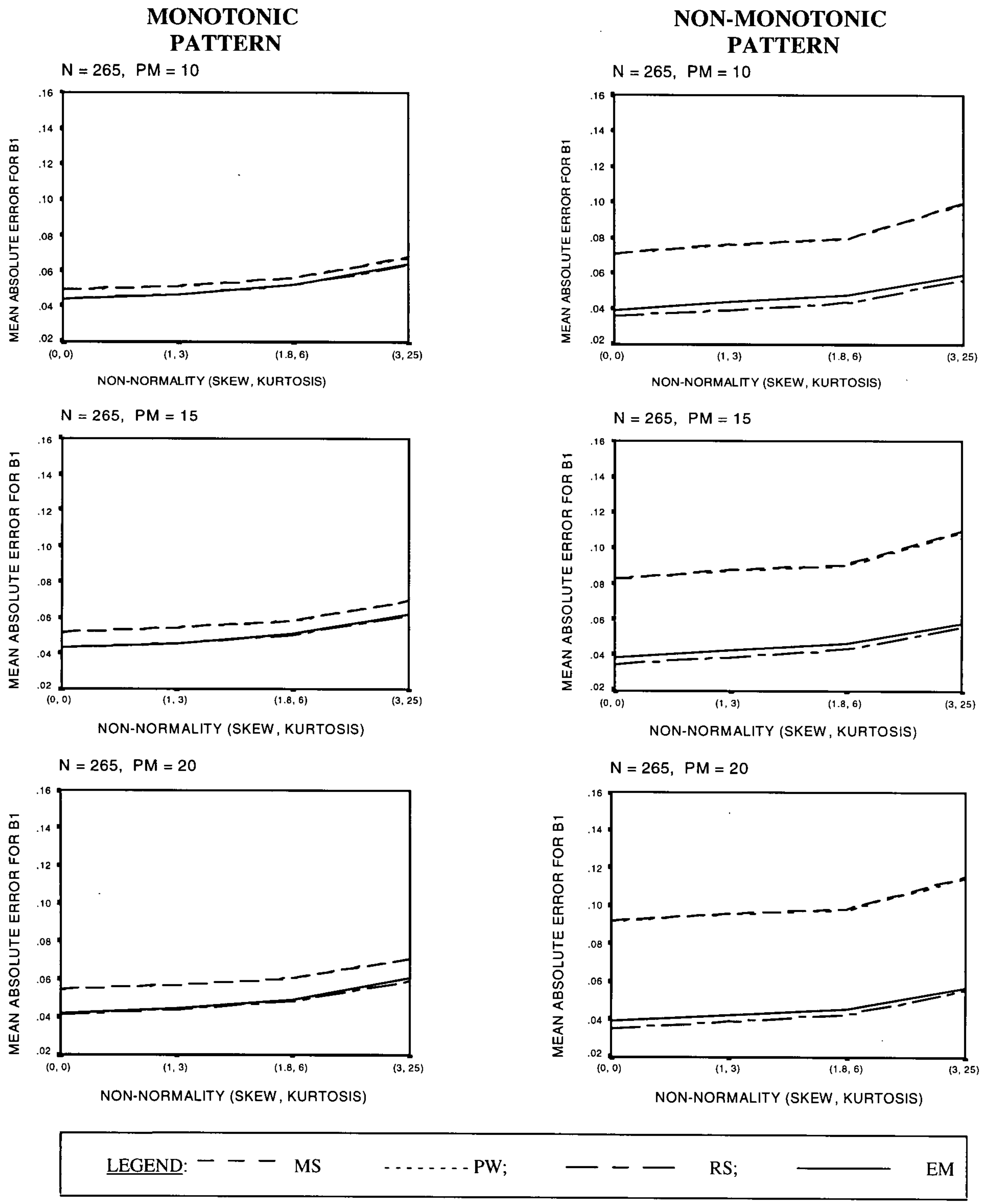


Figure 9: Mean absolute error for b_1 across levels of non-normality (N = 265)
<u>Note</u>: N = Sample size, PM = Percent missing

Coverage Probability for β_1

Coverage probabilities under monotonic pattern were found to differ from those under non-monotonic pattern only for β_1. MDTs were not differentiated in their coverage probabilities for β_2, β_3, and β_4, whose graphs looked the same and the averages extremely close. For this reason, only the coverage probabilities for β_1 are reported.

Table 9 contains mean coverage probabilities for β_1, under various design conditions. These means are plotted for sample size 94, 153 and 265 in Figures 10, 11, and 12, respectively. The graphs show that coverage probability decreased with increasing non-normality at all levels of sample size. Also, there was little differentiation in coverage probability between MS and PW, and between EM and RS, and this was regardless of missing pattern. However, the difference in coverage probability between the pairs MS/PW and EM/RS was larger under non-monotonic pattern than under monotonic pattern. All MDTs had coverage probabilities above 90% with normal data under monotonic missing pattern. MS and PW had poor coverage probabilities (below 90%); with the lowest coverage probability (below 65%) occurring at sample size 265 under non-monotonic pattern when percent missing was 20.

Table 9
Mean coverage probability (%) for β_1

			Monotonic Missing Pattern				Non-monotonic Missing Pattern			
N	Percent Missing	Normality (sk., kurt.)	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation
94	10	(0,0)	96.3	93.6	93.5	95.9	98.8	89.3	88.9	98.8
		(1,3)	95.5	93.2	93.2	94.9	98.7	87.7	87.6	98.6
		(1.8,6)	93.2	90.9	90.7	92.8	97.9	84.5	84.4	97.6
		(3,25)	87.8	84.3	83.8	87.4	94.8	78.0	78.2	95.1
	15	(0,0)	97.4	93.3	93.2	97.0	99.1	85.3	85.2	99.0
		(1,3)	96.0	92.0	91.8	95.0	99.1	83.7	83.5	99.1
		(1.8,6)	93.7	90.0	89.8	93.6	98.6	81.8	81.7	98.4
		(3,25)	89.1	83.8	83.9	88.9	95.3	75.0	74.4	95.2
	20	(0,0)	98.0	93.4	92.6	97.6	99.5	83.3	82.8	99.4
		(1,3)	96.7	91.9	91.0	96.2	99.0	80.6	80.5	99.0
		(1.8,6)	94.7	89.8	89.4	94.5	98.3	79.1	79.1	98.2
		(3,25)	89.4	84.4	83.4	89.3	95.0	73.2	73.0	94.6
153	10	(0,0)	97.0	94.7	94.6	97.0	99.3	85.1	84.8	99.2
		(1,3)	96.4	93.0	93.1	96.1	98.6	82.5	82.1	98.7
		(1.8,6)	93.8	90.0	89.9	93.6	97.8	79.8	79.3	97.7
		(3,25)	88.0	84.4	84.0	87.9	93.6	71.4	71.3	94.0
	15	(0,0)	97.3	94.5	94.1	97.4	99.8	80.2	80.1	99.8
		(1,3)	96.4	92.4	92.2	96.0	99.3	77.4	76.9	99.3
		(1.8,6)	94.3	89.1	89.2	94.0	98.4	74.9	74.8	98.7
		(3,25)	89.3	83.5	83.2	89.1	94.1	68.0	68.4	94.2
	20	(0,0)	98.0	93.5	92.9	97.8	99.9	76.7	75.9	99.9
		(1,3)	97.3	91.4	91.4	97.1	99.3	74.5	73.5	99.4
		(1.8,6)	95.6	88.5	88.1	95.6	98.6	72.1	72.6	98.7
		(3,25)	90.7	83.5	82.8	91.3	95.4	66.7	66.6	95.2
265	10	(0,0)	96.9	94.1	93.9	96.6	98.8	79.2	78.9	98.8
		(1,3)	95.7	93.2	92.9	95.6	97.9	76.3	75.6	98.1
		(1.8,6)	93.1	90.2	90.0	93.1	97.1	74.3	74.0	97.4
		(3,25)	85.0	82.6	82.2	85.2	90.4	64.4	63.8	91.0
	15	(0,0)	97.3	93.1	92.7	97.4	98.9	72.2	71.5	99.2
		(1,3)	96.6	92.7	92.2	96.2	98.1	69.6	69.2	98.6
		(1.8,6)	94.6	90.0	89.2	94.6	97.4	66.8	65.7	97.8
		(3,25)	87.2	82.5	81.6	87.1	91.6	59.4	59.0	91.7
	20	(0,0)	98.1	91.3	90.6	98.1	99.4	64.3	63.9	99.6
		(1,3)	97.3	90.9	90.3	97.1	98.3	63.7	63.0	98.7
		(1.8,6)	95.4	87.7	97.0	95.2	97.8	60.9	60.4	98.2
		(3,25)	88.5	81.7	81.0	88.8	92.6	56.5	56.2	92.9

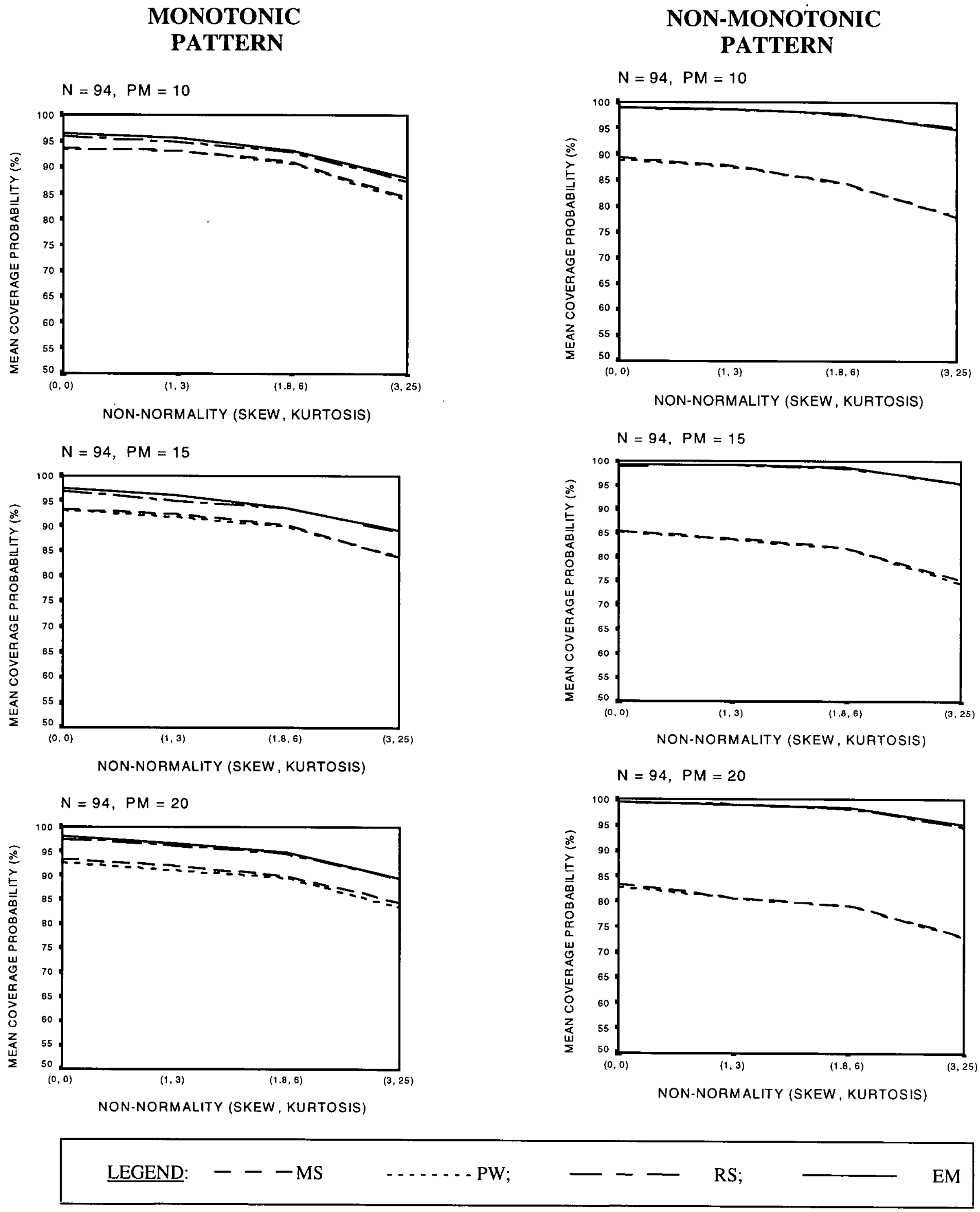


Figure 10: Mean coverage probability for β_1 (N = 94)

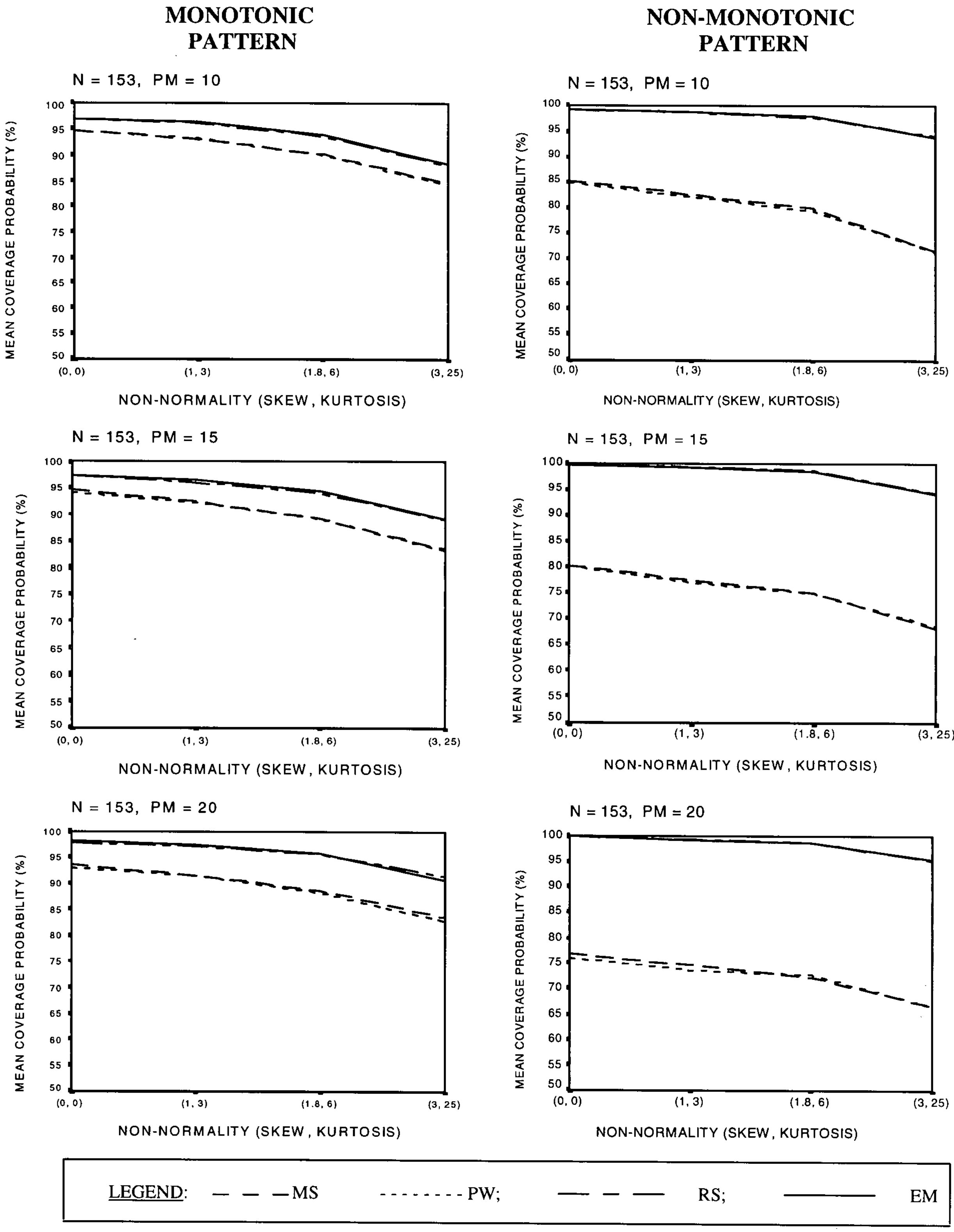


Figure 11: Mean coverage probability for β_1 (N=153)

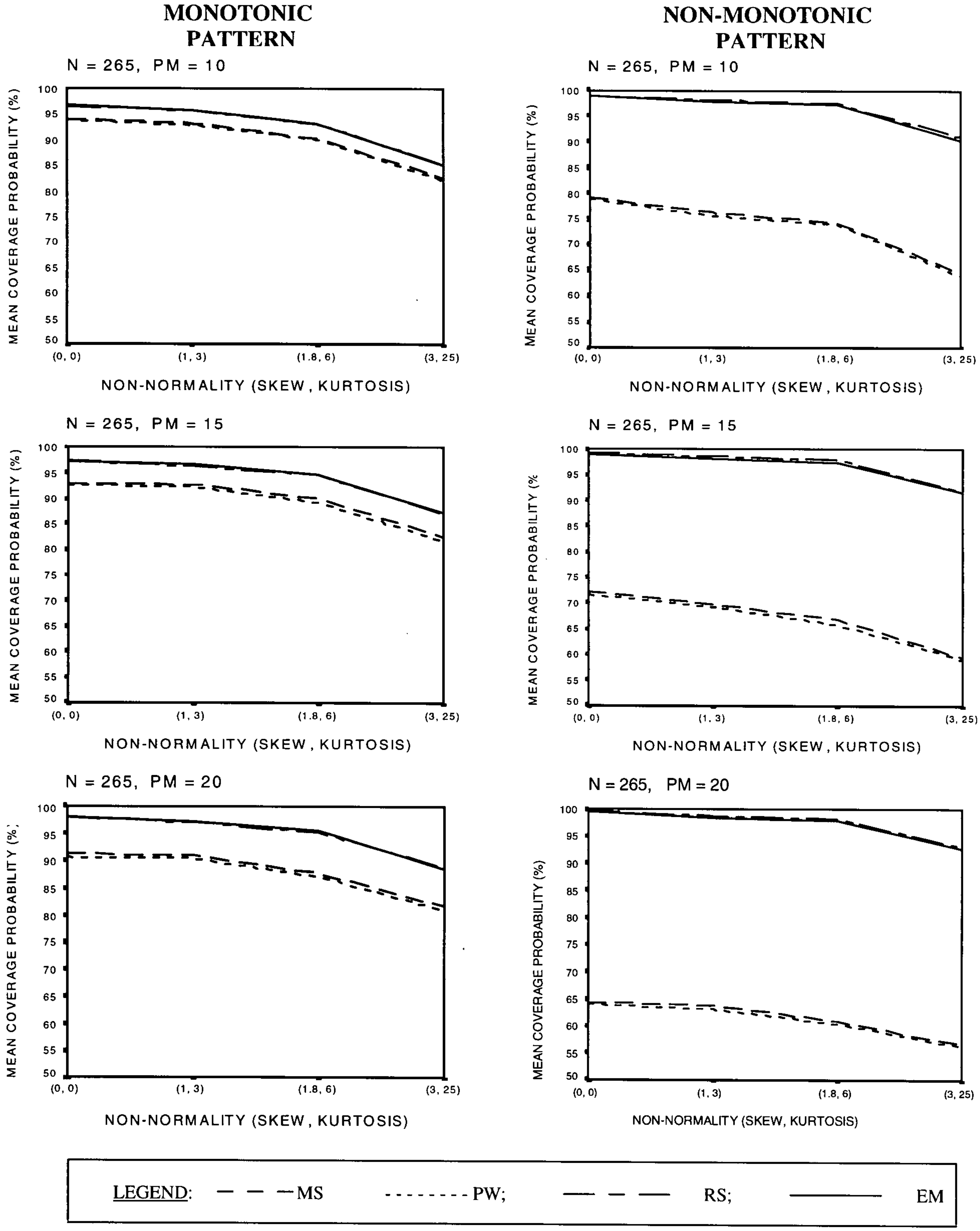


Figure 12: Mean coverage probability for β_1 (N = 265)

Summary

In the foregoing sections, effects on absolute error for $R^2_{estimate}$ and β_1 were reported, together with the relative performance of MDTs. In addition, MDTs were compared with respect to the coverage probabilities of population regression coefficients when nominal alpha is .05.

There were no effects on absolute error for b_1 of practical significance at sample size 94 or 153. However, at sample size 265, the effect of percent missing was of practical significance under EM and RS only.

With respect to mean absolute error for $R^2_{estimate}$, the performance of all MDTs deteriorated with increasing non-normality. The best performance was under EM, the worst under PW. The smallest bias in $R^2_{estimate}$ across different conditions of the design variables was not under a single MDT. There was more differentiation in the performance of MDTs with non-monotonic missing pattern than with monotonic missing pattern. In each case, MDTs differed more at higher levels of non-normality than at lower levels, except at the highest level of sample size with 20% missing.

The effect of missing pattern on absolute error for b_1 was of practical significance under MS and PW, but only when sample size was moderate or large (153 or 265). Non-normality showed an effect of practical significance on absolute error for b_1 under EM and RS when sample size was moderate to large (153 to 265). The strongest effect on absolute error for b_1 was that of missing pattern under both MS and PW when sample size was 265. It appears that with respect to effects on absolute error for b_1, MS and PW were alike in their performance. Also, EM and RS were alike in that the absolute error for b_1 under the two MDTs was affected by non-normality.

As for the relative performance of MDTs in estimating population regression coefficients, MDTs differed mainly on the coefficient of the predictor with no missing data (b_1 in this case). Therefore, only the findings regarding b_1 are reported. For b_1, the performance of MDTs deteriorated with increasing non-normality. This was regardless of missing pattern. Overall, RS and EM outperformed MS and RS in estimating β_1 with respect to accuracy of parameter estimates. The smallest bias in b_1 was also under RS. Whereas MS and PW consistently overestimated β_1, EM and RS consistently underestimated β_1. The performance of MS was undifferentiated from that of PW across the two patterns. However, EM and RS were more differentiated from MS and PW under non-monotonic than under monotonic missing pattern.

Coverage probabilities decreased with increasing non-normality. Both EM and RS provided coverage probabilities that were closer to the nominal value of 95% than did MS and PW.

In sum, the findings from this study suggest that with fewer predictors (four in this case) under low criterion-predictor relationships, EM and RS generally were superior to MS and PW with respect accuracy in parameter estimation. However, the smallest bias was under RS. How the outcomes will change when R^2 is high was examined in the next study.

Study 2: Four Predictors under High R^2 Condition

Effects on Absolute Error for $R^2_{estimate}$

This study was identical to Study 1 in design and analysis, except that the data were generated using high multiple R^2 (0.59437). Table 10 contains the effects of design variables on absolute error for $R^2_{estimate}$ that fulfilled the criterion of $\eta^2_{alt} \geq .03$, showing that at sample size 94, the strongest effect was that of missing pattern under EM. There was no effect of practical significance under MS at sample size 94. Percent missing affected the performance of EM and RS more than it affected MS and PW. On the other hand, non-normality affected the performance of MS and RS more than it affected EM and RS.

At sample size 153, there was no effect of practical significance under MS. Under PW, the effect of non-normality was of practical significance. Under EM and RS, the effects of pattern and non-normality on absolute error were of practical significance. However, their interaction effect was also significant, and thus their main effects have to be qualified. These ordinal interaction effects are displayed in Figures 13 and 14, showing that absolute error were generally smaller under monotonic missing pattern than under non-monotonic missing pattern. However, the difference in absolute error with normal data was smaller than the difference in absolute error at the highest level of non-normality (skew=3 and kurtosis=25).

At sample size 265, the effect of percent missing on absolute error for $R^2_{estimate}$ under MS was of practical significance. Under PW, the effect of non-normality was of practical significance. Under EM and RS, the effects of pattern and non-normality on absolute error for $R^2_{estimate}$ were of practical significance. However, their interaction effect was also significant, and thus their main effects have to be qualified as was the case at sample size 153. These ordinal interaction effects are displayed in Figures 15 and 16, showing that absolute errors were generally smaller under monotonic pattern than under non-monotonic pattern. However, the difference in absolute error across missing pattern was smaller for normal data than for non-normal data.

Relative Performance of MDTs in the Estimation of Population R^2

Table 11 contains means and standard deviations of absolute error in $R^2_{estimate}$ under various design conditions. The means are plotted in Figures 17, 18, and 19 for sample sizes 94, 153 and 265, respectively. The figures show that the performance of all MDTs deteriorated with increasing non-normality as measured by the absolute bias in $R^2_{estimate}$ estimate. Overall, PW had the best performance and EM the worst.

MDTs were more differentiated at the highest level of percent missing (20%). Also, at each level of percent missing, MDTs were most differentiated for normal data, and least differentiated at the highest level of non-normality (skew = 3 and kurtosis = 25).

Table A.3 in Appendix A contains bias in $R^2_{estimate}$ under high R^2 condition with four predictors. Bias in $R^2_{estimate}$ varied from –.25 to .04 across all sample size conditions, giving a range of .29. This suggests a substantial difference in bias among MDTs in the estimation of population R^2. Figure B.3 in Appendix B is a set of graphical plots for the bias in $R^2_{estimate}$ (N = 94), showing that all the MDTs, except PW, tended to underestimate R^2. The smallest bias was under PW at all levels of the design variables.

Table 10

Effects on absolute error for $R^2_{estimate}$ fulfilling the $\eta^2_{alt} \geq .03$ criterion

Sample size	MDT	Effect	SS_{effect}	$SS_{residual}$	df	η^2_{alt}
94	EM	Percent missing	3.60	74.87	2	.046
		Pattern	27.08	74.87	1	.266
	MS	-	-	-	-	-
	PW	Non-normality	2.78	57.82	3	.046
	RS	Pattern	21.43	74.51	1	.223
		Percent missing	3.46	74.51	2	.044
		Non-normality	3.87	74.51	3	.049
153	EM	Pattern	35.68	51.73	1	.408
		Percent missing	3.48	51.73	2	.063
		Non-normality	2.25	51.73	3	.042
		Pattern×Non-normality	1.73	51.73	3	.032
	MS	-	-	-	-	-
	PW	Non-normality	1.62	25.21	3	.061
	RS	Pattern	30.48	52.03	1	.369
		Percent missing	3.65	52.03	2	.066
		Non-normality	4.07	52.03	3	.073
		Pattern×Non-normality	2.44	52.03	3	.045
265	EM	Pattern	41.99	34.06	1	.552
		Percent missing	3.94	34.06	2	.104
		Non-normality	2.03	34.06	3	.056
		Pattern×Non-normality	2.94	34.06	3	.079
	MS	Percent missing	1.54	39.38	2	.038
	PW	Non-normality	.97	15.13	3	.060
	RS	Pattern	37.32	34.68	1	.518
		Percent missing	3.96	34.68	2	.102
		Non-normality	3.66	34.68	3	.095
		Pattern×Non-normality	3.97	34.68	3	.103

Note: EM = EM imputation, MS = Mean substitution, PW = Pairwise deletion, RS = Regression imputation, - = No effect of practical significance.

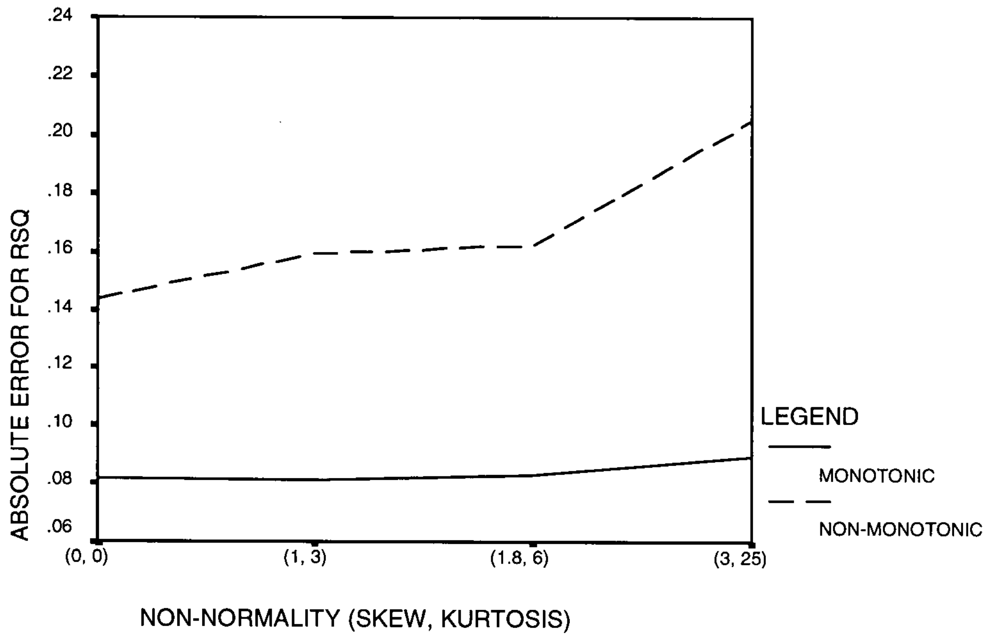


Figure 13: Interaction effect of pattern and non-normality on absolute error for $R^2_{estimate}$ under EM (N = 153)
Note: RSQ = $R^2_{estimate}$.

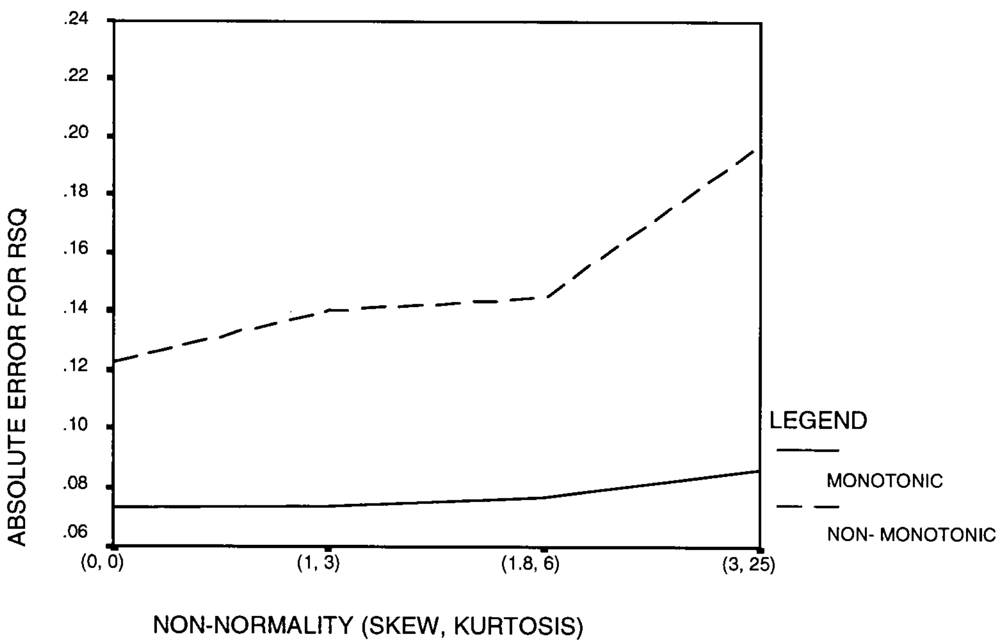


Figure 14: Interaction effect of pattern and non-normality on absolute error in $R^2_{estimate}$ under RS (N = 153)
Note: RSQ = $R^2_{estimate}$.

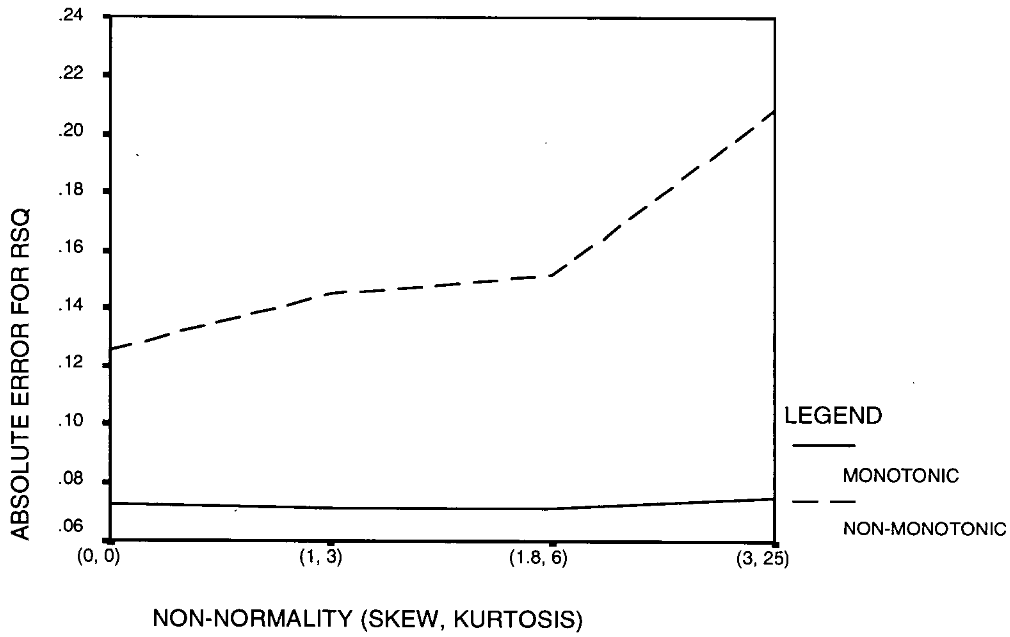


Figure 15 : Interaction effect of pattern and non-normality on absolute error for $R^2_{estimate}$ under EM (N = 265)
Note: RSQ = $R^2_{estimate}$.

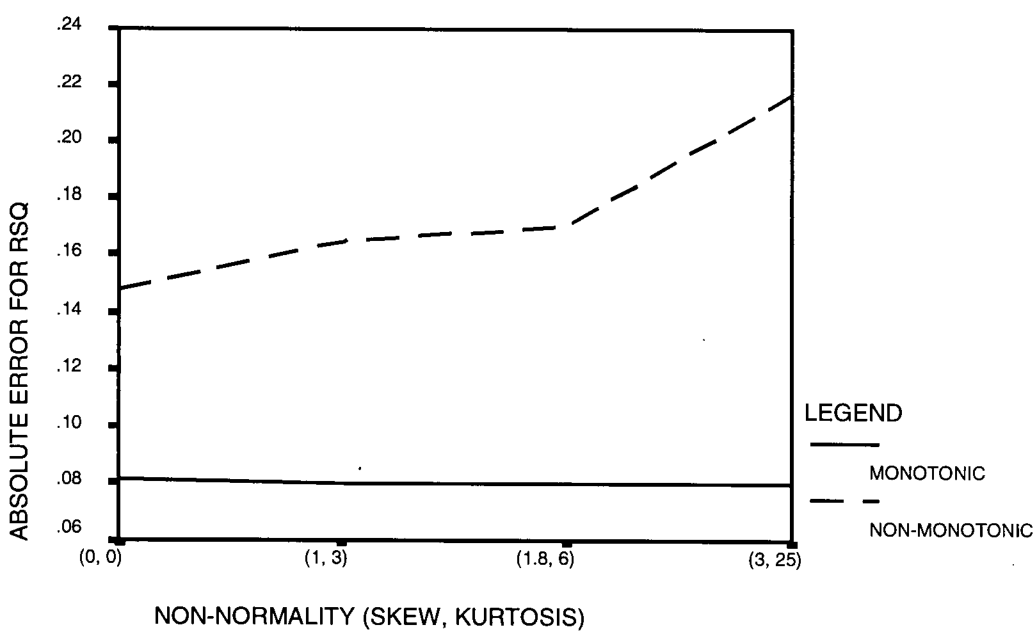


Figure 16: Interaction effect of pattern and non-normality on absolute error for $R^2_{estimate}$ under RS (N = 265)
Note: RSQ = $R^2_{estimate}$.

Table 11
Means and standard deviations* of absolute error for $R^2_{estimate}$

			Monotonic Missing Pattern				Non-monotonic Missing Pattern			
N	Percent Missing	Normality (sk., kurt.)	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation
94	10	(0,0)	0630(049)	0608(047)	0538(040)	0615(047)	0967(055)	0620(048)	0515(039)	0788(054)
		(1,3)	0666(051)	0644(050)	0585(044)	0654(047)	1091(059)	0648(051)	0548(043)	0900(059)
		(1.8,6)	0721(056)	0707(054)	0674(051)	0714(055)	1108(064)	0691(055)	0624(049)	0941(063)
		(3,25)	0884(067)	0902(067)	0913(083)	0898(068)	1511(087)	0878(069)	0796(076)	1400(087)
	15	(0,0)	0813(055)	0720(053)	0535(040)	0752(054)	1388(054)	0729(054)	0510(039)	1167(057)
		(1,3)	0826(058)	0745(054)	0582(044)	0777(056)	1532(058)	0771(058)	0548(043)	1338(060)
		(1.8,6)	0851(062)	0785(058)	0668(051)	0813(060)	1541(066)	0801(062)	0621(049)	1370(067)
		(3,25)	0952(074)	0933(069)	0900(082)	0948(073)	1935(092)	1043(080)	0808(083)	1867(093)
	20	(0,0)	1063(061)	0880(059)	0533(040)	0957(060)	1797(052)	0870(060)	0517(040)	1597(055)
		(1,3)	1060(066)	0897(061)	0582(044)	0969(064)	1924(058)	0923(065)	0559(045)	1765(059)
		(1.8,6)	1069(071)	0922(066)	0671(053)	0990(069)	1930(067)	0961(069)	0629(052)	1788(068)
		(3,25)	1113(085)	1020(077)	0927(115)	1075(082)	2281(095)	1234(088)	0831(080)	2245(097)
153	10	(0,0)	0568(043)	0523(042(	0411(031)	0535(042)	1019(044)	0550(041)	0393(030)	0803(046)
		(1,3)	0581(044)	0542(043)	0436(033)	0554(043)	1177(047)	0585(043)	0408(032)	0959(048)
		(1.8,6)	0619(047)	0591(045)	0502(038)	0600(046)	1208(052)	0627(046)	0461(037)	1005(052)
		(3,25)	0744(056)	0741(056)	0699(052)	0747(056)	1642(075)	0834(063)	0616(050)	1515(075)
	15	(0,0)	0799(049)	0671(048)	0409(031)	0713(049)	1467(042)	0696(047)	0393(031)	1230(045)
		(1,3)	0794(052)	0680(049)	0433(033)	0720(050)	1625(046)	0745(050)	0410(033)	1417(048)
		(1.8,6)	0810(054)	0712(051)	0498(038)	0748(053)	1656(052)	0790(053)	0462(037)	1465(054)
		(3,25)	0873(064)	0817(061)	0691(051)	0843(062)	2091(077)	1059(073)	0618(052)	2012(077)
	20	(0,0)	1074(052)	0867(052)	0408(031)	0953(053)	1837(040)	0862(052)	0402(032)	1643(043)
		(1,3)	1056(056)	0855(055)	0430(033)	0940(057)	1982(045)	0927(054)	0420(034)	1830(046)
		(1.8,6)	1046(061)	0868(058)	0492(038)	0947(060)	2004(051)	0964(058)	0471(038)	1870(052)
		(3,25)	1065(072)	0931(067)	0689(053)	1001(070)	2410(079)	1267(081)	0637(056)	2382(080)
265	10	(0,0)	0541(036)	0479(035)	0307(024)	0495(035)	1067(033)	0516(034)	0299(024)	0832(036)
		(1,3)	0536(037)	0477(036)	0328(026)	0493(037)	1243(036)	0560(038)	0314(026)	1012(038)
		(1.8,6)	0546(040)	0496(039)	0378(029)	0509(040)	1297(041)	0599(042)	0354(029)	1078(043)
		(3,25)	0612(048)	0594(046)	0533(040)	0602(047)	1783(063)	0848(060)	0484(039)	1648(063)
	15	(0,0)	0807(039)	0670(039)	0306(024)	0715(040)	1490(031)	0683(039)	0308(025)	1256(034)
		(1,3)	0792(042)	0659(042)	0326(026)	0704(042)	1665(034)	0752(042)	0325(026)	1462(036)
		(1.8,6)	0782(047)	0659(045)	0375(029)	0702(046)	1719(040)	0797(046)	0363(029)	1531(041)
		(3,25)	0772(059)	0700(054)	0526(040)	0728(056)	2200(061)	1097(066)	0491(040)	2125(062)
	20	(0,0)	1096(040)	0880(042)	0305(024)	0969(041)	1879(030)	0868(042)	0325(026)	1678(032)
		(1,3)	1076(045)	0859(046)	0324(026)	0951(046)	2037(033)	0954(044)	0343(028)	1879(034)
		(1.8,6)	1056(051)	0847(051)	0372(029)	0936(051)	2082(039)	1001(049)	0376(030)	1941(040)
		(3,25)	1003(066)	0839(061)	0519(040)	0913(064)	2509(060)	1336(071)	0505(042)	2474(061)

* All values in the table are preceded by an omitted decimal point. Standard deviations are in parentheses. Standard errors of the means can be obtained by dividing each standard deviation by 1000½.

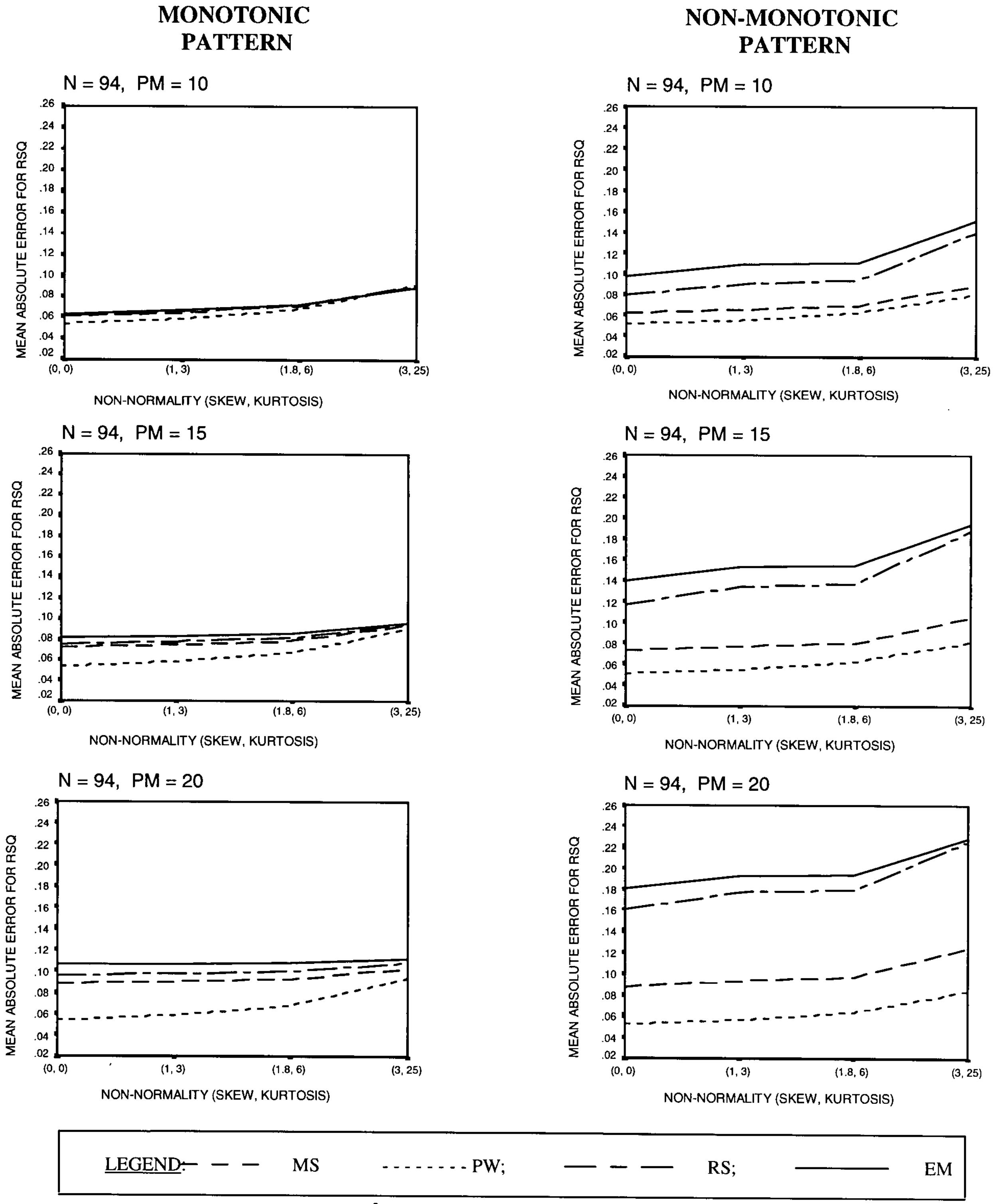


Figure 17: Mean absolute error for $R^2_{estimate}$ across levels of non-normality (N = 94)
Note: N=Sample size, PM=Percent missing, RSQ = $R^2_{estimate}$

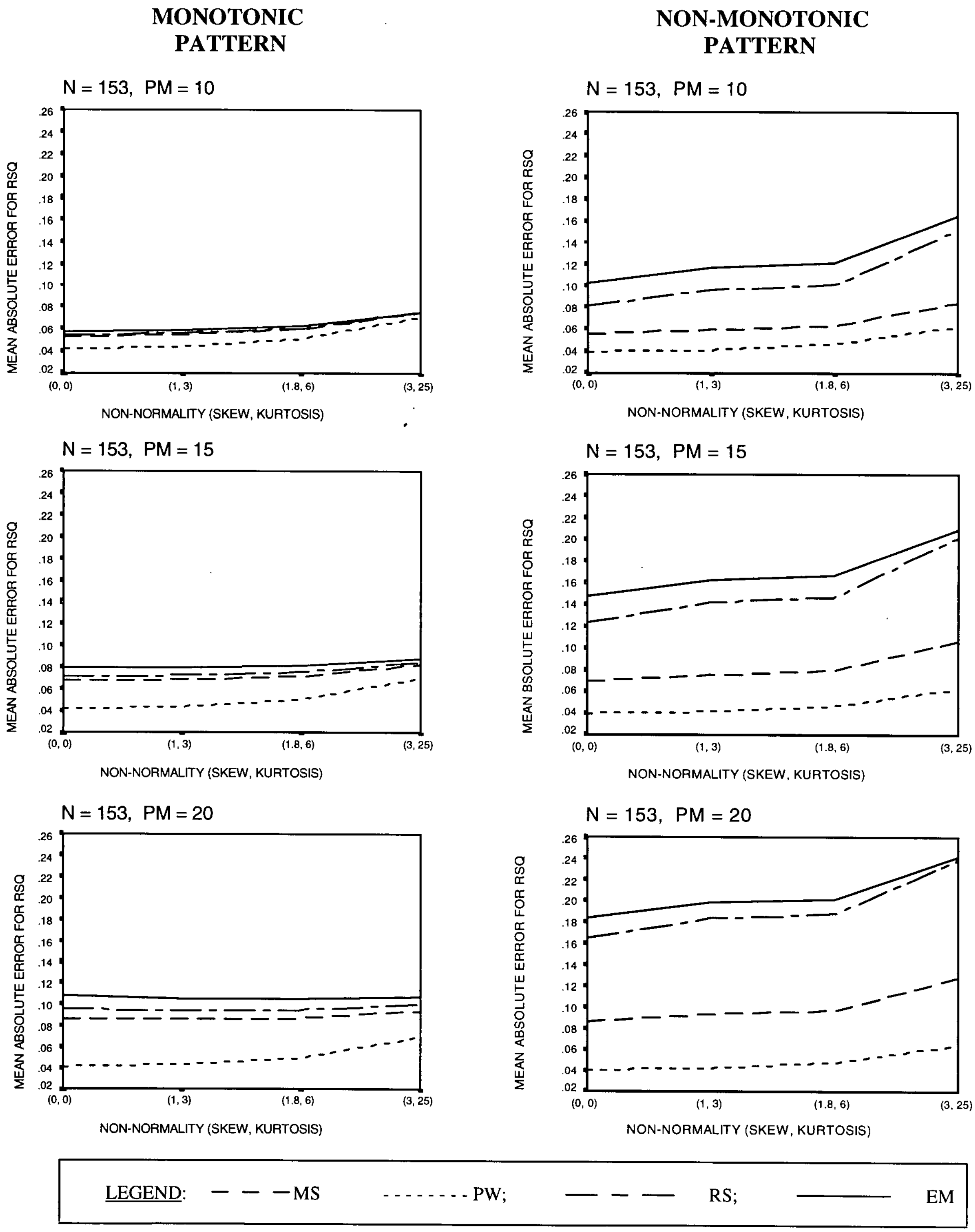


Figure 18: Mean absolute error for $R^2_{estimate}$ across levels of non-normality (N = 153)
Note: N=Sample size, PM=Percent missing, RSQ = $R^2_{estimate}$

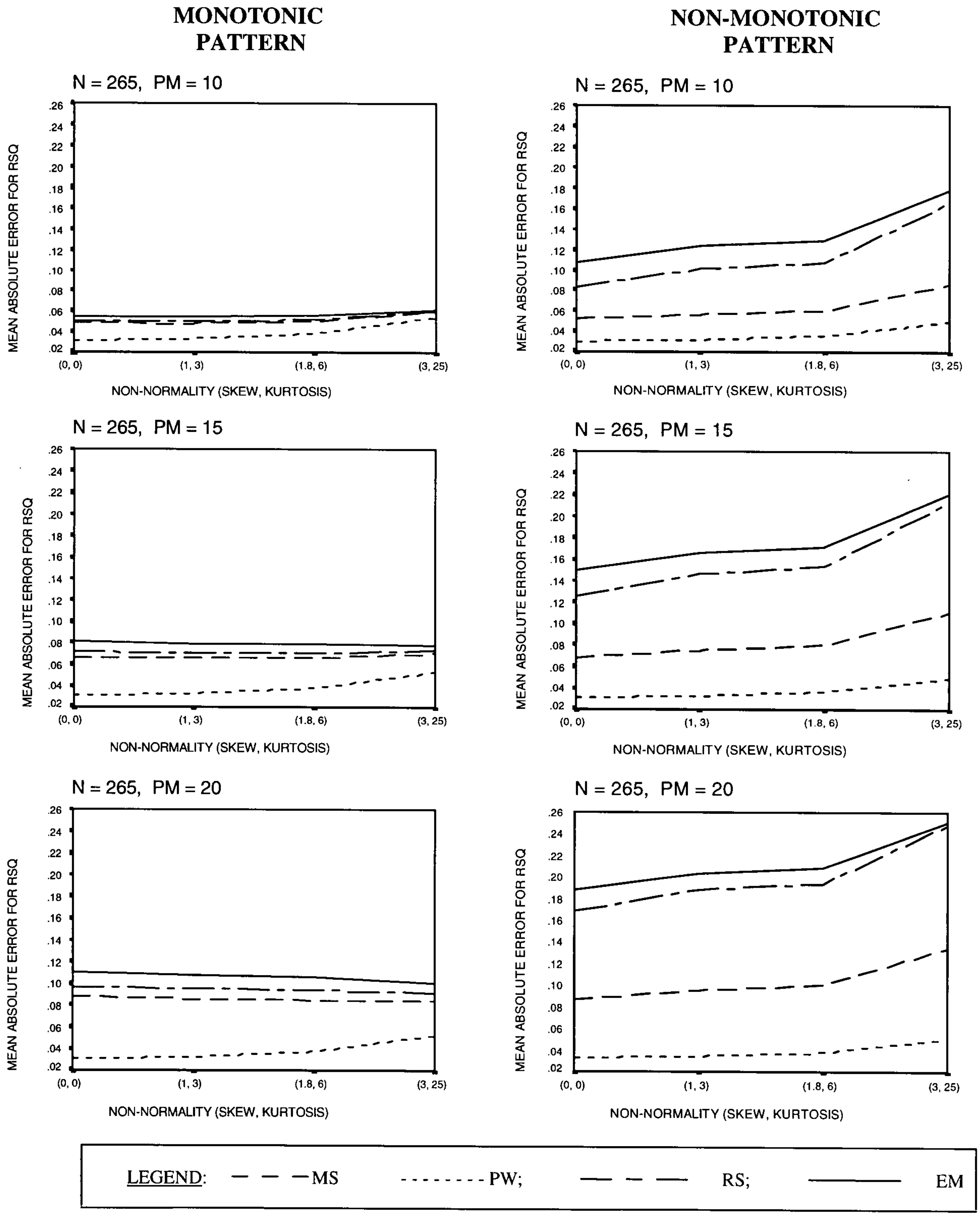


Figure 19: Mean absolute error for $R^2_{estimate}$ across levels of non-normality (N = 265)
Note: N=Sample size, PM=Percent missing, RSQ = $R^2_{estimate}$

Effects on Absolute Error for Regression Coefficients

The effects of design variables on absolute error for regression coefficients, which fulfilled the criterion $\eta^2_{alt} \geq .03$, are in Table 12. Percent missing had no effect of practical value. Missing pattern had an effect of practical significance only for bias in b_1, the intensity of which increased with sample size. This was true for all MDTs. In addition, the effect of non-normality on absolute error for all regression coefficients was of practical significance under all MDTs, except EM. The strength of this effect also increased with sample size. Notice that for absolute error for b_1 under EM, it was only the effect of missing pattern that met the criterion $\eta^2_{alt} \geq .03$.

Table 12
Effects on absolute error for regression coefficients fulfilling the criterion of $\eta^2_{alt} \geq .03$*

Sample size	MDT	b_1	b_2	b_3	b_4
94	EM	PT (157)	NL (050)	NL (074)	NL (033)
	MS	PT (082)	NL (055)	NL (081)	NL (036)
		NL (043)	-	-	-
	PW	PT (082)	NL(056)	NL (080)	NL (037)
		NL (043)	-	-	-
	RS	PT (162)	NL (052)	NL (075)	NL (039)
		NL (050)	-	-	-
153	EM	PT (278)	NL (056)	NL (085)	NL (040)
	MS	PT (164)	NL (057)	NL (096)	NL (049)
		NL (043)	-	-	-
	PW	PT (163)	NL (057)	NL (096)	NL (049)
		NL (042)	-	-	-
	RS	PT (302)	NL (060)	NL (083)	NL (051)
		NL (055)	-	-	-
265	EM	PT (419)	NL (069)	NL (107)	NL (037)
	MS	PT (292)	NL (069)	NL (122)	NL (.061)
		NL (045)			
	PW	PT (291)	NL (068)	NL (123)	NL (061)
		NL (045)	-	-	-
	RS	PT (462)	NL (073)	NL (113)	NL (059)
		NL (042)			

* Values of η^2_{alt} are in parentheses preceded by an omitted decimal point.
Note: EM = EM imputation, MS = Mean substitution, PW = Pairwise deletion, RS = Regression imputation, PT = Pattern, NL = Non-normality, - = No effect of practical significance.

Relative Performance of MDTs in the Estimation of Regression Coefficients

Table 13 contains cell means and standard deviations of absolute error for b_1. The influence of non-normality on the performance of MDTs at sample size 94 is revealed in Figure 20. The figure shows that in general, the mean absolute error was smaller under monotonic pattern than under non-monotonic pattern, and that MDTs were more differentiated as percent missing increased. However, under both missing patterns, the MS and PW methods were undifferentiated.

Overall, RS performed the best under both monotonic and non-monotonic missing patterns, and EM performed the worst. However, at 15% and 20% missing under non-monotonic pattern, EM performed better than MS and PW at the highest level of non-normality (skew = 3 and kurtosis = 25).

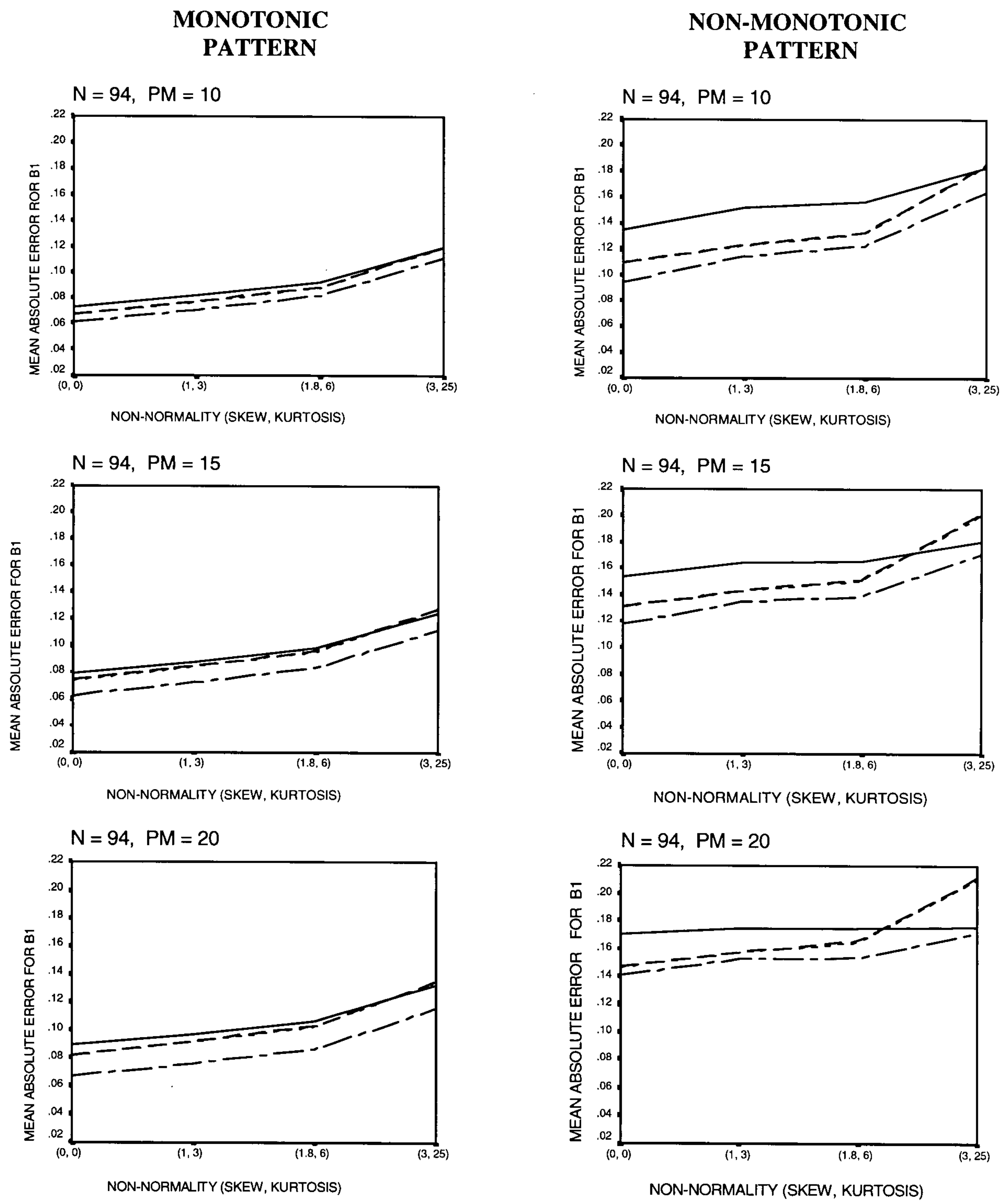
Figure 21 shows mean absolute error for b_1 at sample size 153. Under monotonic missing pattern, RS performed the best and EM the worst. This was true with non-monotonic pattern as well, except that at 15% and 20% missing under non-monotonic pattern, EM performed better than both MS and PW at the highest level of non-normality (skew = 3 and kurtosis = 25). Also, at 15% and 20% missing under non-monotonic pattern, the performance of RS, MS and PW was undifferentiated except at the highest level of non-normality where RS outperformed the rest. Figure 22 is a set of graphs for mean absolute error for b_1 at sample size 265. The figure shows that results at this level of sample size were similar to those under sample size 153.

Table A.4 in Appendix A shows the bias in b_1 under high R^2 with four predictors (N = 94). Figure B.4 in Appendix B shows typical graphs for the bias in b_1. The bias in b_1 varied from –.177 to .195, giving a range of .37. The smallest bias was under RS across all levels of the design variables. However, the bias under EM was very close to that under RS. Whereas MS and PW consistently overestimated β_1, EM and RS consistently underestimated β_1. The range in bias for b_2 to b_4, with a maximum of .08, was far smaller than that for b_1. Therefore, their graphical plots were omitted.

Table 13
Means and standard deviations* of absolute error for b_1

			Monotonic Missing Pattern				Non-monotonic Missing Pattern			
N	Percent Missing	Normality (sk., kurt.)	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation
94	10	(0,0)	0721(053)	0670(051)	0668(051)	0604(046)	1346(067)	1098(072)	1097(072)	0944(054)
		(1,3)	0814(060)	0770(063)	0767(062)	0704(053)	1514(076)	1232(091)	1230(091)	1146(065)
		(1.8,6)	0924(068)	0885(075)	0883(074)	0819(061)	1564(085)	1329(107)	1325(106)	1226(071)
		(3,25)	1198(091)	1197(115)	1195(115)	1108(084)	1825(104)	1845(171)	1840(171)	1641(098)
	15	(0,0)	0790(056)	0742(055)	0739(055)	0625(047)	1533(067)	1311(077)	1310(077)	1175(058)
		(1,3)	0873(063)	0844(067)	0842(067)	0722(054)	1643(076)	1429(096)	1427(096)	1346(069)
		(1.8,6)	0976(071)	0957(079)	0954(079)	0830(062)	1652(084)	1512(113)	1506(112)	1389(076)
		(3,25)	1242(092)	1277(123)	1274(123)	1117(085)	1800(102)	2008(177)	2000(176)	1705(102)
	20	(0,0)	0890(062)	0820(061)	0815(061)	0669(050)	1696(069)	1468(082)	1466(082)	1410(062)
		(1,3)	0964(067)	0918(073)	0914(072)	0759(056)	1745(077)	1573(101)	1570(100)	1522(072)
		(1.8,6)	1059(074)	1026(084)	1023(084)	0860(063)	1741(085)	1652(117)	1642(116)	1533(079)
		(3,25)	1317(096)	1343(129)	1340(129)	1149(086)	1759(101)	2112(180)	2100(180)	1709(102)
153	10	(0,0)	0595(043)	0525(038)	0524(038)	0463(035)	1366(052)	1039(059)	1038(059)	0922(045)
		(1,3)	0664(047)	0587(046)	0586(046)	0543(040)	1523(060)	1158(075)	1156(075)	1123(053)
		(1.8,6)	0758(054)	0677(054)	0677(054)	0642(046)	1548(069)	1230(089)	1223(089)	1177(061)
		(3,25)	1021(073)	0954(084)	0954(083)	0914(067)	1776(089)	1673(142)	1667(141)	1581(083)
	15	(0,0)	0703(048)	0597(043)	0595(043)	0493(037)	1582(053)	1269(064)	1267(064)	1184(047)
		(1,3)	0764(053)	0665(052)	0662(052)	0573(042)	1679(060)	1374(079)	1371(079)	1349(055)
		(1.8,6)	0947(059)	0747(061)	0745(061)	0669(048)	1673(069)	1436(094)	1427(094)	1368(063)
		(3,25)	1082(078)	1024(093)	1022(093)	0935(068)	1738(087)	1845(145)	1834(144)	1628(086)
	20	(0,0)	0811(049)	0672(047)	0668(047)	0553(040)	1695(051)	1423(066)	1422(066)	1413(047)
		(1,3)	0858(055)	0728(056)	0724(056)	0621(045)	1743(058)	1514(081)	1511(081)	1529(056)
		(1.8,6)	0922(061)	0801(065)	0798(065)	0703(050)	1718(067)	1566(096)	1554(096)	1521(065)
		(3,25)	1134(080)	1062(097)	1061(097)	0956(068)	1699(085)	1940(147)	1926(147)	1665(087)
265	10	(0,0)	0537(036)	0424(032)	0424(032)	0383(029)	1378(040)	1023(047)	1023(047)	0903(036)
		(1,3)	0578(040)	0482(036)	0482(036)	0441(033)	1527(047)	1129(059)	1127(059)	1102(043)
		(1.8,6)	0639(045)	0548(042)	0548(042)	0511(038)	1541(056)	1179(070)	1174(070)	1141(051)
		(3,25)	0831(059)	0768(060)	0762(060)	0718(052)	1717(075)	1562(109)	1556(108)	1507(071)
	15	(0,0)	0656(039)	0494(036)	0493(036)	0427(031)	1583(039)	1243(049)	1243(049)	1187(035)
		(1,3)	0682(043)	0548(041)	0548(040)	0473(035)	1672(046)	1333(061)	1331(061)	1346(043)
		(1.8,6)	0726(048)	0609(046)	0609(046)	0534(039)	1664(054)	1374(073)	1366(073)	1360(051)
		(3,25)	0883(062)	0813(066)	0813(066)	0724(053)	1686(075)	1725(111)	1715(111)	1577(073)
	20	(0,0)	0783(042)	0578(039)	0576(039)	0502(033)	1729(038)	1411(050)	1411(050)	1429(035)
		(1,3)	0803(046)	0617(045)	0616(045)	0534(037)	1765(044)	1489(062)	1487(062)	1535(042)
		(1.8,6)	0837(051)	0672(050)	0672(050)	0587(041)	1734(053)	1522(074)	1513(073)	1520(050)
		(3,25)	0970(064)	0862(070)	0862(070)	0754(054)	1637(074)	1829(112)	1818(112)	1585(074)

* All values in the table are preceded by an omitted decimal point. Standard deviations are in parentheses. Standard errors of the means can be obtained by dividing each standard deviation by $1000^{½}$.



LEGEND: — — —MS - - - - - - - - PW; ——— — ——— RS; ———————— EM

Figure 20: Mean absolute error for b_1 across levels of non-normality (N = 94)
Note: N = Sample size, PM = Percent missing.

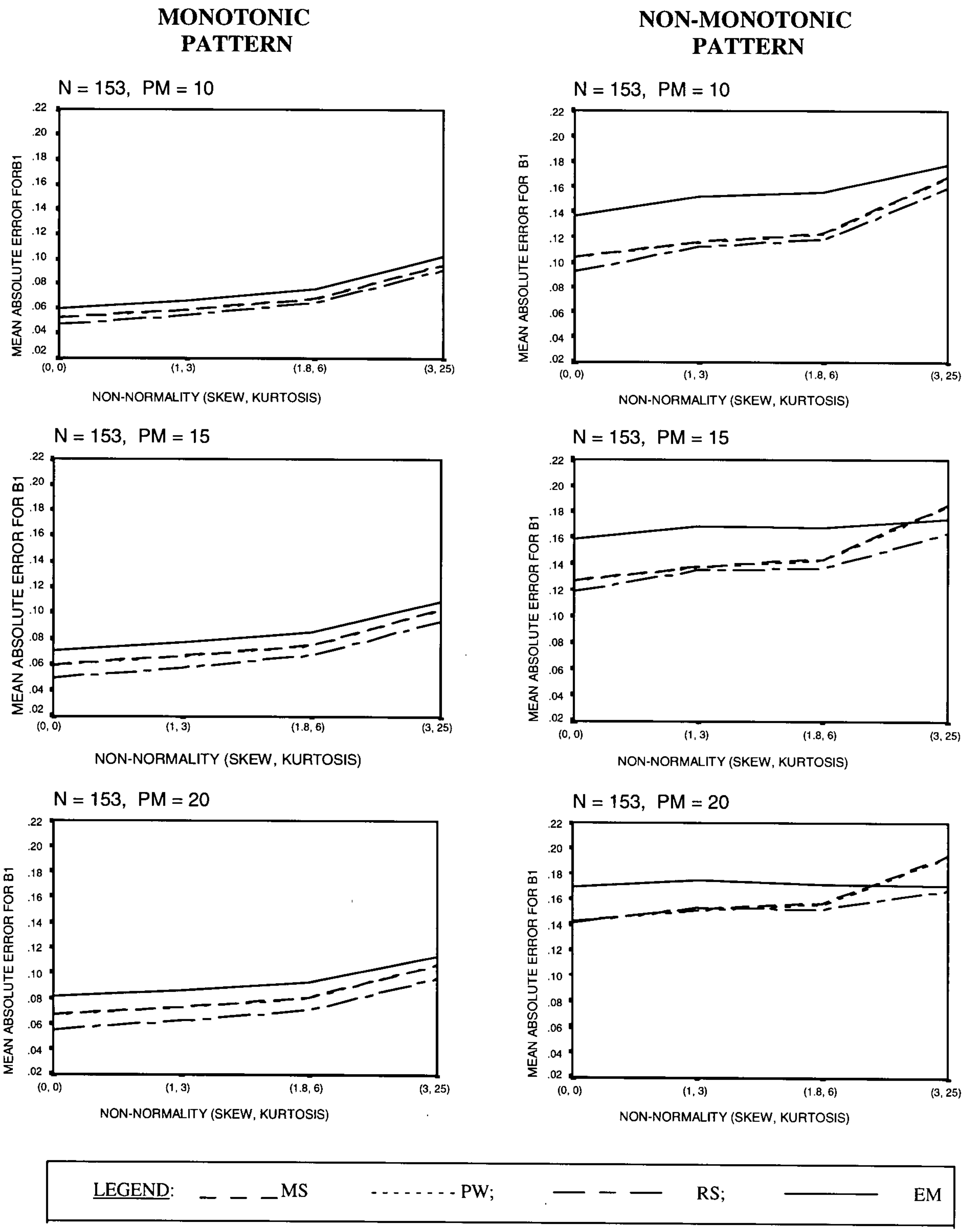


Figure 21: Mean absolute error for b_1 across levels of non-normality (N = 153)
Note: N = Sample size, PM = Percent missing.

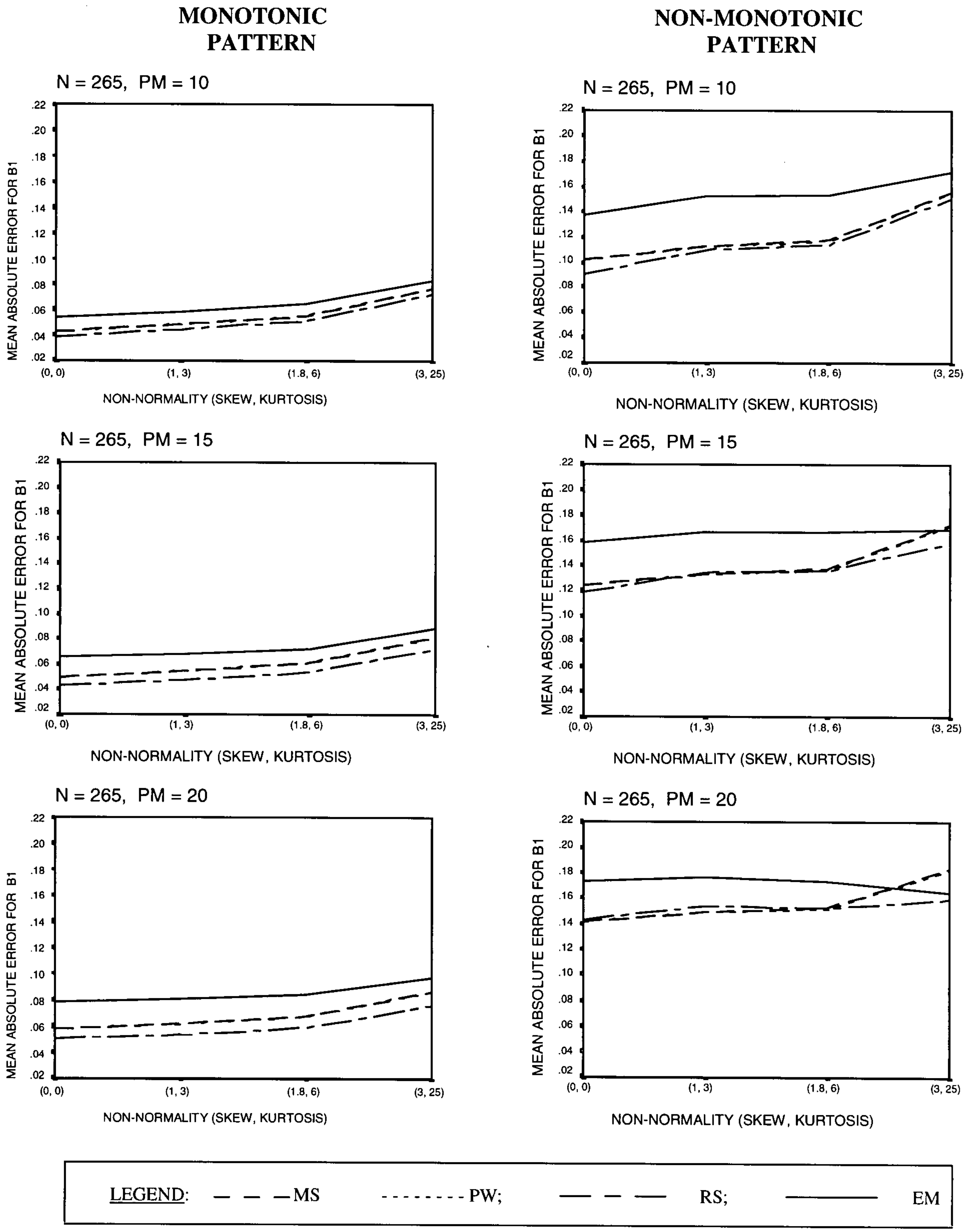


Figure 22: Mean absolute error for b_1 across levels of non-normality (N = 265)
Note: N = Sample size, PM = Percent missing.

Coverage Probability for β_1

MDTs mostly differed in coverage probabilities for β_1 and not for β_2, β_3,...,β_9. For this reason, only the results for β_1 are reported.

Table 14 contains mean coverage probabilities for β_1 at different levels of sample size. The means at sample size 94 are displayed in Figure 23. The figure shows that coverage probability decreased with increasing non-normality for all MDTs under both monotonic and non-monotonic patterns of missing data. RS provided the best coverage probability under both monotonic and non-monotonic missing patterns, followed by EM, MS and PW, in that order. However, under monotonic missing pattern, MDTs differed most in coverage probability of β_1 at 20% missing, and least at 10% missing.

All MDTs performed poorly under non-monotonic missing pattern. Under this pattern, coverage probabilities ranged from 41.8% at the highest level of percent missing (20%) and highest level of non-normality (skew = 3 and kurtosis=25) to 86.4% at the lowest level of percent missing (10%) and normal data. Thus, with non-monotonic missing pattern, the only result close to the nominal value was 86.4%, and this was under RS for normal data at 10% missing.

Figure 24 shows coverage probabilities at sample size 153. RS provided the best coverage probability under both monotonic and non-monotonic patterns. However, all MDTs had poor coverage probabilities with non-monotonic missing pattern. At 10% missing, MDTs had the most differentiated coverage probabilities for normal data, and the least differentiated coverage probabilities at the highest level of non-normality (skew = 3 and kurtosis = 25).

Figure 25 shows coverage probabilities at sample size 265. Under monotonic missing pattern, RS outperformed the other MDTs with respect to coverage probabilities, followed by MS, PW and EM, in that order. Again, all MDTs had poor coverage probabilities under non-monotonic pattern, ranging from 9.6% under EM to 58.1% under RS.

Table 14
Mean coverage probability (%) for β_1

			Monotonic Missing Pattern				Non-monotonic Missing Pattern			
N	Percent Missing	Normality (sk., kurt.)	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation
94	10	(0,0)	92.2	92.3	91.1	94.9	71.8	69.2	66.5	86.4
		(1,3)	88.3	88.6	86.5	91.0	65.0	64.2	62.2	76.8
		(1.8,6)	83.9	83.0	82.0	87.3	61.7	61.5	59.2	73.6
		(3,25)	69.8	69.4	68.4	72.5	55.8	50.3	48.6	60.0
	15	(0,0)	91.6	90.2	88.4	95.8	69.1	58.2	55.2	82.2
		(1,3)	86.2	86.5	83.7	91.7	65.4	55.4	52.4	74.6
		(1.8,6)	82.4	80.7	78.2	87.2	63.1	56.1	52.6	71.3
		(3,25)	69.3	67.6	65.6	74.6	59.6	46.7	44.7	62.0
	20	(0,0)	90.3	87.8	83.7	95.6	65.2	51.6	46.9	75.4
		(1,3)	85.3	82.5	80.4	90.7	64.2	50.0	46.3	70.0
		(1.8,6)	81.3	78.7	75.4	86.5	63.0	49.4	45.6	68.9
		(3,25)	69.0	67.0	64.4	74.2	63.3	44.3	41.8	65.1
153	10	(0,0)	91.4	93.5	92.5	96.4	52.1	58.0	54.5	78.2
		(1,3)	86.7	89.6	88.0	92.6	44.7	53.8	50.9	63.6
		(1.8,6)	80.4	83.3	81.4	86.9	44.7	52.0	49.9	59.5
		(3,25)	65.6	69.1	67.0	70.6	43.9	43.5	41.1	48.4
	15	(0,0)	88.5	90.1	87.1	95.7	42.8	43.4	38.5	67.5
		(1,3)	81.4	85.0	82.8	91.5	41.0	41.5	38.0	56.1
		(1.8,6)	77.2	80.2	76.2	85.5	44.2	42.4	40.2	55.1
		(3,25)	65.0	66.5	64.6	71.3	49.5	36.9	34.1	52.0
	20	(0,0)	86.1	87.3	82.9	94.8	41.3	33.6	29.7	57.6
		(1,3)	79.0	82.7	78.6	90.5	43.4	33.5	30.2	51.0
		(1.8,6)	74.9	77.5	74.0	84.2	44.8	36.7	32.7	51.0
		(3,25)	63.6	66.3	64.5	72.2	53.4	34.6	30.7	54.1
265	10	(0,0)	87.0	90.0	88.7	94.5	21.9	36.4	33.7	58.1
		(1,3)	82.3	86.5	84.4	89.4	20.3	33.8	31.2	43.7
		(1.8,6)	76.4	81.4	78.0	84.2	23.7	34.0	31.9	42.3
		(3,25)	61.4	66.5	65.0	68.8	28.2	26.7	25.3	33.8
	15	(0,0)	80.5	86.8	84.2	93.0	13.2	20.9	18.4	34.2
		(1,3)	75.9	84.0	79.5	88.7	15.6	21.0	18.6	31.3
		(1.8,6)	72.9	77.5	73.9	83.4	20.8	23.8	21.6	32.4
		(3,25)	61.2	65.5	61.1	69.8	32.6	22.1	19.7	35.0
	20	(0,0)	74.5	79.8	76.2	91.1	9.6	13.7	11.4	22.1
		(1,3)	70.2	78.0	72.7	86.0	14.9	15.1	13.1	23.0
		(1.8,6)	66.6	74.0	68.3	81.3	20.5	17.9	15.5	28.5
		(3,25)	57.9	62.7	58.1	69.7	36.7	19.3	17.4	38.7

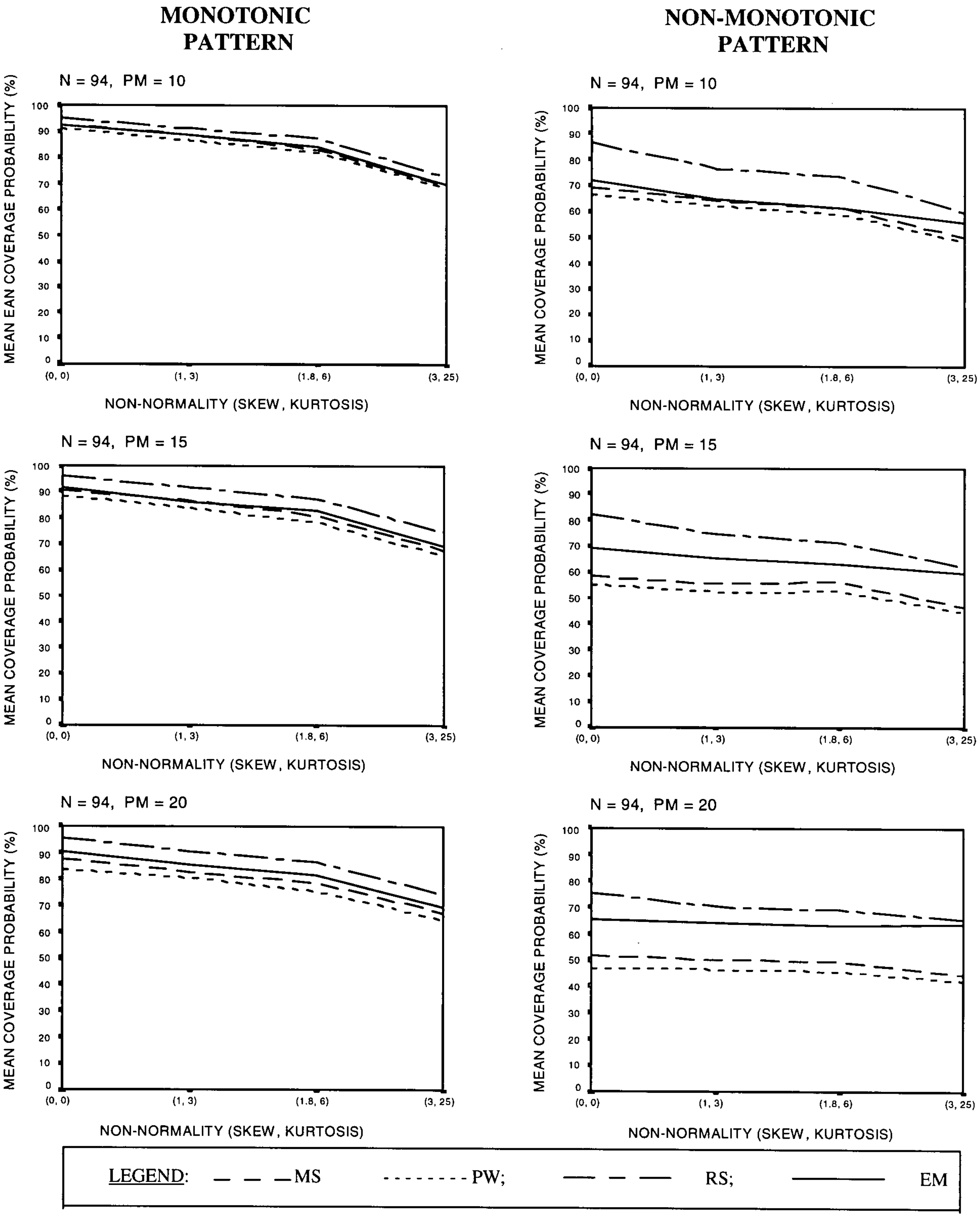


Figure 23: Mean coverage probability for β_1 (N = 94)
Note: N = Sample size, PM = Percent missing.

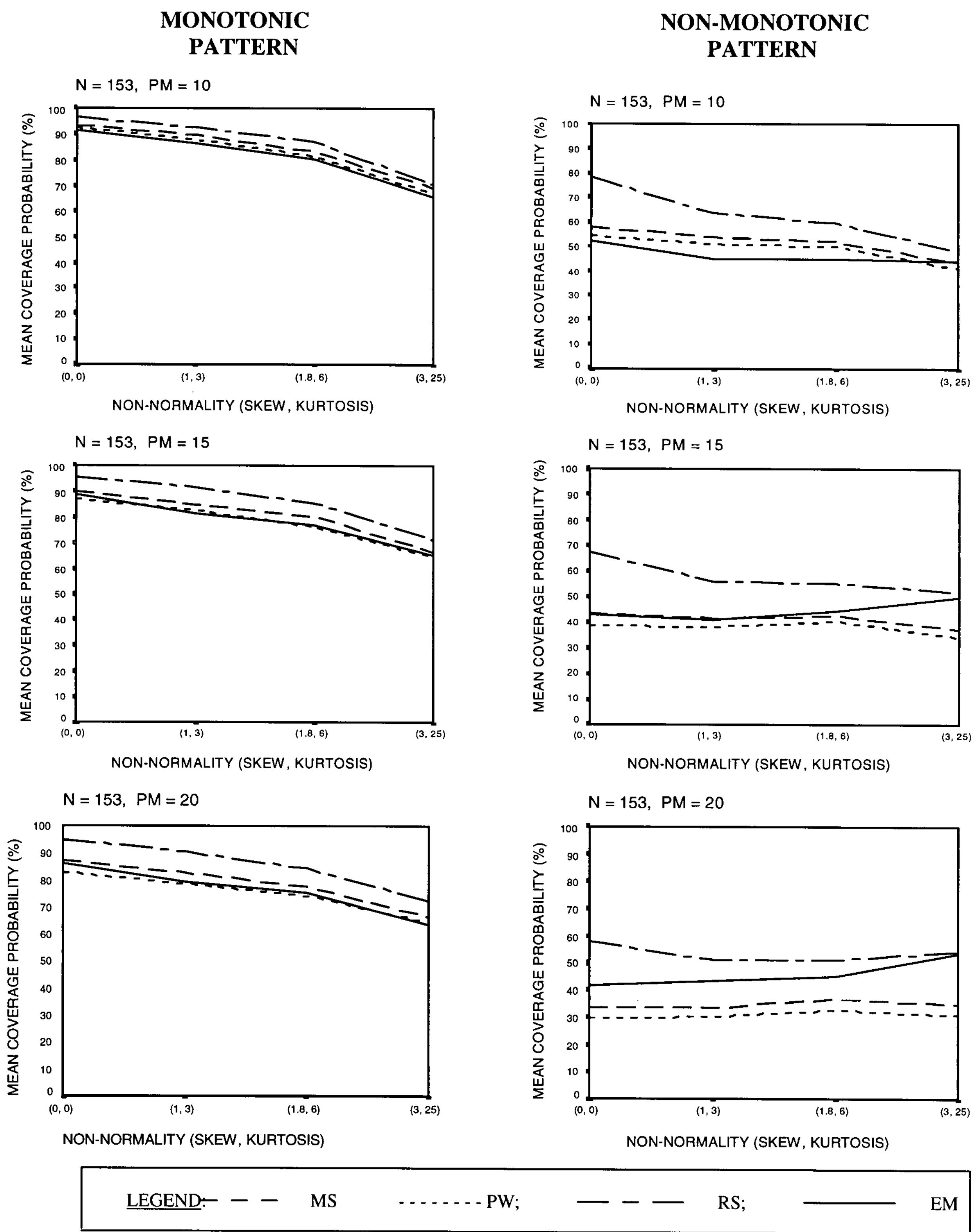


Figure 24: Mean coverage probability for β_1 (N = 153)
<u>Note</u>: N = Sample size, PM = Percent Missing

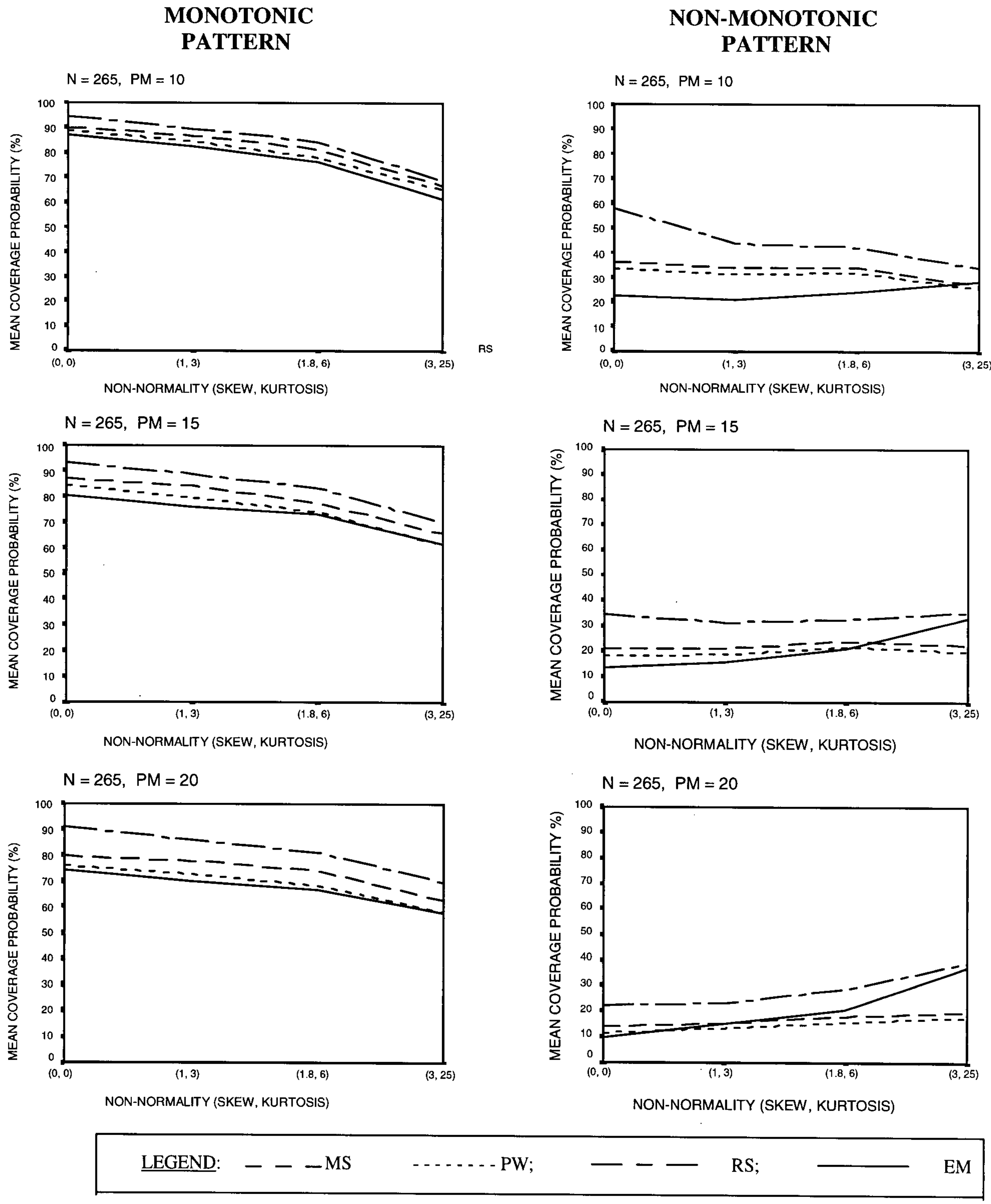


Figure 25: Mean coverage probability for β_1 (N = 265)
Note: N = Sample size, PM = Percent missing.

Summary

In the previous sections, the effects of design variables on absolute error for parameter estimates were presented. Whereas pattern effects were of practical significance under EM and RS at sample sizes 153 and 265, respectively, the interaction effect of pattern and non-normality on absolute error for $R^2_{estimate}$ was also of practical significance under EM and RS. However, this was not the case under MS and PW. Under MS, only the effect of percent missing was of practical significance at sample size 265. Under PW, it was only the effect of non-normality that was of practical significance at all levels of sample size. It appears that MS was the least sensitive to the effects of pattern and non-normality. PW was also insensitive to pattern effects, though it was affected by non-normality.

With respect to absolute error for $R^2_{estimate}$, the performance of all MDTs deteriorated with increasing non-normality. Regardless of missing pattern, PW outperformed the other MDTs, and generally EM produced the largest mean absolute errors for $R^2_{estimate}$. However, the performance of MDTs was more differentiated under non-monotonic missing pattern than under monotonic pattern. MDTs were more differentiated at the highest level of percent missing (20%) than at the lowest level (10%). Also, under monotonic pattern, and at each level of percent missing, MDTs were most differentiated for normal data and least differentiated at the highest level of non-normality.

With respect to $R^2_{estimate}$, the smallest bias was under PW. This was at all levels of the design variables. Except for PW, the MDTs tended to underestimate population R^2. Thus, in the estimation of R^2, PW method was the most accurate and least biased.

It was only for b_1 that missing pattern had an effect of practical significance. This was true for all MDTs. In addition, the effect of non-normality on absolute error for regression coefficients was of practical significance under all MDTs, except EM.

Overall, RS was the most accurate under both monotonic and non-monotonic missing patterns, and EM the least accurate in estimating β_1. The mean absolute error for b_1 was smaller under monotonic pattern than under non-monotonic pattern, and the MDTs became more differentiated as percent missing increased. In the estimation of β_1, RS was the most accurate and least biased.

RS generally provided the best coverage probability under both monotonic and non-monotonic missing patterns. However, all MDTs had poor coverage probabilities under non-monotonic missing pattern. Also, coverage probability generally decreased with increasing non-normality.

Study 3: Nine Predictors under Low R^2 Condition

Effects on Absolute Error for $R^2_{estimate}$

The effects of design variables on absolute error for $R^2_{estimate}$ are listed in Table 15. Using $\eta^2_{alt} \geq .03$ as the criterion, the table shows that with a sample of 94, the effect of non-normality was of practical significance under each MDT, with the largest effect being that of non-normality under MS, and the smallest under PW. With sample size 153, the effect of non-normality on absolute error for $R^2_{estimate}$ was of practical significance under all MDTs. The largest effect was that of non-normality under MS, and the smallest was under PW. With sample size 265, the effect of non-normality was of practical significance under all MDTs. The largest effect was that of non-normality under PW and the smallest under EM.

Table 15
Effects on absolute error for $R^2_{estimate}$ fulfilling the criterion of $\eta^2_{alt} \geq .03$

Sample size	MDT	Effect	SS_{effect}	$SS_{residual}$	df	η^2_{alt}
94	EM	Non-normality	3.57	72.01	3	.047
	MS	Non-normality	5.21	102.34	3	.048
	PW	Non-normality	4.10	105.30	3	.037
	RS	Non-normality	3.57	84.47	3	.041
153	EM	Non-normality	1.55	33.82	3	.044
	MS	Non-normality	2.23	47.65	3	.045
	PW	Non-normality	1.95	54.62	3	.034
	RS	Non-normality	1.60	39.24	3	.039
265	EM	Non-normality	0.62	19.90	3	.030
	MS	Non-normality	0.82	23.87	3	.033
	PW	Non-normality	1.05	27.91	3	.036
	RS	Non-normality	0.73	21.40	3	.033

Note: EM = EM imputation, MS = Mean substitution, PW = Pairwise deletion, RS = Regression imputation.

Relative Performance of MDTs in the Estimation of Population R^2

Table 16 contains means and standard deviations of mean absolute error for $R^2_{estimate}$ under various design conditions. The influence of non-normality on the performance of MDTs at sample size 94 is revealed in Figure 26. The performance of all MDTs deteriorated with increasing non-normality. The smallest mean absolute bias was under EM for both monotonic and non-monotonic missing patterns, and at all levels of percent missing. Also, EM and RS tended to give smaller absolute errors under non-monotonic pattern than under monotonic pattern. Under non-monotonic missing pattern, EM overlapped with RS and MS overlapped with PW. In sum, with a sample of 94, and regardless of missing pattern, EM performed the best.

The influence of non-normality on the performance of MDTs at sample size 153 is revealed in Figure 27. The performance of all MDTs deteriorated with increasing non-normality as measured by the mean absolute error for $R^2_{estimate}$. For both monotonic and non-monotonic missing patterns, EM performed the best, followed by RS, MS and PW, in that order. With sample size 153, EM performed the best regardless of missing pattern and percent missing. However, under non-monotonic missing pattern, the difference in performance between EM and RS became smaller with increasing percent missing.

The influence of non-normality on the relative performance of MDTs at sample size 265 is revealed in Figure 28. The performance of all MDTs deteriorated with increasing non-normality, although mean absolute error was generally much smaller than that under sample size 94 and 153. With monotonic missing pattern, EM performed the best, followed by RS, MS and PW, in that order. MDTs differed most in performance at the highest level of non-normality (skew = 3 and kurtosis = 25). Also, the relative performance among EM, MS and RS became more and more undifferentiated with increasing percent missing.

With non-monotonic missing pattern at 10% missing, EM and RS performed the best. This was true at 15% missing, although MS performed equally well. At 20% missing, MS and PW performed the best, and EM performed the worst.

Table A.5 in Appendix A contains bias in $R^2_{estimate}$ under low R^2 condition with nine predictors. Bias in $R^2_{estimate}$ varied from –.04 to .09 across all conditions of the design variables, giving a range of .13. These values were plotted in figure B.5, a typical set of graphs for bias in $R^2_{estimate}$. At sample size 94, the smallest bias was under RS followed by PW. Similar to the results under low R^2 with four predictors, not a single MDT produced the smallest bias across all design conditions.

Table 16
Means and standard deviations* of absolute error for $R^2_{estimate}$

			Monotonic Missing Pattern				Non-monotonic Missing Pattern			
N	Percent Missing	Normality (sk., kurt.)	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation
94	10	(0,0)	0651(050)	0712(054)	0789(055)	0696(053)	0587(042)	0729(052)	0764(055)	0653(047)
		(1,3)	0699(053)	0766(059)	0822(059)	0747(057)	0596(046)	0761(057)	0789(058)	0656(050)
		(1.8,6)	0807(060)	0883(067)	0917(067)	0859(065)	0663(053)	0843(064)	0881(064)	0719(057)
		(3,25)	1047(085)	1153(093)	1159(092)	1108(090)	0823(070)	1068(086)	1083(086)	0864(074)
	15	(0,0)	0588(043)	0665(050)	0754(054)	0643(048)	0521(039)	0691(051)	0719(052)	0581(043)
		(1,3)	0632(048)	0716(055)	0783(057)	0688(053)	0538(041)	0719(055)	0745(055)	0584(045)
		(1.8,6)	0732(056)	0828(064)	0872(064)	0792(062)	0604(048)	0803(061)	0833(061)	0648(052)
		(3,25)	0972(081)	1103(091)	1093(089)	1041(087)	0764(065)	1015(082)	1033(084)	0806(070)
	20	(0,0)	0531(040)	0616(047)	0717(052)	0582(044)	0479(037)	0651(048)	0678(049)	0517(039)
		(1,3)	0576(044)	0670(053)	0751(055)	0629(049)	0507(038)	0681(052)	0709(051)	0537(041)
		(1.8,6)	0666(052)	0778(062)	0836(063)	0723(058)	0571(044)	0761(058)	0792(058)	0600(047)
		(3,25)	0893(077)	1046(090)	1066(095)	0952(083)	0737(061)	0956(079)	0991(079)	0774(067)
153	10	(0,0)	0467(035)	0506(039)	0569(042)	0499(039)	0396(030)	0492(037)	0552(041)	0445(033)
		(1,3)	0504(036)	0547(040)	0596(043)	0539(040)	0410(031)	0517(039)	0577(042)	0453(034)
		(1.8,6)	0563(041)	0611(045)	0653(048)	0601(044)	0457(034)	0574(044)	0630(046)	0498(038)
		(3,25)	0738(058)	0806(065)	0827(065)	0786(063)	0559(046)	0722(059)	0775(061)	0582(049)
	15	(0,0)	0421(032)	0467(036)	0550(041)	0455(035)	0377(028)	0467(036)	0523(039)	0406(031)
		(1,3)	0453(033)	0505(038)	0579(042)	0490(037)	0389(030)	0484(037)	0548(040)	0410(031)
		(1.8,6)	0507(037)	0565(042)	0631(046)	0547(041)	0426(032)	0534(041)	0599(044)	0447(034)
		(3,25)	0677(054)	0759(061)	0794(063)	0724(058)	0533(043)	0672(055)	0734(060)	0552(046)
	20	(0,0)	0410(030)	0449(035)	0530(039)	0438(033)	0389(029)	0447(034)	0500(038)	0393(030)
		(1,3)	0440(032)	0491(036)	0559(040)	0472(034)	0405(030)	0464(035)	0526(038)	0406(031)
		(1.8,6)	0489(036)	0549(041)	0609(044)	0524(038)	0433(032)	0504(039)	0575(042)	0439(033)
		(3,25)	0653(050)	0738(059)	0770(060)	0694(054)	0544(042)	0629(053)	0720(061)	0559(043)
265	10	(0,0)	0331(024)	0351(026)	0397(030)	0349(026)	0318(023)	0352(026)	0385(030)	0335(024)
		(1,3)	0347(026)	0370(028)	0417(031)	0367(027)	0325(024)	0369(027)	0385(030)	0335(024)
		(1.8,6)	0382(028)	0410(031)	0458(034)	0405(030)	0347(025)	0401(030)	0441(033)	0361(027)
		(3,25)	0508(040)	0549(044)	0589(046)	0538(043)	0427(032)	0495(041)	0554(043)	0431(033)
	15	(0,0)	0335(025)	0350(026)	0388(029)	0348(026)	0352(025)	0352(026)	0373(029)	0346(025)
		(1,3)	0346(026)	0366(028)	0405(031)	0362(027)	0368(026)	0360(027)	0390(029)	0355(026)
		(1.8,6)	0376(028)	0401(030)	0444(033)	0395(029)	0378(026)	0382(029)	0426(032)	0369(027)
		(3,25)	0489(038)	0529(043)	0576(045)	0513(041)	0461(033)	0470(038)	0530(041)	0460(033)
	20	(0,0)	0361(026)	0361(027)	0380(029)	0363(026)	0427(027)	0370(026)	0359(028)	0399(027)
		(1,3)	0370(027)	0373(028)	0395(030)	0373(027)	0442(028)	0369(027)	0376(029)	0417(028)
		(1.8,6)	0390(028)	0403(030)	0430(033)	0399(029)	0447(029)	0386(029)	0410(031)	0429(028)
		(3,25)	0488(036)	0524(042)	0544(043)	0507(039)	0525(033)	0461(037)	0515(040)	0530(033)

* All values in the table are preceded by an omitted decimal point. Standard deviations are in parentheses. Standard errors of the means can be obtained by dividing each standard deviation by 1000½.

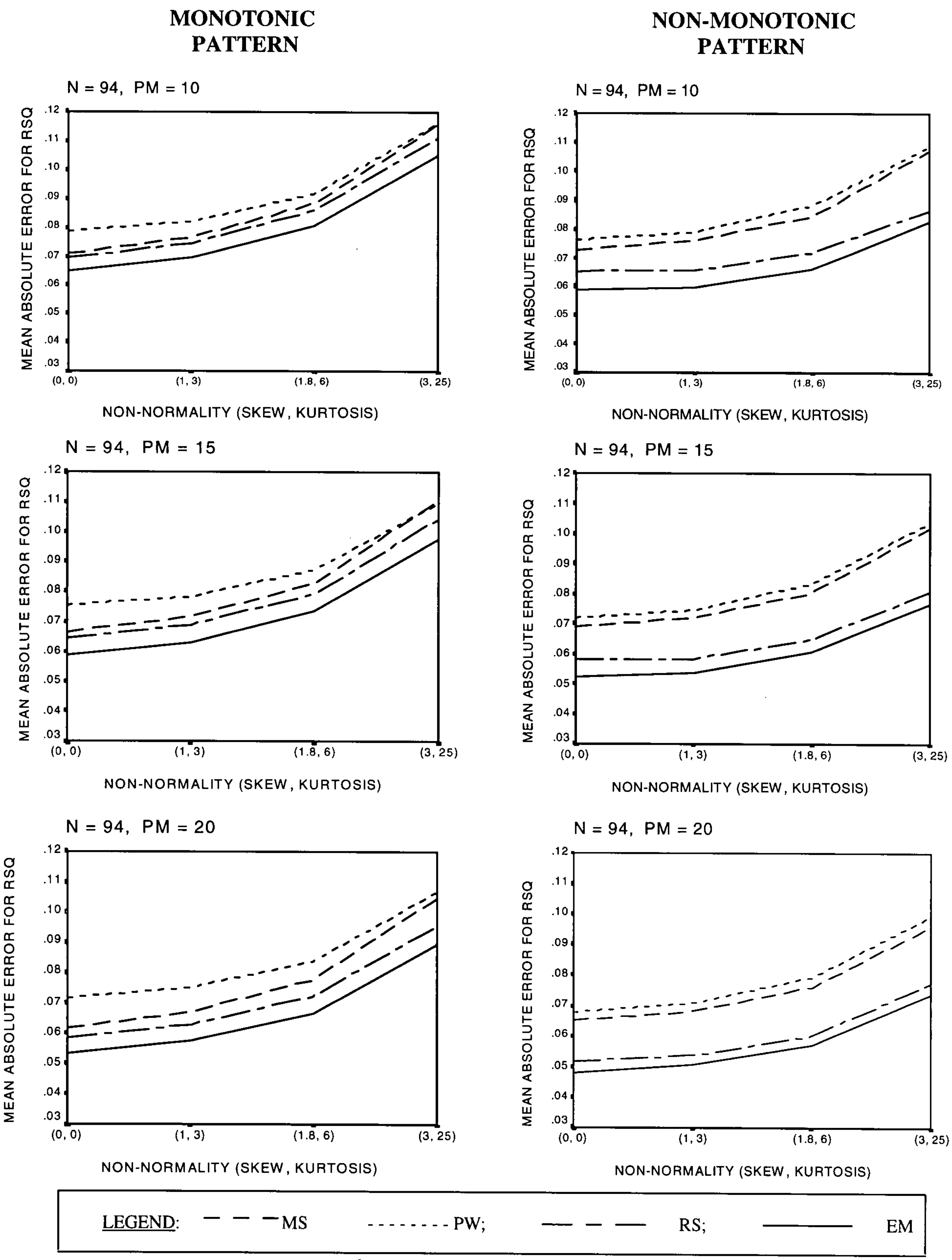


Figure 26: Mean absolute error for $R^2_{estimate}$ across levels of non-normality (N = 94)
Note: N = Sample size, PM = Percent missing.

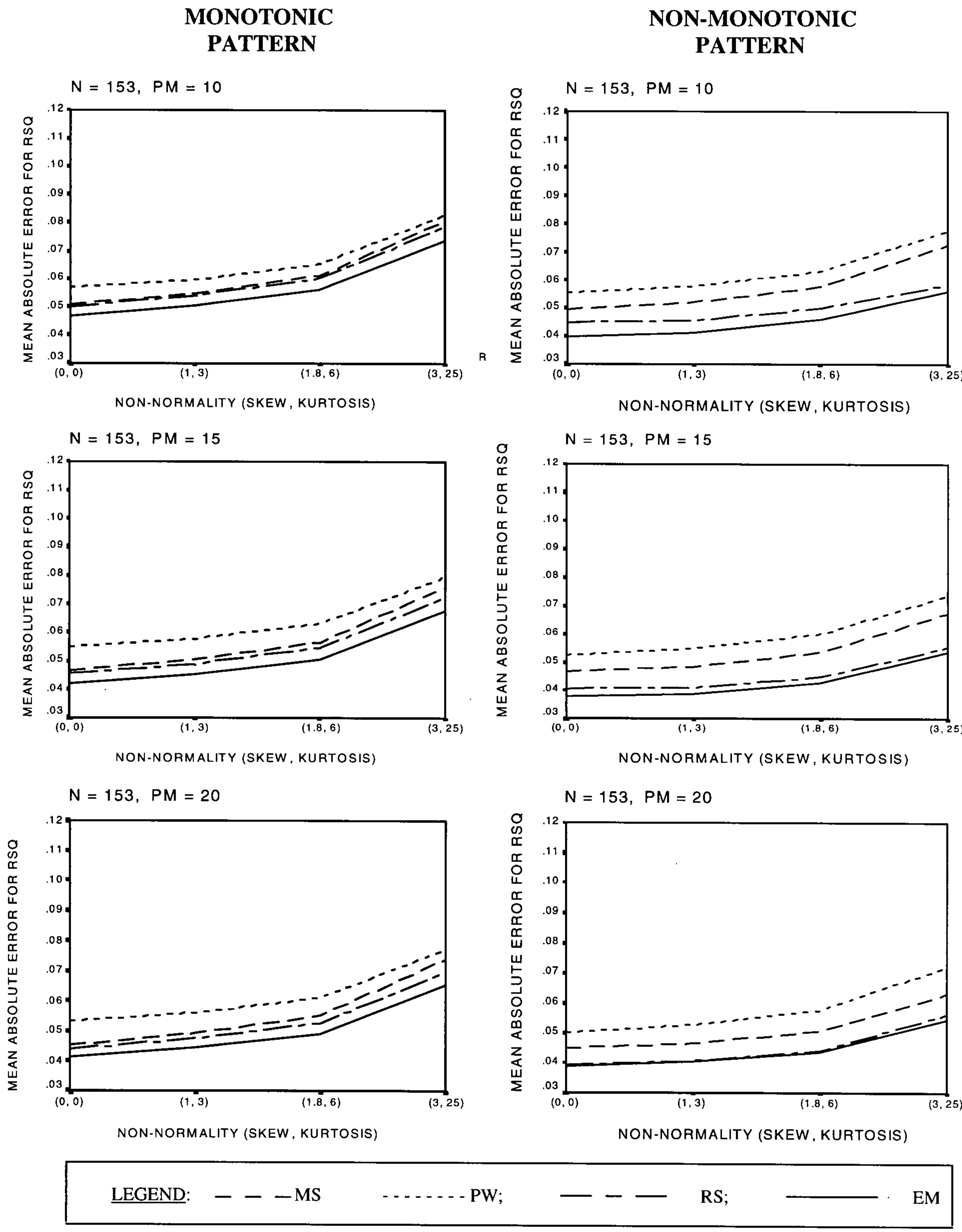


Figure 27: Mean absolute error for $R^2_{estimate}$ across levels of non-normality (N = 153)
Note: N = Sample size, PM = Percent missing.

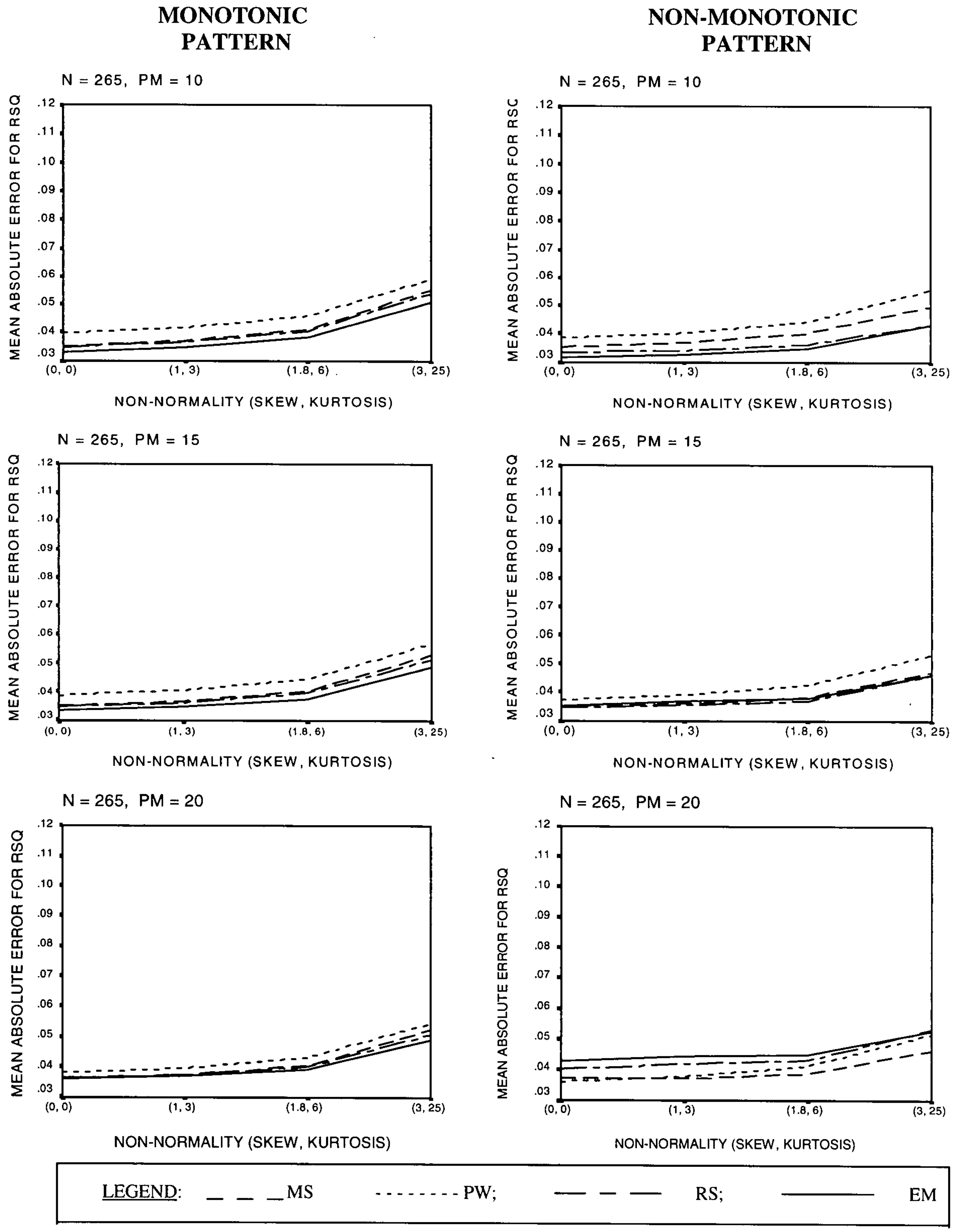


Figure 28: Mean absolute error for $R^2_{estimate}$ across levels of non-normality (N = 265)
Note: N = Sample size, PM = Percent missing.

Effects on Absolute Error for Regression Coefficients

The effects of design variables on absolute error for regression coefficients are listed in Table 17. No effects fulfilled $\eta^2_{alt} \geq .03$ as the criterion for $b_2,\ldots,b_9$. Using the same criterion, the table shows that the effect of non-normality was of practical significance under RS only at sample size 94 and 153. The strength of the effect of non-normality increased with increasing sample size.

With sample size 265, the effect of pattern (PT) on absolute error for b_1 was of practical significance under all MDTs. Pattern effects were strongest under EM algorithm. In addition to pattern effects, the effect of non-normality was of practical significance under RS only at all levels of sample size.

Table 17
Effects on absolute error for regression coefficients fulfilling the criterion of $\eta^2_{alt} \geq .03$*

Sample size	MDT	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9
94	EM	-	-	-	-	-	-	-	-	-
	MS	-	-	-	-	-	-	-	-	-
	PW	-	-	-	-	-	-	-	-	-
	RS	NL(033)	-	-	-	-	-	-	-	-
153	EM	-	-	-	-	-	-	-	-	-
	MS	-	-	-	-	-	-	-	-	-
	PW	-	-	-	-	-	-	-	-	-
	RS	NL(046)	-	-	-	-	-	-	-	-
265	EM	PT(101)	-	-	-	-	-	-	-	-
	MS	PT(039)	-	-	-	-	-	-	-	-
	PW	PT(038)	-	-	-	-	-	-	-	-
	RS	NL(071)	-	-	-	-	-	-	-	-
		PT(045)	-	-	-	-	-	-	-	-

* Values of η^2_{alt} are in parentheses preceded by an omitted decimal point.
Note: EM = EM imputation, MS = Mean substitution, PW = Pairwise deletion, RS = Regression imputation, PT = Pattern, NL= Non-normality, - = No effect of practical significance.

Relative Performance of MDTs in the Estimation of Regression Coefficients

Table 18 contains means and standard deviations of absolute error for b_1 under various design conditions. The influence of non-normality on the performance of MDTs at sample size

94 is revealed in Figure 29. The performance of all MDTs deteriorated with increasing non-normality as measured by the mean absolute error for b_1. This was regardless of missing pattern.

With sample size 94 under monotonic missing pattern, EM and RS both outperformed MS and PW, with EM and RS performing at the same level when percent missing was the lowest (10%). However, RS performed better at moderate to high levels of percent missing (15% to 20%). As for MS and PW, there was little differentiation in performance across all levels of percent missing. Two pairings were apparent, that of MS and PW and that of EM and RS. The gap between the two pairings increased with percent missing.

With non-monotonic missing pattern, RS and EM again outperformed MS and PW. However, RS performed the best under conditions of low to moderate levels of non-normality. At the extreme level of non-normality (skew = 3 and kurtosis = 25), EM performed slightly better than RS. The mean absolute error for b_1 under MS and PW was larger under non-monotonic than under monotonic missing pattern. The MDTs differed most in their performance for normal data.

Figure 30 is a set of graphical plots of the mean absolute error at sample size 153. Under monotonic missing pattern, the results were similar to those under sample size 94, where EM and RS both outperformed MS and PW. There was little differentiation in the performance of MS and PW at all levels of percent missing. As for EM and RS, difference in performance increased with percent missing, with RS doing better than EM as percent missing increased.

With sample size153 under non-monotonic pattern, RS performed the best, followed by EM, with MS and PW tying in third place. However, EM outperformed RS under conditions of extreme non-normality (skew=3 and kurtosis=25). MDTs differed most for normal data, and least at the highest level of non-normality.

Figure 31 is a graphical plot of mean absolute error at sample size 265. With monotonic missing pattern, RS performed the best, although differentiation between EM and RS was smallest at the lowest level of percent missing (10%) and largest at the highest level of percent missing (20%). MS and PW were at par in third place.

With non-monotonic missing pattern, RS performed the best. The MDTs differed least at the extreme level of non-normality (skew = 3 and kurtosis = 25), and most with normal data. Whereas MS and PW performed equally at all levels of percent missing, their performance relative to EM was differentiated most at the highest level of percent missing (20%), and differentiated least at the lowest level of percent missing (10%).

In sum, the effects of missing pattern and non-normality on absolute error for regression coefficients were of practical significance only for b_1. Mean absolute error increased with increasing non-normality. The MDTs were most differentiated with respect to mean absolute error for b_1 only. The best overall performance in the estimation of β_1 was under RS, followed by EM. The performance of MS and PW was undifferentiated under both monotonic and non-monotonic missing pattern. Also, differentiation among MDTs was less prominent under monotonic pattern than under non-monotonic pattern. Under non-monotonic pattern, the MDTs differed most with normal data.

Table A.6 in Appendix A shows the bias in b_1. Figure B.6 is a typical set of graphs of bias in b_1, showing that the smallest bias across all levels of the design variables was under PW. The bias in b_1 varied from –.08 to .09 across all design conditions, giving a range of .17. Whereas EM and MS consistently overestimated β_1, PW and RS consistently underestimated β_1. The bias for b_2 to b_9 was much smaller compared to that for b_1, with a maximum range of .05. Therefore, the graphs for bias in b_2 to b_9 were omitted.

Table 18
Means and standard deviations* of absolute error for b_1

			Monotonic Missing Pattern				Non-monotonic Missing Pattern			
N	Percent Missing	Normality (sk., kurt.)	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation
94	10	(0,0)	0682(050)	0784(059)	0785(059)	0696(052)	0686(041)	0880(067)	0883(067)	0607(041)
		(1,3)	0739(055)	0833(065)	0835(065)	0746(057)	0729(041)	0927(077)	0928(077)	0659(042)
		(1.8,6)	0815(063)	0913(073)	0915(073)	0822(064)	0759(042)	0989(087)	0989(087)	0711(043)
		(3,25)	0929(084)	1035(098)	1037(098)	0938(084)	0860(043)	1142(118)	1142(118)	0903(054)
	15	(0,0)	0683(048)	0808(060)	0810(060)	0679(051)	0709(039)	0923(069)	0928(070)	0614(039)
		(1,3)	0736(053)	0852(067)	0856(067)	0727(055)	0751(039)	0966(079)	0968(080)	0682(040)
		(1.8,6)	0802(061)	0923(075)	0928(075)	0796(062)	0775(041)	1027(090)	1026(090)	0728(044)
		(3,25)	0916(082)	1045(099)	1050(100)	0908(081)	0866(042)	1193(123)	1192(123)	0955(061)
	20	(0,0)	0664(047)	0822(061)	0827(062)	0653(049)	0711(037)	0960(072)	0969(073)	0655(039)
		(1,3)	0714(051)	0859(061)	0869(069)	0698(053)	0743(037)	1000(082)	1004(083)	0723(041)
		(1.8,6)	0780(059)	0935(076)	0944(077)	0765(061)	0759(038)	1052(092)	1053(093)	0758(044)
		(3,25)	0891(079)	1062(102)	1073(103)	0880(080)	0845(044)	1224(125)	1226(126)	0969(062)
153	10	(0,0)	0552(037)	0607(045)	0608(045)	0547(039)	0630(036)	0708(053)	0709(053)	0483(033)
		(1,3)	0568(041)	0628(049)	0630(049)	0547(039)	0680(036)	0755(059)	0755(059)	0537(034)
		(1.8,6)	0622(045)	0688(053)	0690(053)	0624(046)	0699(038)	0807(065)	0807(065)	0577(037)
		(3,25)	0727(061)	0798(070)	0801(071)	0726(063)	0829(038)	0970(092)	0968(092)	0802(044)
	15	(0,0)	0544(037)	0620(046)	0622(047)	0529(038)	0651(034)	0766(056)	0769(056)	0523(033)
		(1,3)	0564(039)	0645(049)	0646(050)	0549(040)	0695(035)	0811(061)	0812(061)	0592(034)
		(1.8,6)	0613(042)	0700(054)	0702(055)	0600(074)	0711(036)	0857(068)	0858(068)	0627(036)
		(3,25)	0712(058)	0803(072)	0806(072)	0696(059)	0821(037)	1018(096)	1015(096)	0868(049)
	20	(0,0)	0544(036)	0627(047)	0630(048)	0507(037)	0679(034)	0819(059)	0822(059)	0576(033)
		(1,3)	0565(038)	0653(051)	0657(051)	0529(039)	0712(034)	0862(064)	0863(064)	0647(035)
		(1.8,6)	0608(042)	0709(056)	0713(056)	0578(043)	0724(035)	0906(071)	0906(071)	0676(037)
		(3,25)	0710(057)	0820(073)	0823(074)	0684(058)	0811(037)	1067(099)	1065(099)	0880(050)
265	10	(0,0)	0408(030)	0445(034)	0445(034)	0390(030)	0598(031)	0591(042)	0591(042)	0398(026)
		(1,3)	0435(031)	0480(037)	0481(037)	0425(031)	0664(032)	0637(048)	0636(048)	0475(028)
		(1.8,6)	0481(034)	0527(041)	0528(041)	0470(035)	0682(033)	0672(054)	0671(054)	0516(030)
		(3,25)	0580(043)	0643(053)	0644(053)	0574(045)	0803(035)	0828(073)	0826(073)	0742(038)
	15	(0,0)	0419(029)	0458(035)	0459(035)	0377(029)	0638(029)	0663(045)	0664(045)	0466(027)
		(1,3)	0440(031)	0498(038)	0499(038)	0409(031)	0690(030)	0709(051)	0709(051)	0551(029)
		(1.8,6)	0481(034)	0544(043)	0545(043)	0453(034)	0702(032)	0737(057)	0735(057)	0582(031)
		(3,25)	0574(042)	0659(055)	0660(055)	0553(044)	0798(034)	0891(077)	0888(077)	0798(040)
	20	(0,0)	0431(029)	0473(036)	0475(036)	0369(028)	0659(028)	0721(047)	0722(047)	0534(027)
		(1,3)	0452(030)	0510(040)	0512(040)	0398(030)	0697(029)	0764(053)	0764(053)	0611(029)
		(1.8,6)	0486(033)	0553(044)	0555(045)	0441(033)	0700(031)	0790(059)	0788(059)	0630(032)
		(3,25)	0574(040)	0667(056)	0668(057)	0538(041)	0777(033)	0939(078)	0935(078)	0821(042)

* All values in the table are preceded by an omitted decimal point. Standard deviations are in parentheses. Standard errors of the means can be obtained by dividing each standard deviation by 1000½.

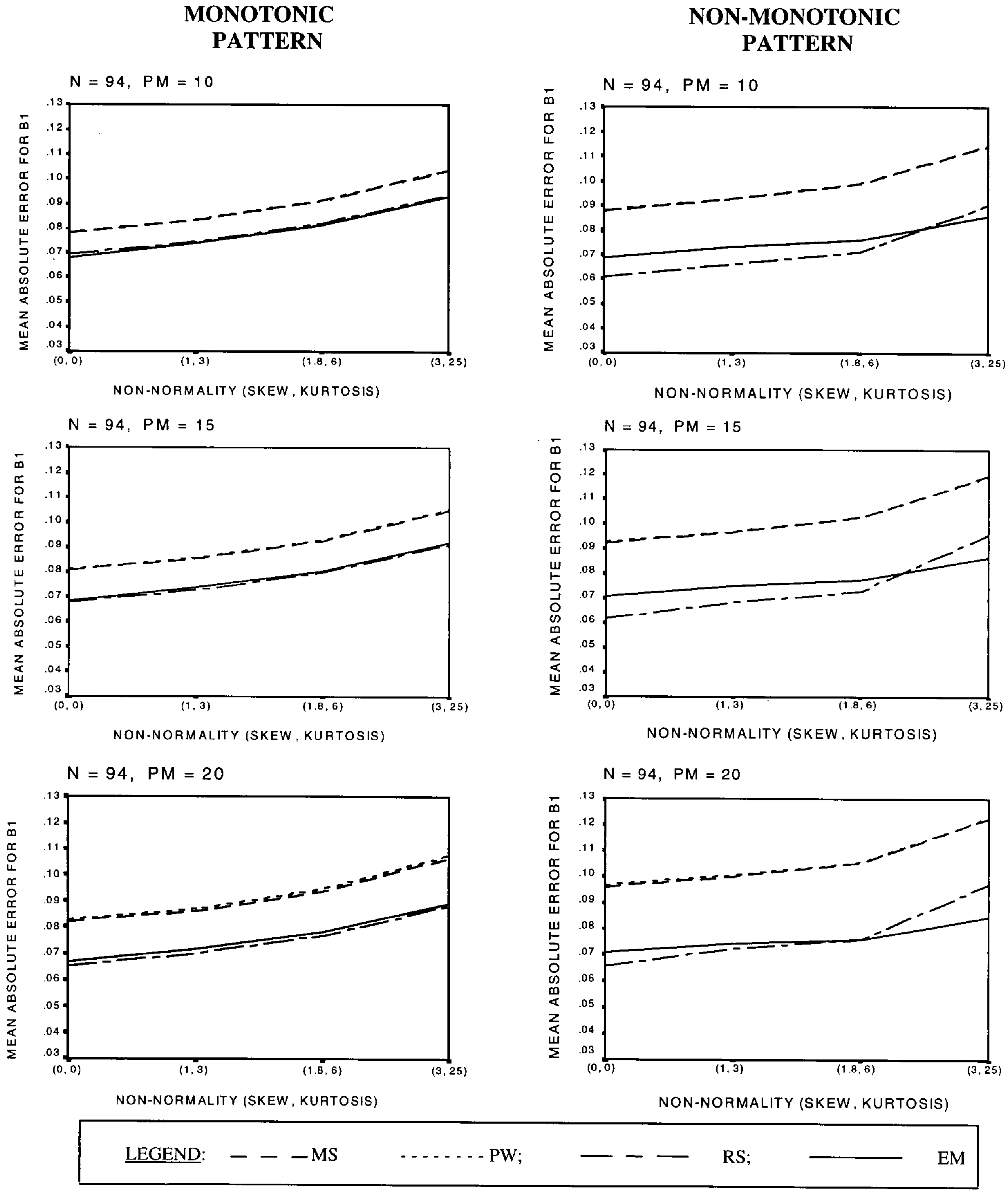


Figure 29: Mean absolute error for b_1 across levels of non-normality (N = 94)
Note: N = Sample size, PM = Percent missing.

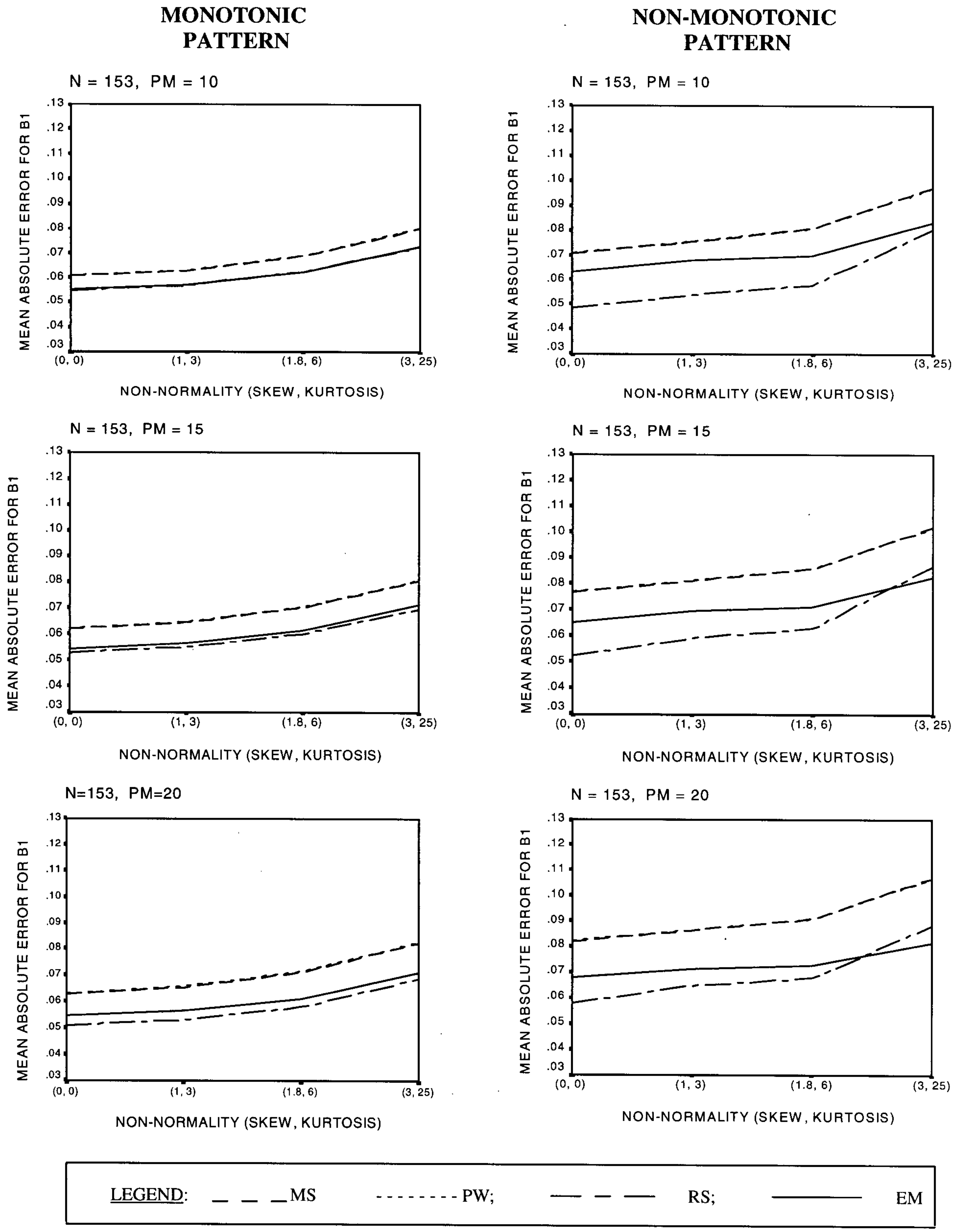


Figure 30: Mean absolute error for b_1 across levels of non-normality (N = 153)
Note: N = Sample size, PM = Percent missing

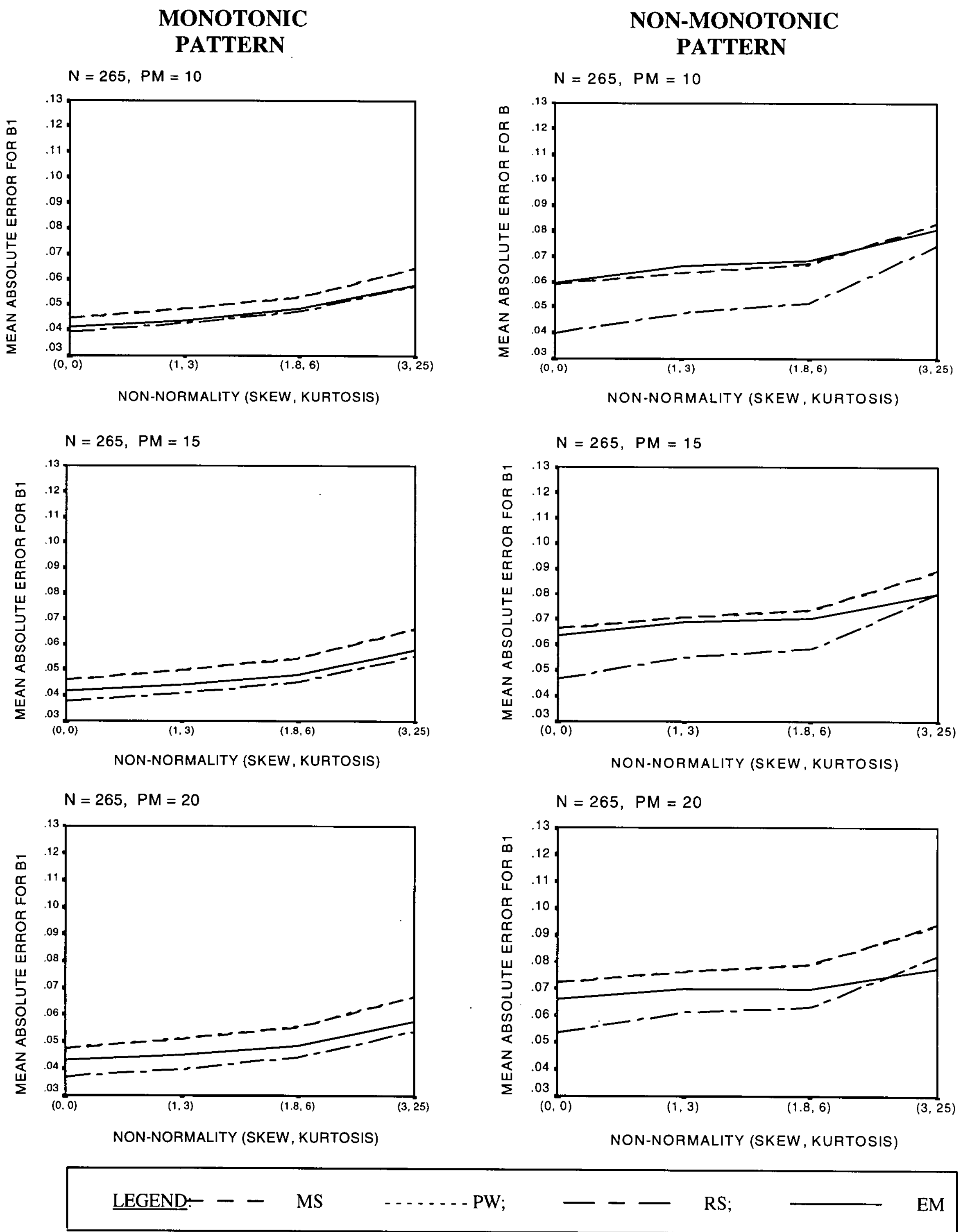


Figure 31: Mean bias in b_1 across levels of non-normality (N = 265)
Note: N = Sample size, PM = Percent missing

Coverage Probability for β_1

Coverage probabilities under monotonic pattern differed from those under non-monotonic pattern only for β_1. In addition, MDTs were not differentiated in their coverage probabilities for β_2, $\beta_3,\ldots,\beta_9$. For this reason, only the findings for β_1 are reported. Table 19 contains mean coverage probabilities for β_1. These coverage probabilities are displayed in Figures 32, 33 and 34 for sample sizes 94, 153 and 265, respectively. The graphs show that coverage probability decreased with increasing non-normality for all MDTs.

At sample size 94, all MDTs had coverage probability above the criterion level of 95% for normal data when missing pattern was monotonic. This suggests that MDTs worked well with monotonic pattern. With non-monotonic pattern, EM and RS were least differentiated with respect to coverage probability which, with the exception of coverage probability at extreme non-normality (skew = 3, kurtosis = 25), were above 95%. This was maintained at all levels of percent missing. For normal data, PW produced coverage probability above 95%, and this was maintained across all levels of percent missing. The coverage probability for MS was the worst, starting below 92% for normal data, and deteriorating with increasing non-normality to 81%.

Coverage probabilities for β_1 at sample size 153 are plotted in Figure 33. The performance of EM, RS, and PW was undifferentiated at all levels of percent missing under monotonic pattern. With non-monotonic pattern, the difference in coverage probabilities for EM and RS was undifferentiated, except at the highest level of non-normality when RS produced a smaller coverage probability than EM. The difference in coverage probabilities among MDTs was largest at extreme non-normality. The graphs also show that the MDTs were most differentiated with respect to coverage probability under non-monotonic pattern than under monotonic pattern. These findings are similar to those under sample size 94. The performance of MDTs at sample size 265 was similar to that under sample sizes 94 and 153.

In sum, all MDTs produced coverage probabilities above 95% for normal data under monotonic missing pattern, with MS producing the smallest coverage probability. However, coverage probabilities decreased with increasing non-normality. The MDTs were most differentiated under non-monotonic missing pattern than under monotonic missing pattern.

Table 19
Mean coverage probability (%) for b_1

N	Percent Missing	Normality (sk., kurt.)	Monotonic Missing Pattern: EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation	Non-monotonic Missing Pattern: EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation
94	10	(0,0)	98.3	95.9	97.7	97.9	99.9	91.7	95.0	99.8
		(1,3)	98.0	94.3	96.6	97.0	99.9	88.9	92.7	99.7
		(1.8,6)	96.5	92.1	94.6	95.6	99.8	86.8	90.2	99.4
		(3,25)	92.4	89.2	92.2	91.8	97.3	82.8	86.8	96.3
	15	(0,0)	98.9	96.4	98.4	98.6	100.0	89.4	95.6	99.9
		(1,3)	98.7	94.7	97.6	98.0	99.9	88.1	93.4	99.7
		(1.8,6)	97.9	92.4	96.7	96.6	100.0	85.9	90.5	99.9
		(3,25)	94.1	89.4	93.4	93.5	97.6	81.3	87.8	96.6
	20	(0,0)	99.3	95.9	98.9	98.8	100.0	88.6	96.5	100.0
		(1,3)	99.0	94.3	98.3	98.0	100.0	86.3	94.7	100.0
		(1.8,6)	98.0	91.6	96.7	97.5	100.0	84.4	91.6	99.6
		(3,25)	95.2	89.4	94.5	94.4	99.5	80.7	88.8	96.8
153	10	(0,0)	98.4	96.2	97.9	98.0	99.8	89.5	93.5	99.4
		(1,3)	98.2	95.6	97.3	97.4	99.4	87.1	92.1	99.3
		(1.8,6)	96.4	92.4	95.4	95.6	99.0	84.6	89.4	98.9
		(3,25)	91.8	88.2	91.3	91.3	95.0	78.9	82.9	94.2
	15	(0,0)	99.0	96.0	98.3	98.5	99.8	87.8	93.9	99.5
		(1,3)	98.7	94.4	97.7	98.2	99.6	84.6	92.1	99.6
		(1.8,6)	97.3	92.4	96.1	96.5	99.3	82.0	90.2	99.2
		(3,25)	93.2	87.9	92.5	92.4	96.3	77.1	83.6	93.9
	20	(0,0)	99.3	95.6	98.6	98.9	96.2	86.4	94.2	99.8
		(1,3)	98.6	94.1	98.0	98.5	95.6	82.8	93.5	99.5
		(1.8,6)	98.3	92.0	97.1	97.5	92.4	79.6	91.1	99.4
		(3,25)	95.0	87.5	93.8	93.9	88.2	75.5	84.6	94.1
265	10	(0,0)	98.8	96.1	97.7	98.2	97.7	88.3	92.7	99.4
		(1,3)	96.8	94.1	96.2	96.7	95.9	84.1	88.2	98.9
		(1.8,6)	95.8	91.5	94.3	94.8	94.3	80.8	84.9	97.6
		(3,25)	90.7	84.7	88.5	90.1	84.7	72.3	75.9	86.3
	15	(0,0)	99.5	95.6	98.2	99.2	98.1	83.4	91.3	99.3
		(1,3)	97.6	93.3	96.4	97.3	96.9	79.5	87.0	98.4
		(1.8,6)	96.1	91.0	94.8	96.0	95.0	76.4	84.3	97.4
		(3,25)	91.4	83.8	90.7	91.4	86.2	68.5	75.9	84.1
	20	(0,0)	99.4	95.4	98.4	99.4	98.8	79.2	91.5	99.4
		(1,3)	98.5	92.7	97.0	98.0	97.9	75.7	88.1	98.5
		(1.8,6)	97.4	90.4	95.6	96.7	96.1	73.6	86.0	97.2
		(3,25)	92.2	83.6	92.3	92.8	89.0	65.6	77.7	83.8

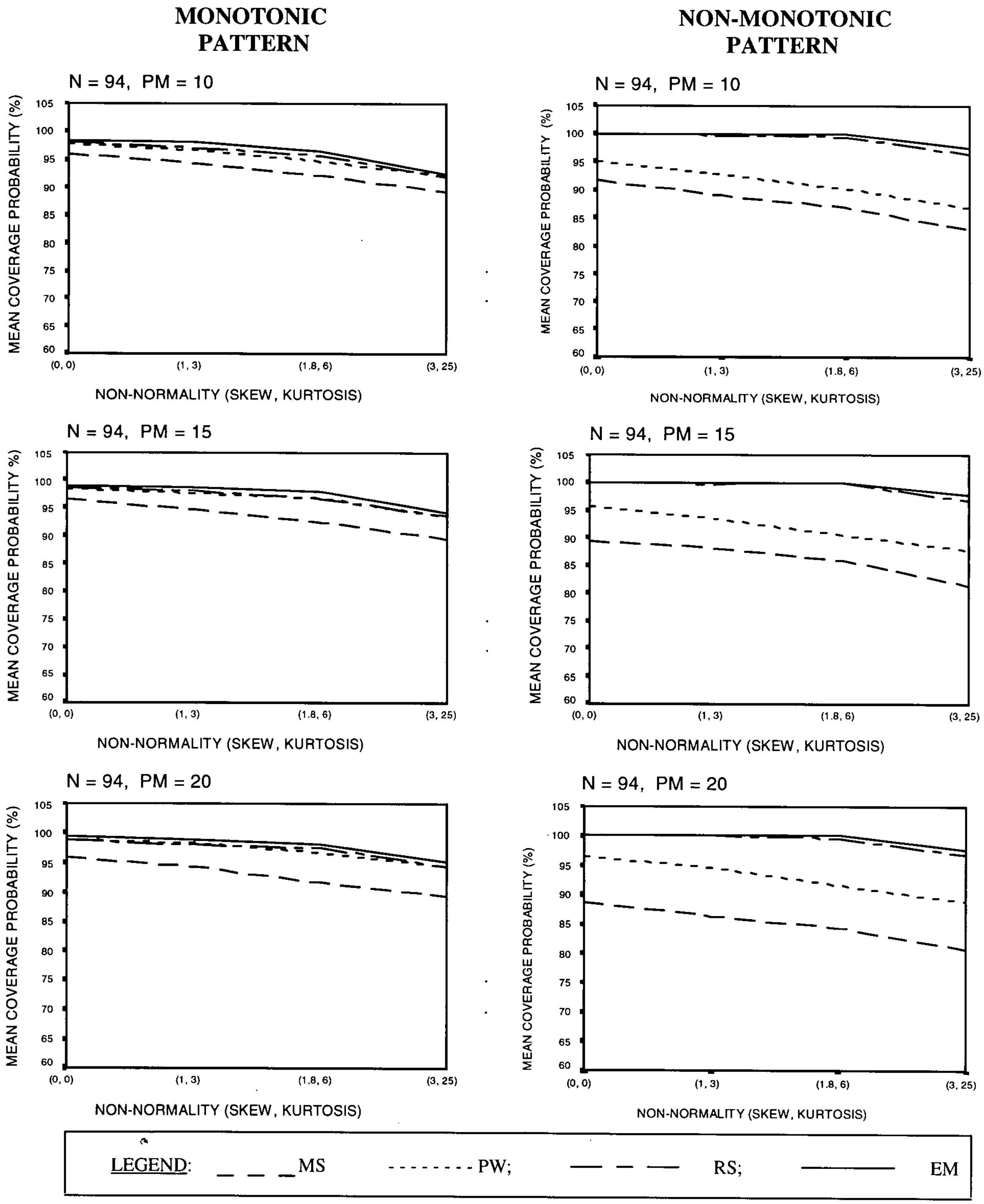


Figure 32: Mean coverage probability for β_1 (N = 94)
Note: N = Sample size, PM = Percent missing

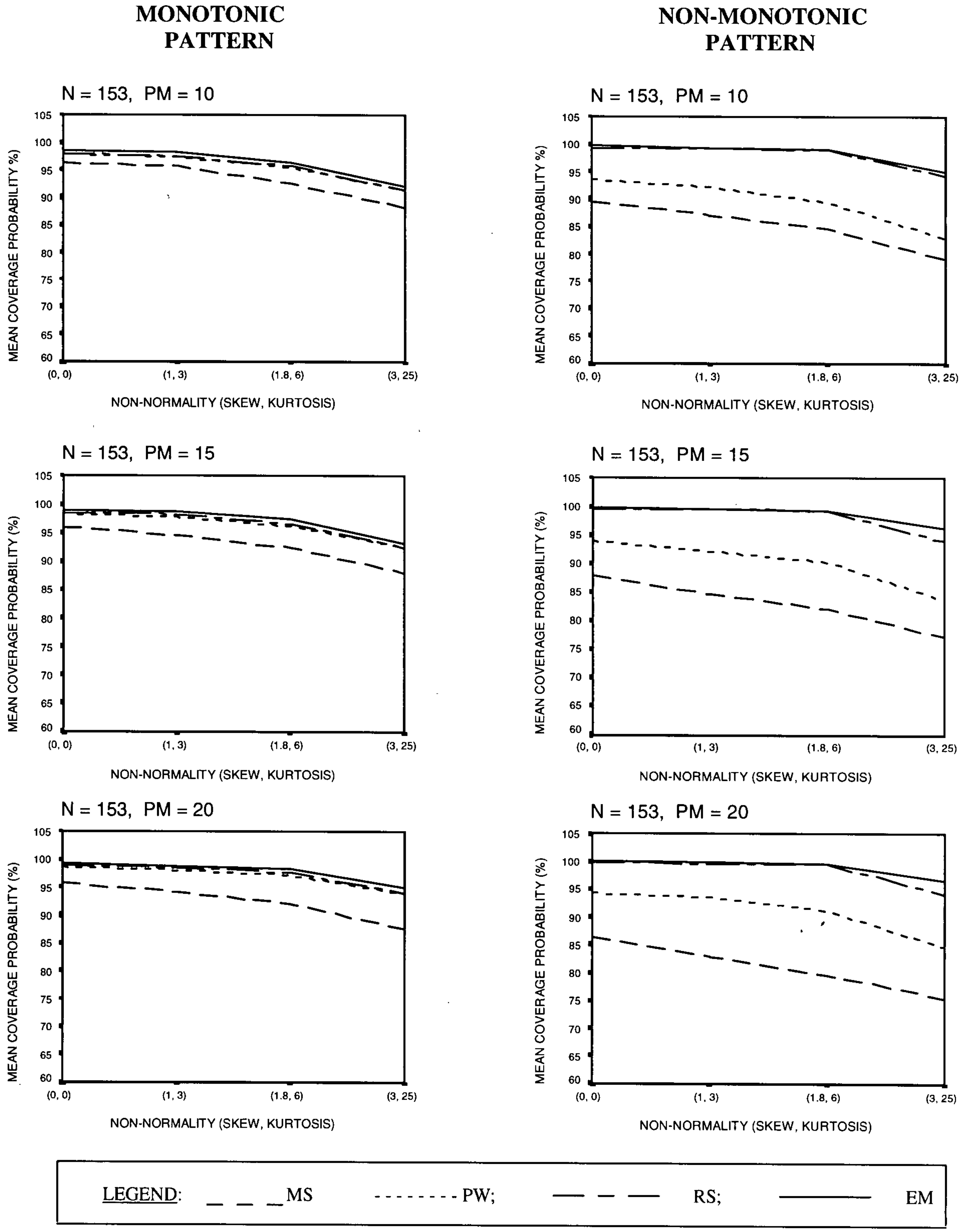
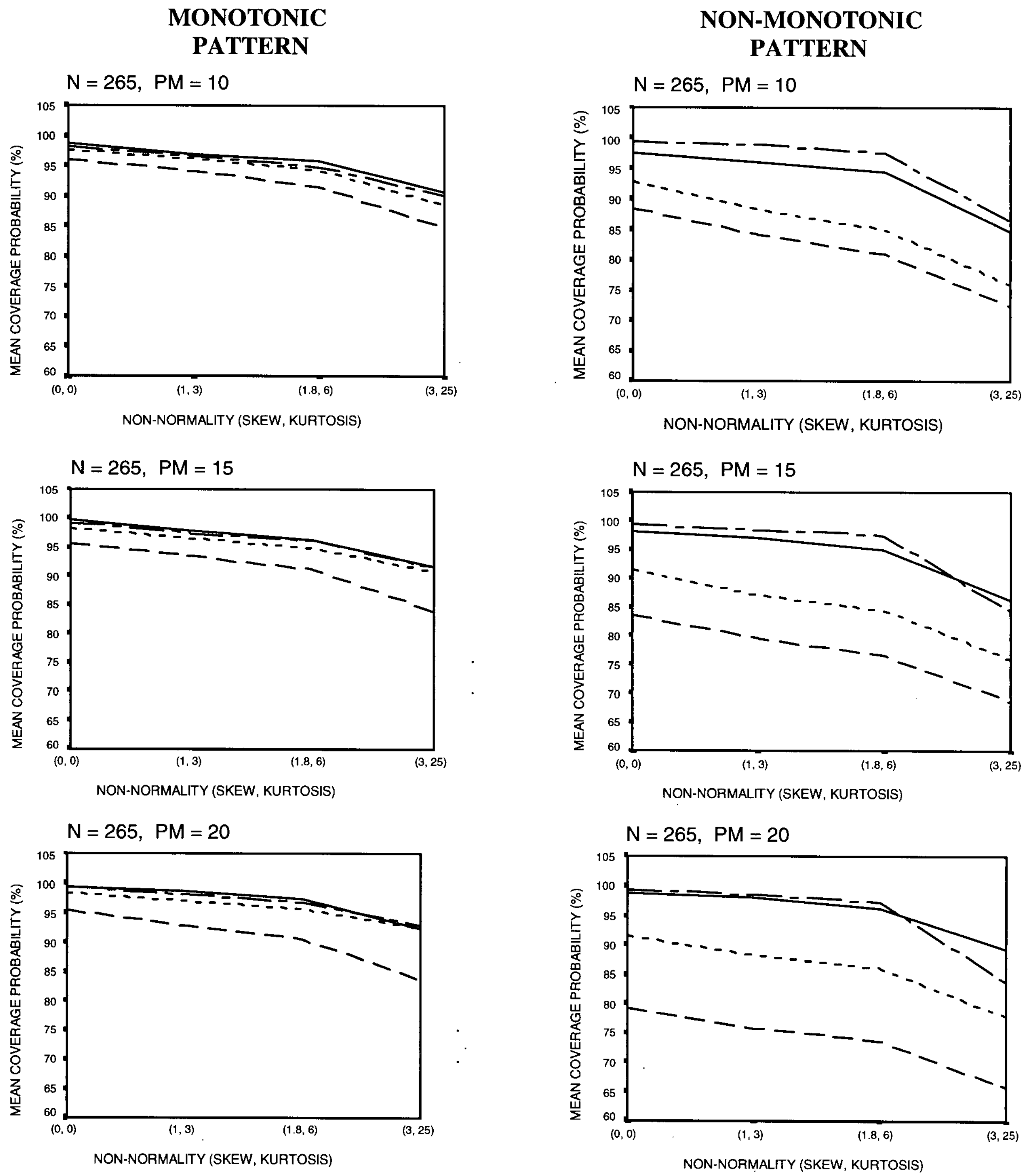


Figure 33: Mean coverage probability for β_1 (N = 153)
Note: N = Sample size, PM = Percent missing.



LEGEND: — — —MS; - - - - - - - - PW; —— — —— RS; ———— EM

Figure 34: Mean coverage probability for β_1 (N = 265)
Note: N = Sample size, PM = Percent missing.

Summary

The effect of non-normality on absolute error for $R^2_{estimate}$ was of practical significance under each MDT at all levels of sample size. The performance of all MDTs with respect to error for $R^2_{estimate}$ decreased with increasing non-normality. EM performed the best under sample size 94 and 153, followed by RS, MS and PW, in that order. At sample size 265, the absolute error for $R^2_{estimate}$ was much smaller than at either sample size 94 or 153. The superiority of EM was for both monotonic and non-monotonic missing patterns. Also, at sample size 265, the relative performance among EM, MS and RS became more and more undifferentiated with increasing percent missing.

The effects of design variables on absolute error for b_1 were also presented in the previous sections. Using $\eta^2_{alt} \geq .03$ as the criterion, the effect of non-normality was of practical significance under RS only. With sample size 265, the effect of pattern was of practical significance under all MDTs. Pattern effects were strongest under EM algorithm, and weakest under PW. In addition to pattern effects, the effect of non-normality was of practical significance under RS. No effects fulfilled the $\eta^2_{alt} \geq .03$ criterion for $b_2,\ldots,b_9$.

The accuracy in parameter estimation under all MDTs deteriorated with increasing non-normality. This was regardless of pattern of missing data. Generally, EM and RS were more accurate than MS and PW. However, RS tended to perform better at moderate to high levels of percent missing (15% to 20%). The smallest bias with respect to R^2 was not under a single MDT across levels of the design variables. At sample size 94, the smallest bias was under RS followed by PW. However, at sample size 153 and 265, the smallest bias was under EM.

In the estimation of β_1, RS was the most accurate under all design conditions. However, the smallest bias in the estimation of β_1 was under PW. This occurred at all levels of the design variables. EM and MS consistently overestimated β_1, and PW and RS consistently overestimated β_1.

Coverage probabilities under all MDTs were above the criterion level of 95% with monotonic missing pattern for normal data, and decreased with increasing non-normality. However, MDTs differed in coverage probability only for β_1. In addition, the graphs for β_2, $\beta_3, \ldots, \beta_9$ were similar to those for β_1. For this reason, only the results for β_1 were reported. EM generally had the best coverage probability and MS the worst.

Study 4: Nine Predictors and High R^2 Condition

Effects on Absolute Error for $R^2_{estimate}$

The effects of design variables on absolute error for $R^2_{estimate}$ are listed in Table 20. Using $\eta^2_{alt} \geq .03$ as the criterion, the table shows that with a sample of 94, only the effect of non-normality was of practical significance under each MDT, with the largest effect being that of non-normality under pairwise deletion.

For sample size 153, the effect of percent missing on absolute error was of practical significance under EM, MS, and RS. With pairwise deletion, it was the effect of non-normality that was of practical significance. The largest effect was that of percent missing under mean substitution.

For sample size 265, the effect of pattern was of practical significance under EM algorithm only. The effect of percent missing was of practical significance under EM, MS, and RS. As with sample of 153, the effect of non-normality was of practical significance only under PW. The largest effect was that of percent missing under EM.

Table 20
Effects on absolute error for $R^2_{estimate}$ fulfilling the $\eta^2_{alt} \geq .03$ criterion

Sample size	MDT	Effect	SS_{effect}	$SS_{residual}$	df	η^2_{alt}
94	EM	Non-normality	3.12	92.12	3	.033
	MS	Non-normality	3.65	76.41	3	.046
	PW	Non-normality	6.77	63.89	3	.096
	RS	Non-normality	4.26	90.91	3	.045
153	EM	% Missing	7.23	76.64	2	.086
	MS	% Missing	3.36	63.42	2	.420
	PW	Non-normality	3.86	38.68	3	.091
	RS	% Missing	5.99	73.29	2	.076
265	EM	% Missing	10.34	59.40	2	.148
		Pattern	2.18	59.40	1	.035
	MS	% Missing	6.07	51.73	2	.105
	PW	Non-normality	2.24	22.22	3	.091
	RS	% Missing	9.28	57.49	2	.139

Note: EM = EM imputation, MS = Mean substitution, PW = Pairwise deletion, RS = Regression imputation.

Relative Performance of MDTs in the Estimation of Population R^2

Table 21 contains cell means and standard deviations of mean absolute errors for $R^2_{estimate}$ under various design conditions. The influence of non-normality on the performance of MDTs at sample size 94 is revealed in Figure 35.

Although the performance of all MDTs deteriorated with increasing non-normality, this deterioration was most pronounced after skew = 1.8 and kurtosis = 6. This was observed for both monotonic and non-monotonic missing pattern and for all levels of percent missing.

The relative performance of MDTs was undifferentiated at the lowest level of percent missing (10%), under both monotonic and non-monotonic missing patterns. However, at higher levels of percent missing (15% and 20%), PW yielded the best estimates and EM the worst for normal data, a trend that continued up to skew = 1.8 and kurtosis = 6. At the extreme level of non-normality (skew = 3 and kurtosis = 25), differentiation of performance of MDTs was least. Whereas mean absolute error for $R^2_{estimate}$ increased with percent missing for EM, MS, and RS at each level of non-normality, the magnitude of the error under PW at each level of non-normality changed marginally for both missing patterns as well as under all conditions of percent missing. Figures 36 and 37 show the influence of non-normality on the performance of MDTs at sample sizes 153 and 265, respectively. The results were similar to those under sample size 94.

In sum, the four MDTs differed most in their performance as percent missing increased, though their performance was most differentiated for normal data. MDTs were also differentiated more under non-monotonic missing pattern. PW had the best performance and EM the worst.

Table A.7 in Appendix A contains bias in $R^2_{estimate}$ under high R^2 condition with nine predictors. Bias in $R^2_{estimate}$ varied from –.13 to .07 across all sample size conditions, giving a range of .20. The values in Table A.7 were plotted in Figure B.7, a typical set of graphs for bias in $R^2_{estimate}$. Although there was no single MDT under which the bias was smallest at all levels of the design variables, PW generally had the smallest bias. Whereas PW consistently overestimated R^2, EM, MS, and RS tended to underestimateR^2.

Table 21
Means and standard deviations* of absolute error for $R^2_{estimate}$

			Monotonic Missing Pattern				Non-monotonic Missing Pattern			
N	Percent Missing	Normality (sk., kurt.)	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation
94	10	(0,0)	0520(042)	0524(041)	0545(039)	0525(041)	0547(043)	0535(040)	0535(039)	0538(041)
		(1,3)	0583(045)	0593(044)	0603(042)	0591(044)	0583(046)	0566(043)	0581(041)	0573(045)
		(1.8,6)	0665(049)	0683(050)	0698(049)	0678(049)	0634(049)	0625(047)	0665(047)	0631(048)
		(3,25)	0963(070)	0975(069)	0998(070)	0958(069)	0887(070)	0867(062)	0951(069)	0897(073)
	15	(0,0)	0668(051)	0627(049)	0529(038)	0641(050)	0751(051)	0618(047)	0522(038)	0679(049)
		(1,3)	0688(055)	0666(053)	0579(042)	0678(054)	0779(057)	0632(049)	0568(041)	0713(054)
		(1.8,6)	0731(058)	0725(056)	0674(048)	0732(057)	0779(059)	0667(051)	0652(046)	0748(057)
		(3,25)	0966(075)	0992(073)	0963(069)	0985(074)	1022(081)	0894(066)	0944(074)	1020(082)
	20	(0,0)	0878(043)	0775(056)	0520(039)	0825(058)	1005(059)	0763(054)	0514(038)	0913(058)
		(1,3)	0871(063)	0781(060)	0563(042)	0829(062)	1014(066)	0750(056)	0554(042)	0936(064)
		(1.8,6)	0876(066)	0807(062)	0653(048)	0847(064)	1011(069)	0756(057)	0648(048)	0952(067)
		(3,25)	1053(081)	1026(076)	0938(068)	1052(079)	1223(093)	0949(071)	0964(081)	1236(099)
153	10	(0,0)	0502(039)	0483(038)	0409(031)	0490(039)	0567(040)	0478(036)	0401(031)	0508(038)
		(1,3)	0527(041)	0515(040)	0451(033)	0521(041)	0607(042)	0497(037)	0441(033)	0544(040)
		(1.8,6)	0567(044)	0565(043)	0523(038)	0569(043)	0637(045)	0534(039)	0510(037)	0584(042)
		(3,25)	0753(059)	0762(058)	0754(055)	0764(058)	0813(064)	0686(053)	0725(056)	0781(062)
	15	(0,0)	0684(046)	0621(045)	0401(031)	0644(046)	0801(046)	0620(043)	0392(030)	0705(045)
		(1,3)	0692(050)	0633(048)	0441(033)	0659(049)	0827(051)	0628(044)	0432(033)	0741(049)
		(1.8,6)	0705(053)	0655(051)	0513(037)	0680(052)	0836(055)	0641(046)	0499(037)	0760(052)
		(3,25)	0828(065)	0806(062)	0742(055)	0822(064)	1003(075)	0779(061)	0707(055)	0971(074)
	20	(0,0)	0951(052)	0843(052)	0398(030)	0886(052)	1094(050)	0829(050)	0387(031)	0980(050)
		(1,3)	0937(058)	0833(057)	0436(033)	0878(058)	1113(057)	0821(053)	0428(033)	1014(056)
		(1.8,6)	0928(063)	0832(060)	0505(037)	0879(061)	1101(064)	0814(056)	0499(037)	1017(062)
		(3,25)	0979(077)	0913(070)	0727(054)	0956(074)	1254(088)	0930(070)	0704(058)	1225(087)
265	10	(0,0)	0481(033)	0448(032)	0311(023)	0457(032)	0588(035)	0459(033)	0306(023)	0503(034)
		(1,3)	0489(035)	0456(035)	0335(025)	0466(035)	0617(039)	0470(036)	0330(025)	0535(037)
		(1.8,6)	0507(039)	0480(038)	0385(029)	0490(038)	0628(044)	0494(038)	0381(029)	0557(041)
		(3,25)	0601(049)	0594(047)	0572(043)	0602(048)	0790(059)	0621(047)	0557(041)	0743(056)
	15	(0,0)	0733(039)	0657(039)	0307(023)	0680(039)	0872(040)	0666(040)	0301(023)	0767(040)
		(1,3)	0730(043)	0654(043)	0331(025)	0679(043)	0904(045)	0664(043)	0326(025)	0802(044)
		(1.8,6)	0724(048)	0655(047)	0380(029)	0680(047)	0901(051)	0665(046)	0375(028)	0809(049)
		(3,25)	0743(059)	0703(055)	0558(042)	0724(057)	1049(072)	0769(059)	0542(040)	1003(070)
	20	(0,0)	1008(041)	0899(042)	0304(023)	0940(042)	1156(041)	0895(043)	0302(023)	1047(042)
		(1,3)	0995(047)	0884(047)	0327(025)	0928(047)	1189(048)	0896(048)	0324(025)	1095(048)
		(1.8,6)	0978(052)	0868(052)	0375(028)	0915(053)	1187(055)	0887(052)	0370(028)	1103(054)
		(3,25)	0930(069)	0839(065)	0548(041)	0887(067)	1324(080)	0965(068)	0533(042)	1288(079)

* All values in the table are preceded by an omitted decimal point. Standard deviations are in parentheses. Standard errors of the means can be obtained by dividing each standard deviation by $1000^{1/2}$.

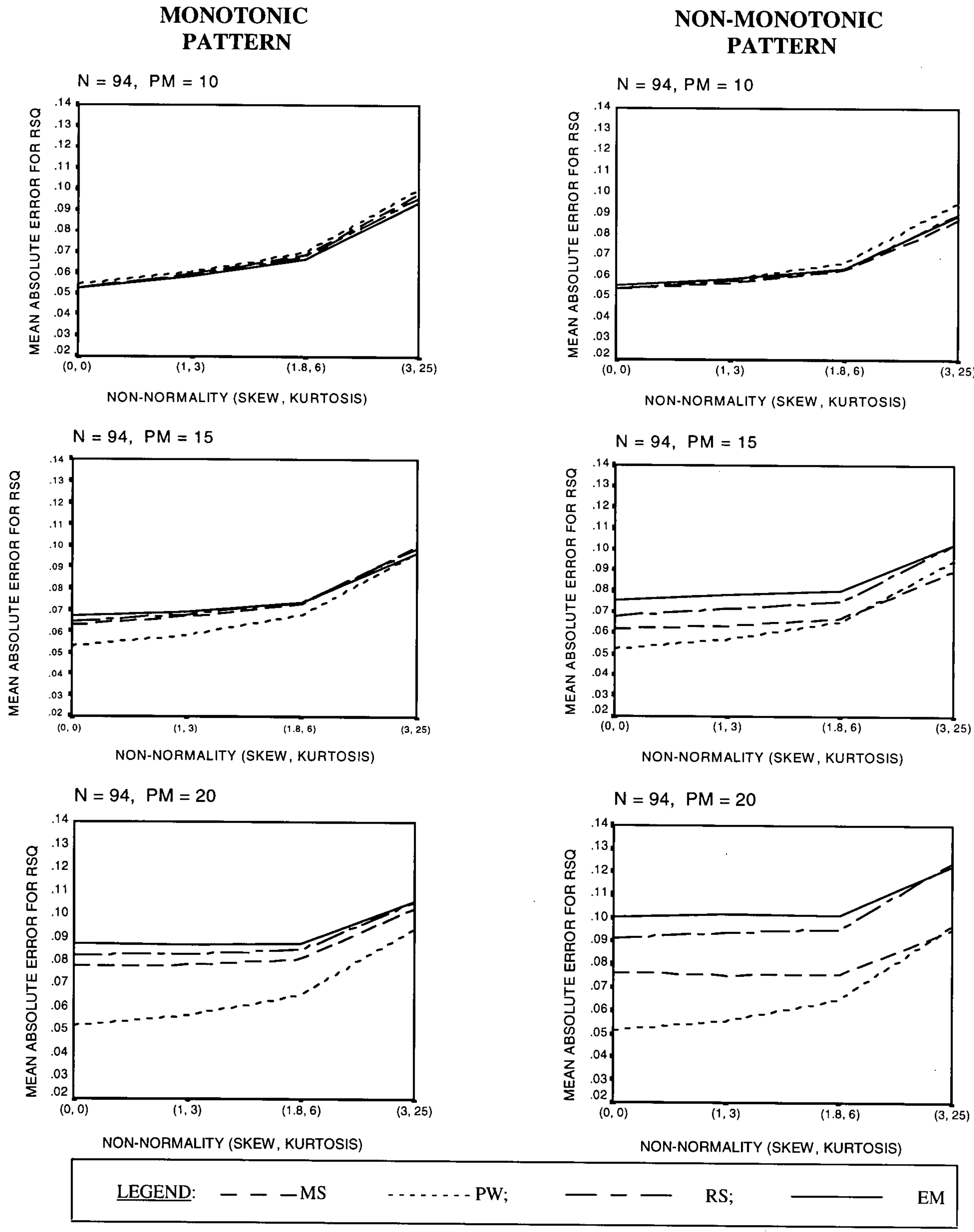


Figure 35: Mean absolute error for $R^2_{estimate}$ across levels of non-normality (N = 94)
Note: N = Sample size, PM = Percent missing, RSQ = $R^2_{estimate}$

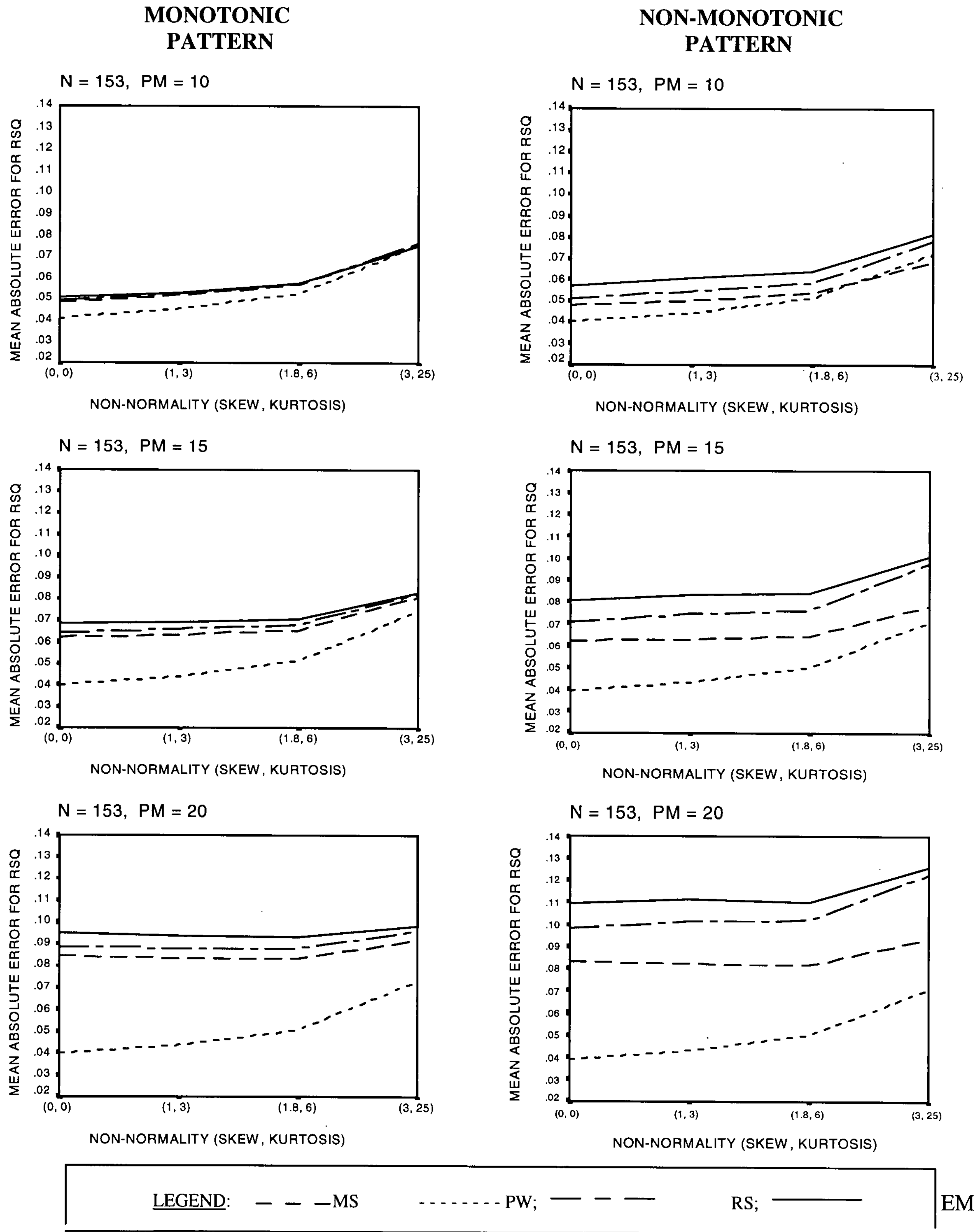


Figure 36: Mean absolute error for $R^2_{estimate}$ across levels of non-normality (N = 153)
Note: N = Sample size, PM = Percent missing, RSQ = $R^2_{estimate}$

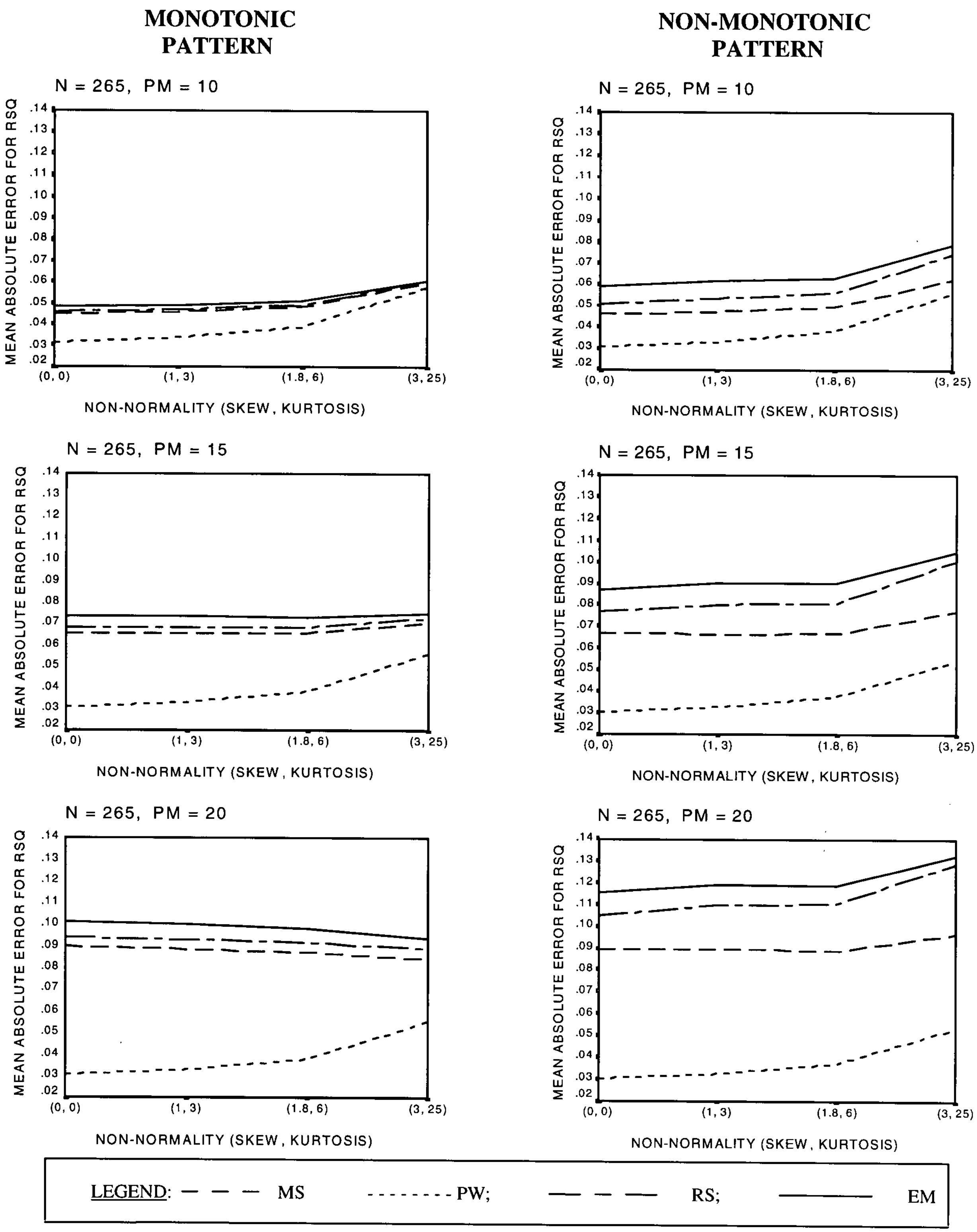


Figure 37: Mean absolute error for $R^2_{estimate}$ across levels of non-normality (N = 265)
Note: N = Sample size, PM = Percent missing, RSQ = $R^2_{estimate}$

Effects on Absolute Error for Regression Coefficients

Table 22 shows the effects of design variables on absolute error in b_1, b_2,…,b_9. The effect of missing pattern was strongest for b_1, the regression coefficient of the predictor with no missing data. More specifically, under EM, the effect of missing pattern on absolute error for b_1 was of practical significance at all levels of sample size. Under MS and PW, pattern effects were significant only for sample size 265. There were no effects of practical significance for b_5.

Under MS and PW, the effect of non-normality on absolute error of b_1 was of practical significance under sample size 265 only. However, under RS, the effect of non-normality was of practical significance at all levels of sample size. The effect of non-normality on absolute error for b_1 under EM was not of practical significance. Only non-normality had an effect of practical significance for b_2 to b_9.

Table 22
Effects on absolute error for regression coefficients fulfilling the criterion of $\eta^2_{alt} \geq .03$*

Sample size	MDT	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9
94	EM	PT(042)	NL(063)	NL(038)	NL(050)	-	NL(068)	NL(042)	NL(045)	-
	MS	-	NL(061)	NL(032)	NL(048)	-	NL(073)	NL(039)	NL(045)	-
	PW	-	NL(060)	NL(032)	NL(047)	-	NL(074)	NL(039)	NL(043)	-
	RS	NL(038)	NL(066)	NL(034)	NL(051)	-	NL(071)	NL(038)	NL(047)	-
153	EM	PT(182)	NL(057)	NL(036)	NL(054)	-	NL(087)	NL(056)	NL(056)	-
	MS	-	NL(056)	NL(031)	NL(056)	-	NL(089)	NL(050)	NL(061)	-
	PW	-	NL(056)	NL(030)	NL(054)	-	NL(089)	NL(050)	NL(060)	-
	RS	PT(058)	-	-	-	-	-	-	-	-
		NL(054)	NL(061)	NL(033)	NL(058)	-	NL(089)	NL(048)	NL(057)	-
265	EM	PT(382)	-	-	-	-	-	-	-	-
		-	NL(082)	NL(047)	NL(071)	-	NL(098)	NL(069)	NL(073)	NL(039)
	MS	PT(050)	-	-	-	-	-	-	-	-
		NL(031)	NL(083)	NL(041)	NL(063)	-	NL(105)	NL(059)	NL(077)	NL(035)
	PW	PT(050)	-	-	-	-	-	-	-	-
		NL(031)	NL(084)	NL(041)	NL(063)	-	NL(105)	NL(060)	NL(077)	NL(034)
	RS	PT(182)	-	-	-	-	-	-	-	-
		NL(088)	NL(089)	NL(045)	NL(071)	-	NL(100)	NL(058)	NL(075)	NL(038)

* Values of η^2_{alt} are in parentheses preceded by an omitted decimal point.
Note: EM = EM imputation, MS = Mean substitution, PW = Pairwise deletion, RS = Regression imputation, PT = Pattern, NL = Non-normality, - = No effect of practical significance.

Relative Performance of MDTs in Estimation of β_1

Table 23 contains means and standard deviations of absolute error for in b_1 at various design conditions. The means at sample size 94 are plotted in Figure 38. The figure shows that mean absolute error for b_1 increased with non-normality. RS performed the best, regardless of missing pattern. However, under monotonic pattern of missing data, MDTs were least differentiated in performance at 10% missing, the degree of differentiation increasing with percent missing. RS consistently gave the best performance, followed by EM.

With non-monotonic pattern, again RS outperformed the other MDTs, although its performance relative to EM was most prominent with normal data. At the extreme level of non-normality, little difference in performance was observed between RS and EM. Whereas EM consistently outperformed both MS and PW at 10% missing under non-monotonic missing pattern, MS and PW had a slight advantage over EM at 15% and 20% missing at low to moderate levels of non-normality. MDTs were more differentiated under non-monotonic pattern than under monotonic missing pattern.

Figure 39 shows mean absolute error for b_1 at sample size 153 across levels of non-normality. Under monotonic pattern, the MDTs differed more in performance as percent missing increased. Unlike the finding at sample size 94, where MS and PW performed poorer than EM with monotonic pattern, little differentiation in performance was observed between MS, PW and EM at sample size 153 under low to moderate percent missing. At 20% missing, MS and PW did better than EM at all levels of non-normality. Under non-monotonic pattern, RS performed the best, followed by MS and PW. EM performed the worst. The MDTs differed most in performance with normal data, the differentiation becoming more blurred with increasing non-normality. At the most extreme non-normality and 20% missing, the performance of all MDTs was least differentiated.

Graphical plots of mean absolute error in b_1 at sample size 265 are displayed in Figure 40, showing that under monotonic pattern, RS was best, followed by MS and PW. EM performed the worst. Under non-monotonic pattern, however, RS only excelled with normal data at ten percent missing. The performance of MS and PW was best. However, the relative performance of MS and RS as well as PW and RS was least differentiated with normal data, and most differentiated at the highest level of non-normality (skew = 3, kurtosis = 25).

In sum, RS performed the best under monotonic missing pattern at sample size of 265. Under non-monotonic pattern, MS and PW outperform RS at moderate to high levels of percent missing (15% to 20%).

Table A.8 shows the bias in b_1 under nine predictors and low R^2 condition. Figure B.8 is the corresponding set of graphs for bias in b_1 for N = 94, showing that for non-monotonic pattern, the smallest bias across all levels of the design variables was under MS and PW. The bias in b_1 varied from –.10 to .07 across all design conditions, giving a range of .17. Whereas MS and PW consistently overestimated β_1, EM and RS consistently underestimated β_1. The bias for b_2 to b_9 were much smaller compared to that for b_1, with a maximum range of .10 under b_7. Therefore, the graphs for bias in b_2 to b_9 were omitted.

Table 23
Means and standard deviations* of absolute error for b_1

			Monotonic Missing Pattern				Non-monotonic Missing Pattern			
N	Percent Missing	Normality (sk, kurt.)	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation
94	10	(0,0)	0573(041)	0625(046)	0625(046)	0579(043)	0682(040)	0726(053)	0727(053)	0587(041)
		(1,3)	0639(046)	0669(052)	0670(052)	0631(047)	0761(039)	0796(063)	0797(063)	0663(041)
		(1.8,6)	0696(052)	0728(058)	0730(058)	0687(052)	0793(040)	0863(074)	0865(074)	0717(044)
		(3,25)	0784(071)	0816(077)	0817(078)	0770(068)	0932(043)	1004(104)	1006(104)	0929(057)
	15	(0,0)	0591(040)	0654(049)	0654(049)	0575(043)	0835(038)	0791(059)	0794(059)	0605(038)
		(1,3)	0647(044)	0693(054)	0696(054)	0618(046)	0891(038)	0863(070)	0867(070)	0699(037)
		(1.8,6)	0699(051)	0747(060)	0752(060)	0664(051)	0903(038)	0928(082)	0931(082)	0740(040)
		(3,25)	0797(068)	0833(080)	0840(080)	0747(064)	0992(038)	1101(115)	1106(115)	0970(050)
	20	(0,0)	0609(040)	0677(051)	0679(052)	0567(043)	0865(035)	0834(062)	0841(062)	0686(038)
		(1,3)	0653(043)	0709(055)	0716(056)	0609(045)	0915(036)	0893(073)	0901(073)	0781(039)
		(1.8,6)	0701(048)	0760(061)	0767(062)	0660(049)	0912(037)	0953(085)	0961(085)	0806(041)
		(3,25)	0783(063)	0844(083)	0854(084)	0741(063)	0981(039)	1136(118)	1146(118)	1009(055)
153	10	(0,0)	0477(033)	0496(036)	0496(036)	0452(033)	0794(035)	0580(044)	0581(044)	0498(032)
		(1,3)	0503(035)	0519(038)	0519(038)	0476(035)	0873(034)	0644(051)	0645(051)	0582(034)
		(1.8,6)	0550(039)	0564(043)	0565(043)	0521(038)	0882(036)	0709(058)	0711(058)	0624(035)
		(3,25)	0639(053)	0634(056)	0634(056)	0593(051)	0989(036)	0871(086)	0872(087)	0864(042)
	15	(0,0)	0505(034)	0509(037)	0510(037)	0452(033)	0855(031)	0621(047)	0622(047)	0564(033)
		(1,3)	0531(036)	0534(039)	0535(040)	0478(035)	0917(031)	0687(053)	0687(053)	0656(034)
		(1.8,6)	0574(039)	0578(044)	0579(044)	0524(039)	0915(033)	0741(061)	0743(061)	0684(035)
		(3,25)	0661(052)	0657(058)	0657(058)	0608(050)	0991(033)	0921(090)	0922(090)	0909(043)
	20	(0,0)	0553(035)	0517(039)	0520(039)	0450(034)	0922(029)	0672(050)	0675(050)	0625(031)
		(1,3)	0576(036)	0540(041)	0542(042)	0476(034)	0965(028)	0735(056)	0737(056)	0716(033)
		(1.8,6)	0616(038)	0582(046)	0588(046)	0519(038)	0956(031)	0790(064)	0792(064)	0737(036)
		(3,25)	0689(051)	0661(060)	0655(060)	0600(049)	0992(034)	0974(093)	0976(094)	0939(043)
265	10	(0,0)	0370(027)	0350(027)	0350(027)	0323(025)	0792(028)	0472(034)	0473(034)	0424(027)
		(1,3)	0391(028)	0379(029)	0379(029)	0346(026)	0881(028)	0532(041)	0473(034)	0527(029)
		(1.8,6)	0425(030)	0415(033)	0415(033)	0383(029)	0890(030)	0572(047)	0574(047)	0562(030)
		(3,25)	0508(038)	0508(043)	0510(043)	0470(037)	1002(030)	0729(047)	0730(066)	0815(035)
	15	(0,0)	0428(028)	0367(028)	0368(028)	0333(025)	0868(025)	0533(037)	0534(037)	0522(027)
		(1,3)	0437(029)	0395(031)	0395(031)	0355(026)	0929(025)	0587(045)	0587(045)	0627(028)
		(1.8,6)	0460(031)	0435(034)	0435(034)	0387(029)	0930(027)	0625(051)	0627(051)	0653(030)
		(3,25)	0534(037)	0527(044)	0528(044)	0470(037)	1001(029)	0793(071)	0793(071)	0886(035)
	20	(0,0)	0481(029)	0372(029)	0373(029)	0334(025)	0913(022)	0573(040)	0574(040)	0604(025)
		(1,3)	0488(030)	0400(031)	0401(031)	0357(026)	0958(023)	0628(047)	0630(047)	0704(026)
		(1.8,6)	0506(032)	0443(035)	0443(035)	0390(028)	0953(025)	0664(053)	0666(053)	0722(029)
		(3,25)	0574(036)	0535(045)	0537(045)	0475(035)	0997(028)	0834(073)	0835(073)	0922(035)

* All values in the table are preceded by an omitted decimal point. Standard deviations are in parentheses. Standard errors of the means can be obtained by dividing each standard deviation by $1000^{½}$.

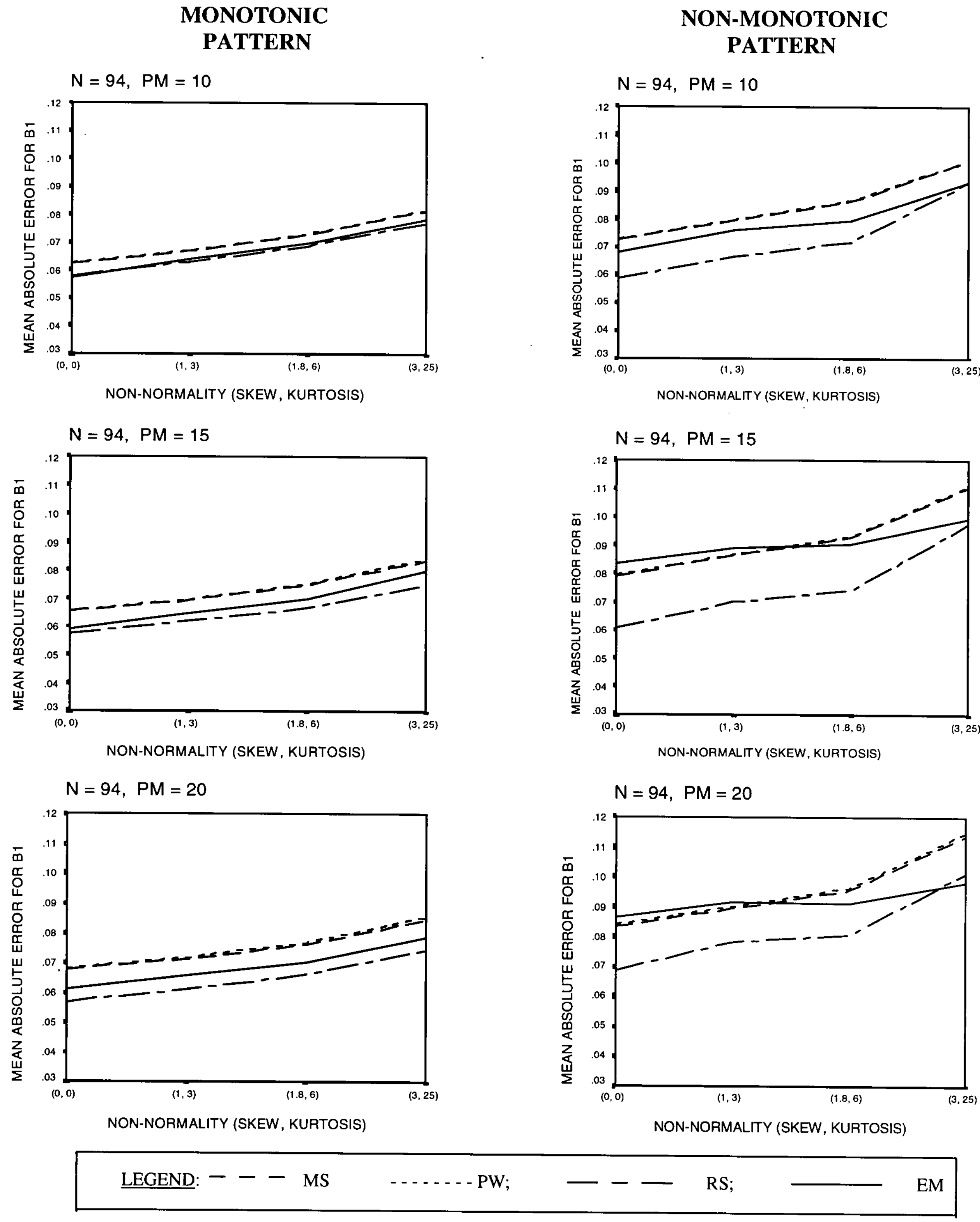


Figure 38: Mean absolute error for b_1 across levels of non-normality (N = 94)
Note: N = Sample size, PM = Percent missing.

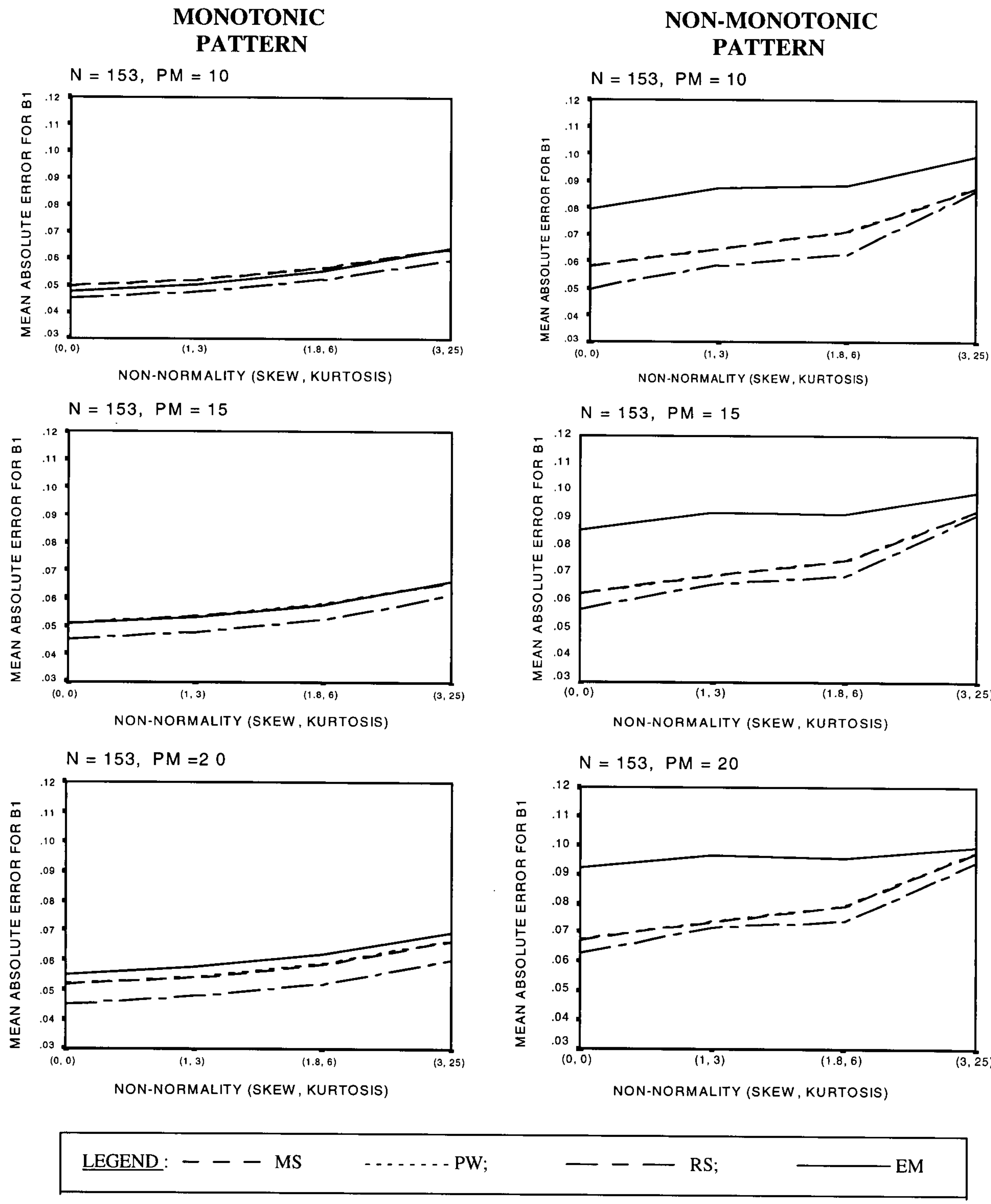


Figure 39: Mean absolute error for b_1 across levels of non-normality (N = 153)
Note: N = Sample size, PM = Percent missing.

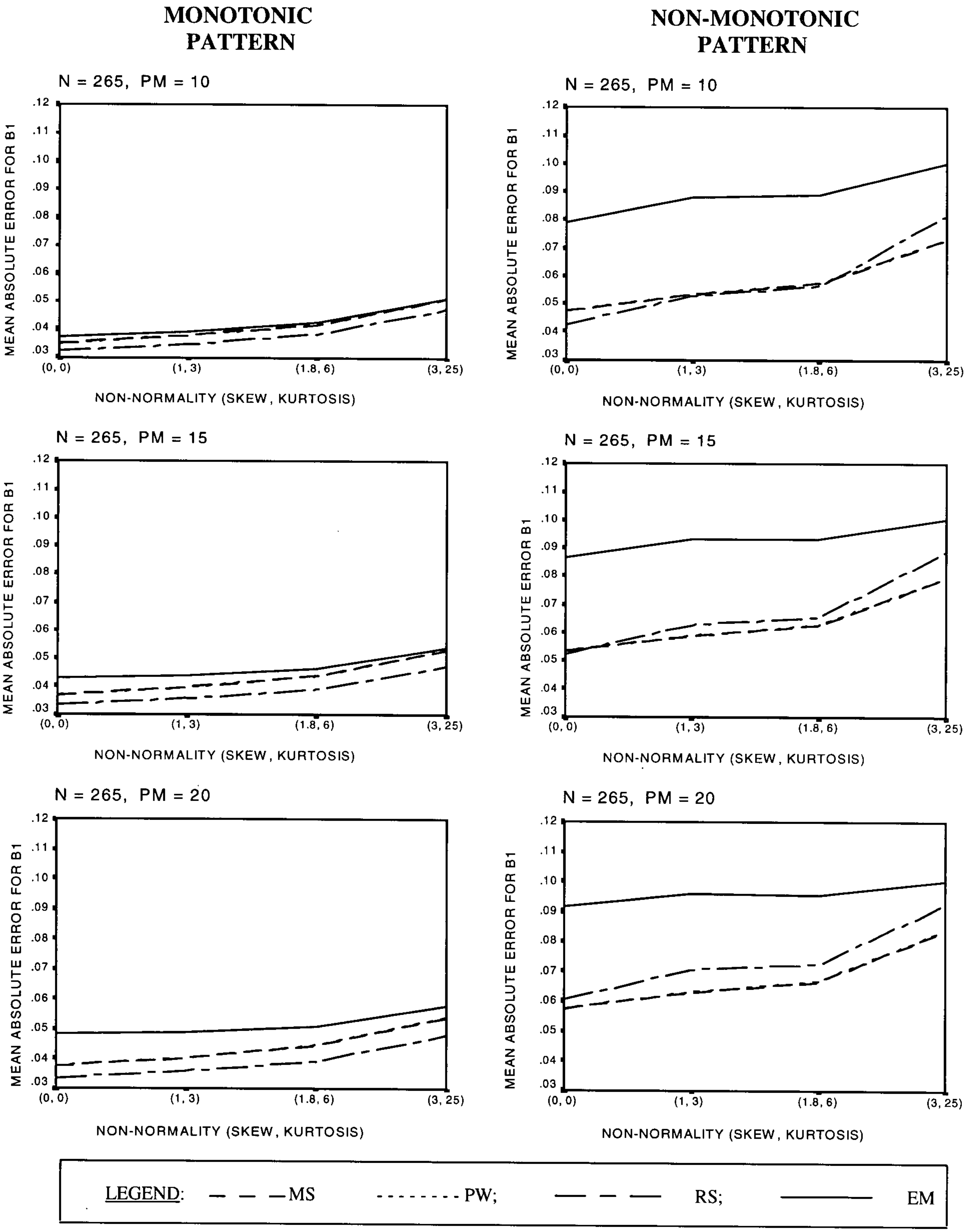


Figure 40: Mean absolute error for b_1 across levels of non-normality (N = 265)
Note: N = Sample size, PM = Percent missing.

<u>Coverage Probability for β_1</u>

The MDTs differed in their coverage probabilities for β_1 only. The graphs for coverage probability of $\beta_1, \beta_2, \beta_3, \ldots, \beta_9$ were similar to those of β_1. For this reason, only the findings for β_1 are reported.

Table 24 contains coverage probabilities at various levels of the design variables. Figure 41 is a set of line graphs at sample size 94. The graphs show that coverage probability decreased with increasing non-normality. Coverage probabilities under all MDTs were above the criterion level of 95% for normal data. This suggests that all MDTs worked well for normal data irrespective of missing pattern or percent missing. MDTs were most differentiated under non-monotonic pattern. EM and RS were the best and MS the worst.

Figure 42 shows coverage probabilities for β_1 at sample size 153, indicating that coverage probabilities decreased with increasing non-normality. The relative performance of MDTs was most differentiated under non-monotonic than under monotonic missing pattern. RS had the best coverage probabilities and MS the worst.

Figure 43 is a set of graphical plots of mean coverage probabilities for β_1 at sample size 265, showing that coverage probability decreased with increasing non-normality. The graphs also show that the MDTs were most differentiated under non-monotonic than under monotonic missing pattern. All MDTs performed close to the criterion level of 95% for normal data under monotonic pattern.

Under non-monotonic pattern, all MDTs had coverage probabilities below the expected value. At ten percent missing, RS had coverage probability closest to the expected value, except at the highest level of non-normality (skew = 3 and kurtosis = 25). At this point, PW had the best coverage probability. This outcome was also observed at fifteen percent missing. At 20% missing, PW had the best coverage probability and EM the worst (below 65%).

Table 24
Mean coverage probability (%) for β_1

			Monotonic Missing Pattern				Non-monotonic Missing Pattern			
N	Percent Missing	Normality (sk, kurt.)	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation
94	10	(0,0)	97.9	95.9	97.1	97.4	98.6	89.5	92.2	97.2
		(1,3)	96.3	93.4	95.3	95.6	96.5	86.6	89.3	96.2
		(1.8,6)	93.8	90.7	92.6	93.6	94.3	82.9	86.1	93.5
		(3,25)	89.2	87.2	89.4	89.9	85.3	76.6	79.3	84.4
	15	(0,0)	98.7	95.7	97.2	98.4	98.8	86.1	91.1	98.7
		(1,3)	97.2	93.5	96.2	96.7	96.3	83.9	88.3	97.4
		(1.8,6)	95.8	91.0	94.0	94.9	94.8	81.6	85.0	95.9
		(3,25)	90.6	86.8	90.0	91.2	85.9	75.1	79.4	84.5
	20	(0,0)	99.3	95.6	97.6	98.4	99.3	85.7	91.8	98.6
		(1,3)	97.9	93.0	96.2	97.3	98.2	83.0	89.2	97.8
		(1.8,6)	97.3	90.9	94.4	96.0	96.1	81.2	87.2	96.1
		(3,25)	92.5	86.5	91.2	92.5	88.5	74.9	81.7	85.7
153	10	(0,0)	97.9	96.4	97.2	97.3	89.0	88.5	91.0	97.5
		(1,3)	96.5	94.5	95.4	96.3	82.2	84.4	86.9	94.4
		(1.8,6)	93.8	90.6	93.2	93.9	79.9	81.5	83.5	91.7
		(3,25)	87.5	85.2	87.9	88.4	64.5	73.3	76.1	73.6
	15	(0,0)	98.2	95.6	97.9	98.0	90.1	85.9	90.3	96.8
		(1,3)	96.4	94.4	96.7	97.0	85.1	81.7	86.7	93.7
		(1.8,6)	93.9	91.0	93.8	95.3	81.7	79.0	83.8	90.6
		(3,25)	87.4	85.0	89.8	89.3	70.1	72.4	76.4	71.9
	20	(0,0)	98.2	95.2	98.1	98.4	90.8	84.8	90.5	97.2
		(1,3)	96.3	93.2	97.1	97.2	85.2	80.6	87.3	93.5
		(1.8,6)	94.6	90.5	95.2	95.2	81.5	77.4	83.8	91.0
		(3,25)	88.6	85.7	91.9	90.6	72.8	70.5	77.3	74.4
265	10	(0,0)	96.4	95.7	97.0	97.5	67.5	87.0	89.1	93.3
		(1,3)	95.2	93.7	95.4	96.1	53.9	81.4	84.2	87.3
		(1.8,6)	92.8	91.3	92.9	93.8	51.6	77.3	79.7	81.8
		(3,25)	84.6	84.1	87.1	87.0	38.6	68.1	70.3	56.2
	15	(0,0)	96.1	95.6	97.6	97.6	61.8	81.4	87.1	91.3
		(1,3)	94.3	94.0	96.3	97.2	51.6	77.1	81.4	82.2
		(1.8,6)	91.1	90.9	94.7	94.4	50.2	73.8	78.1	77.2
		(3,25)	84.0	84.0	88.7	87.0	41.8	65.4	67.9	53.8
	20	(0,0)	95.3	95.7	98.3	98.1	59.4	80.1	87.3	89.1
		(1,3)	92.6	93.4	97.1	97.6	50.3	74.7	82.0	79.1
		(1.8,6)	90.8	91.0	95.2	95.4	51.3	71.2	78.7	74.5
		(3,25)	83.8	95.4	90.2	89.4	47.1	64.1	69.5	53.4

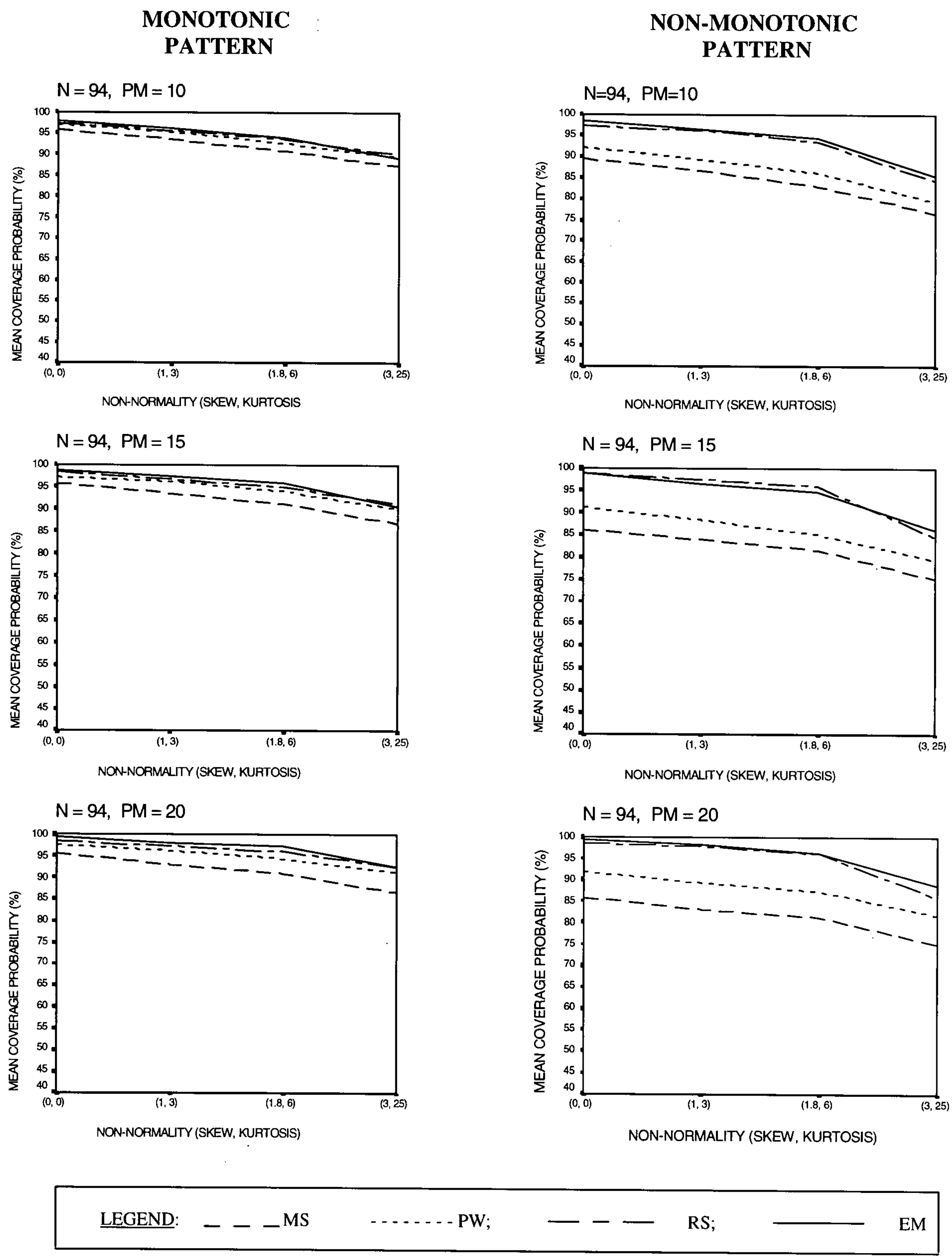


Figure 41: Mean coverage probability for β_1 (N = 94)
Note: N = Sample size and PM = Percent Missing.

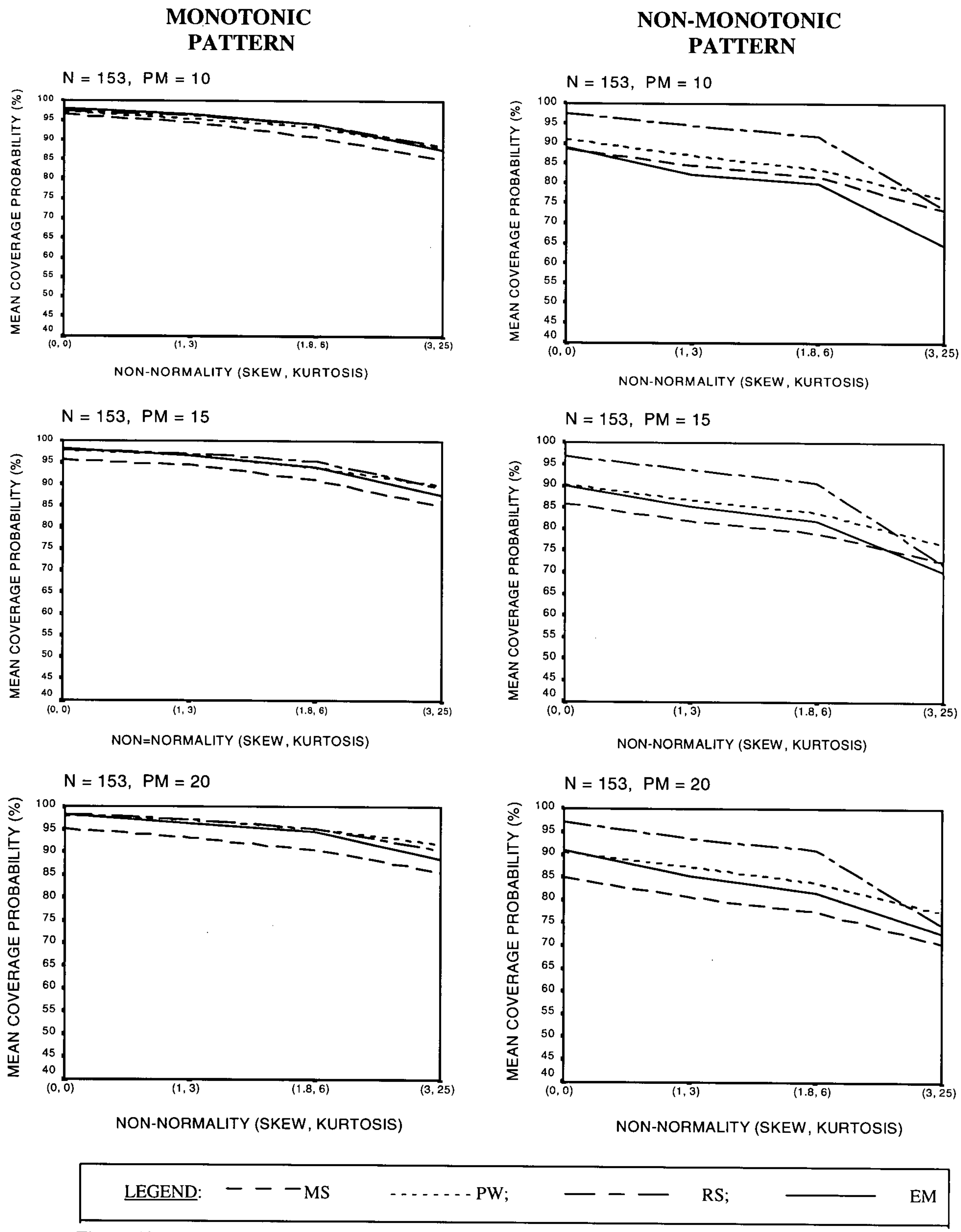


Figure 42: Mean coverage probability (N = 153)
Note: N = Sample size and PM = Percent Missing

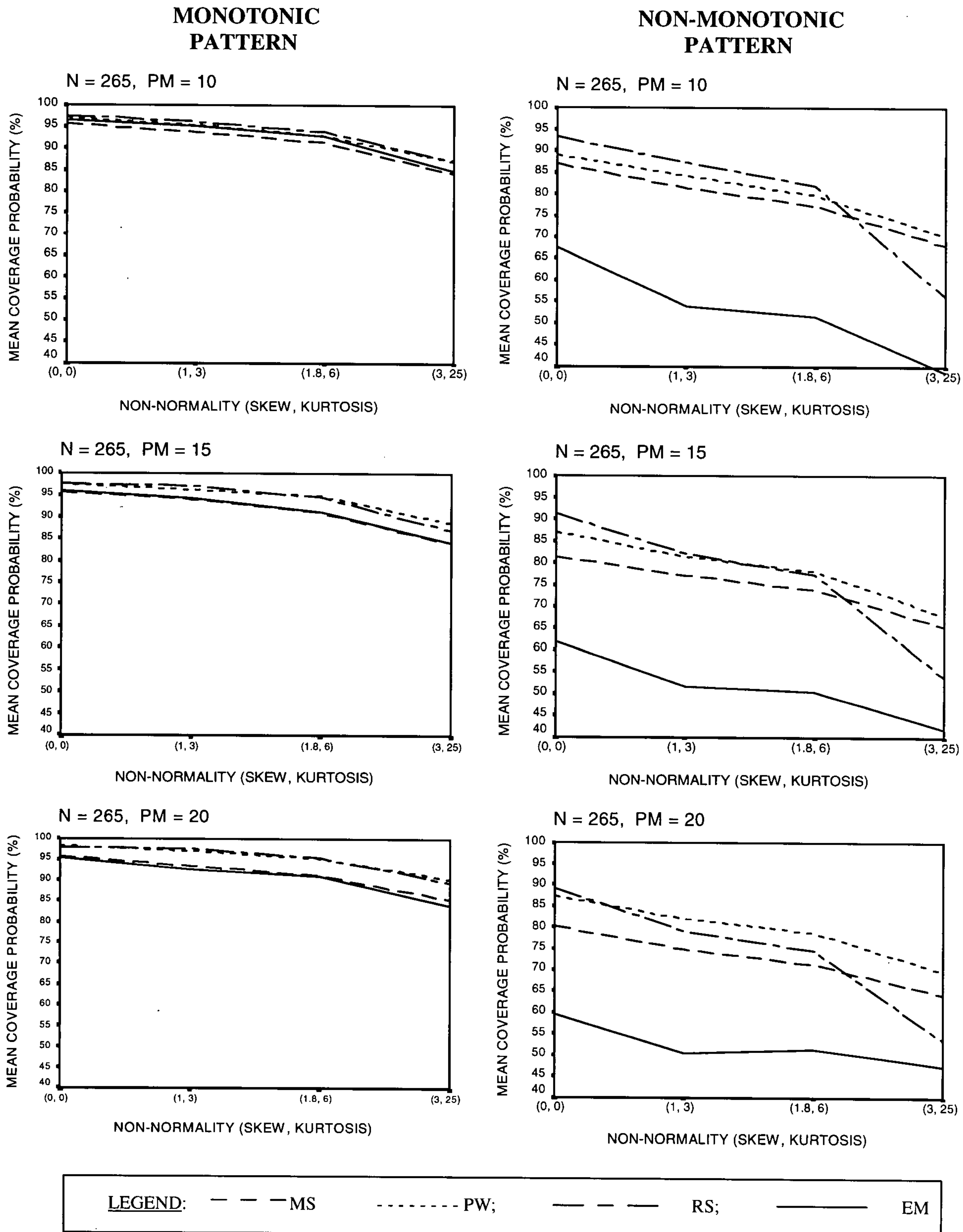


Figure 43: Mean coverage probability (N = 265)
Note: N = Sample size and PM = Percent Missing

Summary

In the previous sections, the effects of pattern of missing data, percent missing and non-normality on absolute error for parameter estimates were reported. The effects of non-normality on absolute error for $R^2_{estimate}$ were the most prominent across MDTs at sample size 94. However, at sample size 153, the effect of percent missing was the most dominant across MDTs. The effect of missing pattern only showed at sample size 265 under EM.

In the estimation of population R^2, MDTs differed marginally when 10% of the data was missing, regardless of missing pattern. MDTs were differentiated more under non-monotonic than under monotonic missing pattern. In addition, MDTs were most differentiated for normal data, and also when percent missing was highest. Overall, the performance of PW was the best and EM the worst.

The effects of design factors on regression coefficients were also reported in the previous sections. Of all factors, the effect of missing pattern on absolute error for b_1 was the strongest. For the remaining regression coefficients, except for b_5, it was the effect of non-normality that was of practical significance, with the strongest effect being at sample size 265.

With respect to mean absolute error for regression coefficients, the MDTs differed most in the estimation of β_1, and least in the estimation of β_2, β_3,...β_9. In general, RS performed the best with respect to accuracy. However, with non-monotonic missing pattern at sample size 265, and 15% to 20% missing, MS and PW tended to outperform the other MDTs. Also, MDTs were more differentiated under non-monotonic pattern than under monotonic pattern of missing data.

The smallest bias in R^2 at different levels of the design variables was not under a single MDT. However, the bias under PW was in most cases the smallest. Whereas PW consistently overestimated R^2, EM, MS, and RS tended to underestimate R^2. For β_1, the smallest bias was under MS and PW with non-monotonic pattern. MS and PW consistently overestimated β_1,and EM and RS consistently underestimated β_1.

With respect to coverage probability of β_1, all MDTs performed well under monotonic missing pattern for normal data. Coverage probabilities decreased with increasing non-normality. Both EM and RS had similar coverage probabilities at sample size 94. At sample size 153, the performance of RS was the best and MS the worst. For non-normal data under non-monotonic missing pattern at sample size 265, all MDTs generally had poor coverage probabilities, the worst being under EM at extreme non-normality.

CHAPTER 5
DISCUSSION

In the present study, the relative performance of MS (mean substitution), PW (pairwise deletion), RS (regression imputation) and EM (expectation-maximization) methods was investigated in four studies under different conditions of sample size, missing pattern, percent missing and non-normality. Four similar studies were conducted using multiple regression models. Study 1 had four predictors with multiple R^2 of .19, Study 2 four predictors with multiple R^2 of .59, Study 3 nine predictors with multiple R^2 of .21, and Study 4 nine predictors with R^2 of .58. What follows in this final chapter is a comparative discussion of findings from the four studies in relation to existing literature.

Effects on Absolute Error of Parameter Estimates

The effects of missing pattern, percent missing and non-normality on absolute error in $R^2_{estimate}$ were of practical significance upon treatment of missing values with MS, PW, RS and EM. As expected, there were sizeable effects for missing pattern, percent missing and non-normality. Mean absolute error was generally smaller under monotonic pattern, and smallest at 10% missing. In addition, mean absolute error was smallest for normal data.

When the number of predictors was four and multiple R^2 was .19, percent missing seemed to be the most important factor related to absolute error for $R^2_{estimate}$. This was especially so under EM and RS. However, this effect was only of practical significance with large samples ($N = 265$).

When the number of predictors was four and multiple R^2 was .59, the effect of percent missing was of practical significance under MS at sample size 265, and under RS and EM at all levels of sample size. Whereas non-normality seemed to have an influence on absolute error mainly under PW, it was the ordinal interaction effect of missing pattern and non-normality that had effects of practical significance under RS and EM. Under RS and EM, mean absolute error for $R^2_{estimate}$ was smaller for monotonic pattern than for non-monotonic pattern. However, the difference in mean absolute error across patterns was much greater at higher levels of non-normality.

When the number of predictors was nine and multiple R^2 was .21, the effect of non-normality on absolute error for $R^2_{estimate}$ seemed to be the most important, and this was under all MDTs. At sample size 94 and 153, the effect of non-normality was strongest under MS. However, at sample size 265, the strongest effect was under PW.

Under multiple R^2 of about .58 with nine predictors, non-normality seemed to be an important factor only at sample size 94. This was with the exception of PW that was affected by non-normality at all levels of sample size. At sample sizes 153 and 265, percent missing appeared to be the most important factor. However, with EM, pattern effects dominated at sample size 265.

The above results show that percent missing was the most important factor when number of predictors was four. However, with nine predictors, non-normality became the most important factor. This was regardless of the strength of the criterion-predictor relationship. Missing pattern effect was the most important under EM.

The effect of non-normality on absolute error for regression coefficients appeared to dominate under all MDTs. Missing pattern only affected b_1, the coefficient of the variable with no missing data. The strength of these effects tended to get larger with increasing sample size.

Relative Performance of MDTs

Table 25 shows the best MDTs in the estimation of R^2 and β_1 with respect to maximizing accuracy as measured by the mean absolute error of estimation, as well as MDTs that provided coverage probability closest to the expected value of 95% for each study. Table 26 shows the least biased MDTs in the estimation of R^2 and β_1. The tables indicate that under four predictors and low R^2 condition, EM was the best with respect to accuracy in the estimation of population R^2. This finding also held true under nine predictors and low R^2 condition. However, there was no MDT that was uniformly least biased under the low R^2 condition.

Under high population R^2 (about .60), PW had the most accurate and least biased estimates, with a tendency to overestimate R^2. This result was consistent with that of Azen et al. (1989) who found that although EM was generally the preferred method, PW performed better than EM for systematically missing data when population R^2 was 0.5. The limitation with PW is the fact that it can yield correlations outside the acceptable range. In a similar manner, covariance matrices based on PW may not be positive definite. Since many analyses based on the covariance matrix, including multiple regression, require a positive-definite matrix, modifications may be necessary when this condition is not satisfied.

Table 25: Best MDTs in the estimation of parameters and coverage probability

	4 Predictors			9 Predictors		
	$R^2_{estimate}$	b_1	Coverage Probability	$R^2_{estimate}$	b_1	Coverage Probability
Low R^2	EM	RS	EM/RS	EM	RS	EM
High R^2	PW	RS	RS	PW	RS	RS

Table 26: Least biased MDTs in the estimation of parameters

	4 Predictors		9 Predictors	
	$R^2_{estimate}$	b_1	$R^2_{estimate}$	b_1
Low R^2	None	RS	None	PW
High R^2	PW	RS	PW	MS/PW

Kromrey and Hines (1994) reported that PW and RS performed better in most situations than MS in the estimation of R^2. The findings in the present study seem to be in agreement with that of Kromrey and Hines, but only as far as the relative accuracy of PW and MS is concerned. As for the relative accuracy of RS and MS, the present study showed that RS outperformed MS under the condition of low R^2. However, under the condition of high R^2, MS generally outperformed RS.

Generally, the mean absolute error for $R^2_{estimate}$ increased with increasing non-normality, a finding consistent with that of Graham, Hofer and MacKinnon (1996) who showed that estimates obtained for skewed data were somewhat less accurate than for normal data. MDTs also appeared to be more differentiated under non-monotonic than under monotonic pattern, suggesting that the choice of MDT may be more crucial when the missing pattern is non-monotonic. Also, MDTs were most differentiated at the highest level of percent missing (20%), suggesting that the choice of MDT was most important at this level. This finding is consistent with that of Kromrey and Hines (1994) and Raymond and Roberts (1987) who found that the differences among MDTs increased as the proportion of missing data increased, although Raymond and Roberts (1987) used randomly deleted data. Therefore, the choice of MDTs is less critical if the amount of missing data is small (about 10%). In the present study, EM had the best overall performance under low multiple R^2 condition (R^2 about .2), and PW had the best performance under high R^2 condition (R^2 about .6). However, under low R^2 condition, at sample size 265, and 15% to 20% missing data, PW and MS also performed well. No MDT consistently provided the smallest bias under the low R^2 condition under four predictors and nine predictors. Under high R^2 condition, PW consistently provided the smallest bias regardless of number of predictors. Whereas PW had the tendency to overestimate R^2, the other MDTs had the tendency to underestimate R^2.

In the estimation of regression coefficients, the greatest accuracy was under RS for each study. However, RS had the least biased estimates under four predictors only. With nine predictors, PW had the least biased estimates. The bias under MS with nine predictors and high R^2 condition was least differentiated from that of PW.

In the estimation of regression coefficients, the performance of all MDTs deteriorated with increasing non-normality, and this was regardless of missing pattern. This result may be due to the fact that MDTs are based on normal theory, and as such, their performance is expected to deteriorate with increasing non-normality. Also, the mean absolute error for b_1 tended to be smaller under monotonic pattern than under non-monotonic pattern. There was a tendency for MDTs to differ more at higher levels of percent missing than at lower levels. Of particular interest was the fact that the MDTs were most differentiated with respect to accuracy in the estimation of β_1, the regression coefficient of the predictor with no missing values.

Some researchers (e.g., Little, 1988; Muthen, Kaplan, & Hollis, 1987) have suggested the superiority of the EM method in the treatment of missing data. Allison (1987) reported that even with non-normal data, maximum-likelihood estimates should have reasonably good properties relative to competing estimators. Graham et al. (1996) found that estimates obtained with non-normal data using EM treatment were satisfactory. However, they were cautious to make no claim that the parameter estimates based on the EM method would be adequate when data were more non-normal than in their study. In their study, five variables were used, with univariate skew ranging from –.68 to 3.30 and univariate kurtosis ranging from –.25 to 13.11. In the present study, EM outperformed RS in the estimation of population R^2 under low R^2 (about .2). However, in the estimation of β_1, RS had more accurate estimates than EM regardless of size of correlation. RS also had the least bias with four predictors. Both RS and EM underestimated β_1, and this can be explained by the fact that the marginal distribution of the predictor is distorted because the imputations do not reflect variation in the distribution of the predictor given the criterion. With nine predictors under high R^2 condition, PW and MS had the least bias, both MDTs overestimating β_1.

One possible explanation for the overwhelmingly favoured EM method (see Beale & Little, 1975; Gleason & Staelin, 1975; Little, 1979), is that the proportions of missing data investigated in previous studies were often disturbingly high, ranging from about 10 percent to 40 percent. With 40 percent of the data missing, one would have to question seriously the appropriateness of conducting any analysis. The finding in the present study that the RS method performed about as well as the EM method is important in the sense that RS is less complex than EM method. With regard to comparisons of the accuracy of parameter estimates after different treatments of missing data, there is need to exercise some caution because different criteria were used in different studies.

MDTs and Coverage Probabilities

The relative effectiveness of MDTs in producing coverage probabilities that captured β_1 at alpha = .05 under various design conditions was investigated. Overall, coverage probabilities under RS were closest to the nominal value. However, EM too had coverage probabilities close to the nominal value under low R^2 condition. It was also found that MDTs produced the best coverage probabilities with normal data, and the coverage probabilities decreased with increasing non-normality. This expected result might be attributed to the fact that the confidence intervals constructed were based on the assumption of normality. If this assumption was violated, then Type I error rates were likely to increase. Coverage probabilities under monotonic missing pattern were generally closer to the nominal value of 95% than under non-monotonic missing pattern, with MDTs being less differentiated under monotonic than non-monotonic missing pattern. Generally, for non-normal data, coverage probabilities fell below the nominal value with non-monotonic missing pattern. As reflected by the estimation and coverage probability for β_1, the EM method broke down under non-monotonic pattern at sample size 265.

Implications of the Study

Considering that the choice of MDT in a multiple regression context not only depends on the magnitude of R^2 and number of predictors, but also on the parameter estimate of interest, the following implications for researchers seem pertinent:

(1) The choice of MDT should be of little concern to researchers if the proportion of missing data is 10 percent or less and the missing pattern is monotonic. In particular, in the estimation of R^2 and regression coefficients, as well as coverage probability, the results showed that MDTs were more differentiated with increasing percent missing. However, the differentiation was much larger under non-monotonic missing pattern than under monotonic missing pattern, implying that the choice of MDT is more crucial when the missing pattern is non-monotonic.

(2) The accuracy of parameter estimates after treating missing data was found to deteriorate with increasing non-normality. The results showed that the mean absolute error of parameter estimates was largest at the highest level of non-normality, and smallest for normal data. Coverage probability was found to decrease with increasing non-normality. In general, coverage probability was closest to nominal value for normal data.

(3) With the estimation of R^2 as the goal of analysis, use of EM is recommended if the anticipated R^2 is low (about .2), regardless of number of predictors. However, if the anticipated R^2 is high (about .6), use of PW is recommended, regardless of number of predictors. Although PW performed the worst and EM the best under four predictors and low R^2 condition, EM was the best and PW the worst under four predictors and high R^2 condition. Under nine predictors and low R^2 condition, EM was the best, except at sample size 265 and 20 percent missing under non-monotonic pattern where MS performed the best. Under nine predictors and high R^2 condition, results showed

that PW had the best performance in terms of both accuracy and bias, and EM had the worst performance at all levels of the design variables.

(4) With the estimation of regression coefficients as the goal of analysis, the choice of MDT is most crucial for the variable with the least amount of missing data. Based on the accuracy and coverage probability of regression coefficients, the RS method is recommended regardless of the number of predictors or the magnitude of the anticipated R^2. However, PW and MS methods tended to give less biased estimates than RS when the number of predictors is large (i.e., nine predictors). Researchers should avoid the use of EM when the missing pattern is non-monotonic and sample size is moderate to large (153 and 265) under nine predictors and high R^2 condition because the method breaks down.

A contribution of the present study is that it addressed important and meaningful research questions on the relative effectiveness of selected MDTs. This should advance our understanding of the performance of the MDTs under various design conditions, enabling researchers to select MDTs more appropriately. In the present study, use was made of simulated data. It is important that real data sets are used in future research to validate these results. Also, four and nine predictors were used in the study. A future investigator could modify the number of predictors. Finally, whereas the 95% confidence was used in the calculation of coverage probabilities in the present study, a future investigator could use the 99% confidence interval to extend generalization of results.

Strengths and Limitations of the Present Study

As noted by Kromrey and Hines (1994), a critical issue in the generalizability of results involving MDTs is the consideration of regression models with more predictor variables in which missing data occur on more than one predictor. However, previous researchers tended to use two-predictor regression models in which missing data occurred on only one predictor. Such situations are rare in psycho-educational research. In the present study, four and nine predictor models were used, with data missing on all but one predictor, thus allowing for more realistic generalizations.

Also, unlike some of the previous studies with few replications, 1000 replications was utilized in the present study. This means that the results of the study more accurately represent the relative performance of MDTs that were examined.

Knowledge of coverage probabilities is an important component in the evaluation of any statistical procedure. Previous studies on MDTs failed to investigate this component. In the present study, this vital component was included.

Another important aspect of the present study was that levels of independent variables were determined by consulting the literature. This made the results of the present study more generalizable and more consistent with reality. However, a limitation of the present study was that the levels of non-normality were restricted by the parabolic parameter space as presented in Fleishman (1978). This implies that it was not possible to combine a level of skew with any other level of kurtosis. Such a restriction may sacrifice generalizability, but only as far as levels of non-normality are concerned.

In his review of missing data research, Roth (1994) stressed the conspicuous deficiency of studies contrasting simple MDTs like mean substitution and pairwise deletion to the more complex ones like expectation-maximization (EM) or other maximum likelihood methods. Also, data were randomly deleted in most previous research, and there has been little work with different patterns of missing data. An attempt was made in the present study to address these issues by including the EM method, and comparing MDTs within and across different patterns of missing data.

In conclusion, it was found that percent missing, missing pattern, and multivariate non-normality had an effect on absolute error for standard multiple linear regression estimates. Under low R^2 condition, EM was the best in estimating population R^2. Under high R^2 condition, PW method had the best overall performance in estimating population R^2. In the estimation of regression coefficients, RS had the best overall performance irrespective of magnitude of multiple R^2 and number of predictors. However, the estimates under RS were least biased only under the four-predictor analyses. Coverage probabilities under EM were closest to nominal value under low R^2 condition. Under high R^2 condition, RS had coverage probabilities closest to nominal value.

REFERENCES

Afifi, A. A., & Elashoff, R. M. (1969). Missing observations in multivariate statistics III: Large sample analysis of simple linear regression. American Statistical Association Journal, 64, 337-359.

Allison, P. D. (1987). Estimation of linear models with incomplete data. In C. Clogg (Ed.), Sociological Methodology (pp. 71-103). San Francisco: Jossey-Bass.

Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. Psychometrika, 49, 155-173.

Anderson, L. E., Stone-Romero, E. F., & Tisak, J. (1996). A comparison of bias and mean squared error in parameter estimates of interaction effects: Moderated multiple regression versus errors-in-variables regression. Multivariate Behavioral Research, 31(1), 69-94.

Anderson, T. W. (1957). Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. Journal of the American Statistical Association, 52, 200-203.

Arminger, G., & Sorbel, M. E. (1990). Pseudo-maximum likelihood estimation of mean and covariance structures with missing data. Journal of the American Statistical Association, 85(409), 195-203.

Azen, S. P., Guilder, M., & Hill, M. A. (1989). Estimation of parameters and missing values under a regression model with non-normally distributed and non-randomly incomplete data. Statistics in Medicine, 8, 217-228.

Barford, N. C. (1985). Experimental measurements: Precision, error and truth. NY: John Wiley.

Beale, E. M. L., & Little, R. J. A. (1975). Missing values in multivariate analysis. Journal of the Royal Statistical Society, 37, Series B, 129-145.

Bentler, P. M. (1989). EQS [Computer software]. BMDP Statistical Software Inc: Los Angeles, CA.

Brockmeier, L. L., Kromrey, J. D., & Hines, C. V. (1996). Missing data treatments for nonrandomly missing data and multiple regression analysis. A paper presented at the Annual Meeting of the American Educational and Research Association, New York, April 8-12, 1996.

Brown, R. L. (1994). Efficacy of the indirect approach for estimating structural equation models with missing data: A comparison of five methods. Structural Equation Modeling, 287-316.

Buck, S. F. (1960). A method of estimation of missing values in multivariate data suitable for use with an electronic computer. Journal of the Royal Statistical Society, 22, Series B, 302-307.

Calkins, D. S. (1974). Some effects of non-normal distributions shape on the magnitude of the Pearson product moment correlation coefficient. Interamerican Journal of Psychology, 261-288.

Chan, L. S., Gilman, J. A., & Dunn, O. J. (1976). Alternative approaches to missing values in discriminant analysis. Journal of the American Statistical Association, 71, 842-844.

Cochran, W. G. (1983). Historical Perspective. In William, G. Madow, Ingram Olkin, & Donald B. Rubin (Eds.). Incomplete data in sample surveys: Theory and Bibliographies (Vol. 2, pp. 11-25). NY: Academic Press.

Cohen, J. (1977). Statistical power analysis for the social sciences (Rev. ed.). NY: Academic Press.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Hillsdale, NJ: Earlbaum.

Cohen, J., & Cohen, P. (1983). Applied multiple regression/correlation analysis for the behavioral sciences. Hillsdale, NJ: Lawrence Erlbaum Associates.

Coleman, J. S., Hoffer, T., & Kilgore, S. (1981). Contractors report; Public and private schools: An analysis of High School and Beyond. National Center for Educational Statistics, Washington, DC.

Dear, R. E. (1959). A principal component missing data method for multiple regression models (Technical Report SP-86), System Development Corporation.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, 39, Series B, 1-38.

Dillon, W. R., & Goldstein, M. (1984). Multivariate analysis: Methods and applications. NY: John Wiley.

Dixon, W. J. (1988). BMDP Statistical software [Computer software]. Los Angeles, CA.

Donner, A. (1982). The relative effectiveness of procedures commonly used in multiple regression analysis for dealing with missing values. The American Statistician, 36(4), 378-381.

Donoghue, J. R. (1995). The effects of within-group covariance structure on recovery in cluster analysis. Multivariate Behavioral Research, 30(2), 227-254.

Draper, N. R., & Smith, H. (1981). Applied regression analysis. NY: John Wiley.

Dunteman, G. H. (1984). Introduction to linear models. Beverly Hills, CA: Sage Publications.

Edwards, A. L. (1985). Multiple regression and the analysis of variance and covariance. NY: W. H. Freeman and Company.

Efron, B. (1980). Censored data and the bootstrap. Journal of the American Statistical Association, 76, 312-319.

Engel, U., & Meyer, W. (1996). Structural analysis in the study of social change. In U. Engel & J. Reinecke (Eds.), Analysis of Change: Advanced Techniques in Panel Data Analysis, 221-252. Berlin: Walter de Gruyter & Co.

Finkbeiner, C. (1979). Estimation for the multiple factor model when data are missing. Psychometrika, 44(4), 409-420.

Fleishman, A. I. (1978). A method for simulating non-normal distributions. Psychometrika, 43(4), 521-532.

Frane, J. (1988). Description and estimation of missing data. In Dixon, W. J., Brown, M. B., Engelman, L., Hill, M. A. & Jennrich, R. I. (Eds.). BMDP Statistical Software Manual (pp. 813-829). Berkeley: University of California Press.

Frane, J. W. (1976). Some simple procedures for handling missing data in multivariate analysis. Psychometrika, 41(3), 409-415.

Fraser, C., & Halperin, S. (1998). Comparison of complete-case, pairwise available-case, and maximum likelihood missing data methods. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego.

Gerbing, D. W., & Anderson, J. C. (1985). The effects of sampling error and model characteristics on parameter estimation for maximum likelihood confirmatory factor analysis. Multivariate Behavioral Research, 20, 255-271.

Gilley, O. W., & Leone, R. P. (1991). A two-stage imputation procedure for item nonresponse in surveys. Journal of Business Research, 22, 281-291.

Gleason, T. C., & Stealin, R. (1975). A proposal for handling missing data. Psychometrika, 40(2), 229-252.

Goldberger, A. S. (1991). A course in econometrics. Cambridge, Massachusetts: Harvard University Press.

Graham, J. W., & Donaldson, S. I. (1993). Evaluating interventions with differential attrition: The importance of nonresponse mechanisms and use of follow-up data. Journal of Applied Psychology, 78, 119-128.

Graham, J. W., Hofer, S. M., & Piccinin, A. M. (1994). Analysis with missing data in drug prevention research. In Collins, L. M., & Seitz, L. A. (Eds.), Advances in data analysis for prevention intervention research, NIDA Research Monograph 108, Washington, DC:National Institute on Drug Abuse, 13-63.

Graham, J. W., Hofer, S. M., and MacKinnon, D. P. (1996). Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures. Multivariate Behavioral Research, 31(2), 197-218.

Haitovsky, Y. (1968). Missing data in regression analysis. Journal of the Royal Statistical Society, 30, Series B, 67-82.

Hawkins, M. R., & Merriam, V. H. (1991). An overmodeled world. Direct Marketing, 21-24.

Hocking, R. R., & Smith, W. B. (1968). Estimation of parameters in the multivariate normal distribution with missing observations. Journal of the American Statistical Association, 63, 159-173.

Johnson, M. E. (1987). Multivariate statistical simulation. NY: John Wiley.

Jöreskog, K. G., & Sörbom, D. (1993). LISREL 8: Structural equation modeling with the SIMPLIS command language. [Computer software]. Chicago: Scientific Software International.

Judge, G. G., Hill, R. C., Griffiths, W. E., Lütkepohl, H., & Lee, T. (1988). Introduction to the theory and practice of econometrics. NY: John Wiley.

Kaiser, H. F., & Dickman, K. (1962). Sample and population score matrices and sample correlation matrices from an arbitrary population correlation matrix. Psychometrika, 27, 179-182.

Kim, J., & Curry, J. (1977). The treatment of missing data in multivariate analysis. Sociological Methods & Research, 6(2), 215-240.

King, B. F. (1983). Quota sampling. In Madow, W. G., Nisselson, H., & Olkin, I. (Eds.). Incomplete data in sample surveys, Vol. 2: Theory and Bibliographies. New York: Academic Press.

Kiveeri, H. T. (1987). An incomplete data approach to the analysis of covariance structures. Psychometrika, 52(4), 539-554.

Kolb, R. R., & Dayton, C. M. (1995). Correcting for nonresponse in latent class analysis. Multivariate Behavioral Research, 31(1), 7-32.

Kromrey, J. D., & Hines, C. V. (1994). Nonrandomly missing data in multiple regression: An empirical comparison of common missing-data treatments. Educational and

Psychological Measurement, 54(3), 573-593.

Laird, N. M. (1988). Missing data in longitudinal studies. Statistics in Medicine, 7, 305-315.

Larsen, R. J., & Marx, M. L. (1981). An introduction to mathematical statistics and its applications. Englewood Cliffs, NJ: Prentice-Hall

L'Ecuyer, P. (1988). Efficient and portable combined random number generators. Communications of the ACM, 31, 742-774.

L'Ecuyer, P. (1990). Random numbers for simulation. Communications of the ACM, 33, 85-97.

L'Ecuyer, P. (1994). Recent advances in uniform random number generation. Paper presented at the Winter Simulation Conference, 1994: Lake Buena Vista; FL.

Lepkowski, J. M., Landis, J. R., & Stehouer, S. A. (1987). Strategies for the analysis of imputed data from a sample survey: The national medical care utilization and expenditure. Medical Care, 25, 705-716.

Little, R. J. A. (1979). Maximum likelihood inference for multiple regression with missing values: A simulation study. Journal of the Royal Statistical Society, 41, Series B, 76-87.

Little, R. J. A. (1983). The ignorable case. In Madow, W. G., Olkin, I., & Rubin, D. B. (Eds.) Incomplete data in sample survey, (Vol. 2, pp. 341-382.). NY: Academic Press.

Little, R. J. A. (1988). Missing data adjustments in large surveys. Journal of Business and Economic Statistics, 6, 1-15.

Little, R. J. A. (1992). Regression with missing X's: A review. Journal of the American Statistical Association, 87(420), 1227-1237.

Little, R. J. A., & Rubin, D. B. (1987). Statistical analysis with missing data. NY: John Wiley.

Lyon, A. J. (1970). Dealing with data. NY: Pergamon Press.

Malhotra, N. K. (1987). Analyzing marketing research data with incomplete information on the dependent variable. Journal of Marketing Research, 24, 74-84.

Marascuilo, L. A., & Levin, J. R. (1983). Multivariate statistics in the social sciences: A researcher's guide. Belmont, CA: Brooks/Cole Publishing.

Marini, M. M., Olsen, A. R., Rubin, D. B. (1980). Maximum likelihood estimation in panel studies with missing data. In K. F. Schuessler (Ed.). Sociological Methodology (pp. 314-357). San Francisco, CA: Jossey-Bass.

Marsh, H. W. (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: A Multidisciplinary Journal, 5(1), 22-36.

Micceri, T. (1986). A futile search for that statistical chimera of normality. Paper presented at the 31st Annual Convention of the Florida Educational Research Association, Tampa.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. Psychological Bulletin, 105, 156-166.

Milligan, G.W. (1980). An examination of the effects of of six types of error perturbation on fifteen clustering algorithms. Psychometrika, 45, 325-342.

Milligan, G.W. (1981). A Monte Carlo study of thirty internal criterion measures for cluster analysis. Psychometrika, 46, 187-199.

Milligan, G.W. (1989). A study of the beta-flexible clustering method. Multivariate Behavioral Research, 24, 163-176.

Mooney, C. Z., & Duval, R. D. (1993). Bootstrapping: A nonparametric approach to statistical inference. Newbury Park, CA: Sage Publications.

Muthen, B., Kaplan, D., & Holis, M. (1987). On structural equation modeling with data that are not missing completely at random. Psychometrika, 52, 431-462.

Neter, J., Wasserman, W., & Kutner, M. H. (1983). Applied linear regression models. Homewood, IL: Richard D. Irwin.

Odell, P. L., & Feiveson, A. H. (1966). A numerical procedure to generate a sample covariance matrix. Journal of the American Statistical Association, 61, 199-203.

Pearson, E. S., & Please, N. W. (1975). Relation between the shape of population distribution of four simple test statistics. Biometrika, 62, 223-241.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). Numerical recipes in FORTRAN: The art of scientific computing. NY: Cambridge University Press.

Raymond, M. R. (1986). Missing data in evaluation research. Evaluation and the Health Profession, 9, 395-420.

Raymond, M. R., & Roberts, D. M. (1987). A comparison of methods for treating incomplete data in selection research. Educational and Psychological Measurement, 47, 13-26.

Reinecke, J. & Schimdt, P. (1996). Model specification and missing value treatment in panel data. In U. Engel & J. Reinecke (Eds.), Analysis of Change: Advanced Techniques in Panel Data Analysis, (pp. 161-189). Berlin: Walter de Gruyter & Co.

Roth, P. L. (1994). Missing data: A conceptual review for applied psychologists. Personnel Psychology, 47, 537-560.

Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. Annals of Statistics, 12, 1151-1172.

Rubin, D. B. (1991). EM and beyond. Psychometrika, 56(2), 241-254.

Sawilosky, S. S., & Blair, R. C. (1992). A more realistic look at the robustness and Type Error properties of the t test to departures from population normality. Psychological Bulletin, 3(2), 352-360.

Seaman, S. L., Algina, J., & Olejnik, S. F. (1985). Type I error probabilities and power of the rank and parametric ancova procedures. Journal of Educational Statistics, 10(4), 345-367.

Shchigolev, B. M. (1965). Mathematical analysis of observations (H. Eagle, Trans.). NY: Elsevier Publishing Co. (Original work published 1960).

Simonoff, J. S. (1988). Regression diagnostics to detect nonrandom missingness in linear regression. Technometrics, 30, 205-214.

SPSS [Computer software] (1993). SPSS Inc., Chicago.

Stumpf, S. A. (1978). A note on handling missing data. Journal of Management, 4, 65-73.

Sudman, S. (1967). Reducing the cost of surveys. Chicago: Aldin.

Tabachnick, B. G., & Fidell, L. S. (1983). Using multivariate statistics. NY: Harper & Row.

Timm, N. H. (1970). The estimation of variance-covariance and correlation matrices from incomplete data. Psychometrika, 35, 417-438.

Tirri, H., & Silander, T. (1998). Stochastic complexity based estimation of missing elements in questionnaire data. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego.

Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. Psychometrika, 48, 465-471.

Wilks, S. S. (1932). Moments and distributions of estimates of population parameters from fragmentary samples. Annals of Mathematical Statistics, 3, 163-195.

APPENDIX A
BIAS IN PARAMETER ESTIMATES

Table A.1: Bias in $R^2_{estimate}$: 4 predictors and low R^2 condition

			Monotonic Missing Pattern				Non-monotonic Missing Pattern			
N	Percent Missing	Normality (sk., kurt.)	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation
94	10	(0,0)	.014266	.015673	.026846	.014973	.005228	.013706	.022897	.007649
		(1,3)	.018106	.019812	.031224	.018943	.007193	.018463	.025382	.009785
		(1.8,6)	.024917	.026993	.037877	.025909	.011700	.024425	.030684	.014437
		(3,25)	.044387	.048456	.055844	.046140	.018653	.041704	.039049	.021012
	15	(0,0)	.004107	.006573	.023575	.005265	-.006121	.006203	.017708	-.003531
		(1,3)	.007670	.010544	.027176	.008997	-.004370	.010777	.019654	-.001796
		(1.8,6)	.014246	.017606	.032999	.015753	-.000070	.016326	.024421	.002472
		(3,25)	.032844	.038732	.048037	.035135	.004499	.030531	.027681	.006057
	20	(0,0)	-.006420	-.002734	.019325	-.004837	-.017661	-.001691	.012267	-.015170
		(1,3)	-.003002	.001253	.022340	-.001200	-.016012	.002508	.013185	-.013770
		(1.8,6)	.003556	.008459	.027135	.005566	-.011744	.007848	.016764	-.009607
		(3,25)	.020912	.029022	.039035	.023670	-.008049	.020940	.014695	-.007253
153	10	(0,0)	.000732	.001733	.015641	.001221	-.007253	-.007671	-.000044	.011790
		(1,3)	.003823	.005051	.018485	.004405	-.005508	-.006549	.003639	.012977
		(1.8,6)	.008844	.010328	.022917	.009531	-.004189	-.003493	.007838	.016326
		(3,25)	.024834	.027645	.037445	.026056	-.001080	-.000199	.020305	.022359
	15	(0,0)	-.009078	-.007223	.013557	-.008211	.001855	-.019266	-.007887	.007804
		(1,3)	-.005969	-.003768	.016143	-.004966	-.016849	-.018915	-.005019	.008681
		(1.8,6)	-.001081	.001506	.020180	.000068	-.016576	-.016417	-.001591	.011663
		(3,25)	.014051	.018620	.033150	.015916	-.014807	.008383	.015062	-.013461
	20	(0,0)	-.018985	-.016116	.010855	-.017809	-.029701	-.015390	.003777	-.027609
		(1,3)	-.015807	-.012512	.013049	-.014487	-.029573	-.013031	.004287	-.027720
		(1.8,6)	-.010946	-.007183	.016538	-.009472	-.027707	-.010339	.006982	-.025957
		(3,25)	.002832	.009186	.027708	.005117	-.026062	-.000917	.009241	-.025397
265	10	(0,0)	-.007386	-.006653	.008374	-.007020	-.015480	-.008296	.004675	-.013391
		(1,3)	-.005643	-.004833	.010253	-.005246	-.015897	.006384	.005419	-.013636
		(1.8,6)	-.002768	-.001867	.013036	-.002336	-.015145	-.004821	.007477	-.012881
		(3,25)	.008313	.009743	.023768	.008959	-.014766	.002636	.011517	-.012973
	15	(0,0)	-.017002	-.015550	.007074	-.016337	-.026665	-.016087	.001471	-.024455
		(1,3)	-.015346	-.013766	.008743	-.014634	-.027575	-.014810	.001919	-.025484
		(1.8,6)	-.012460	-.010733	.011215	-.011693	-.027056	-.013548	.003764	-.025039
		(3,25)	-.001774	.000732	.020888	-.000691	-.028262	-.008132	.006587	-.027139
	20	(0,0)	-.026380	-.024027	.005421	-.025409	-.037143	-.023661	-.001555	-.035151
		(1,3)	-.024784	-.022240	.006816	-.023751	-.038160	-.022753	-.001321	-.036445
		(1.8,6)	-.021948	-.019174	.008922	-.020844	-.037845	-.021912	.000417	-.036238
		(3,25)	-.011708	-.007810	.017423	-.010320	-.039238	-.017029	.001553	-.038617

Table A.2: Bias in b_1: 4 predictors and low R^2 condition

			Monotonic Missing Pattern				Non-monotonic Missing Pattern			
N	Percent Missing	Normality (sk., kurt.)	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation
94	10	(0,0)	-.003553	.017797	.017631	.001535	-.015127	.063690	.063519	-.004284
		(1,3)	-.004239	.017267	.017107	.000842	-.023164	.067521	.067273	-.013102
		(1.8,6)	-.004028	.017329	.017177	.000737	-.025374	.069068	.068474	-.015662
		(3,25)	-.004694	.020150	.019912	-.000223	-.040202	.092975	.092001	-.034090
	15	(0,0)	-.002918	.028086	.027854	.003538	-.016335	.079435	.079254	-.007360
		(1,3)	-.003073	.028157	.027920	.003369	-.023698	.082609	.082403	-.015818
		(1.8,6)	-.002566	.028646	.028395	.003654	-.024992	.084241	.083585	-.017569
		(3,25)	-.003445	.033081	.032771	.002614	-.036547	.107390	.106197	-.034251
	20	(0,0)	-.004045	.037464	.037158	.003816	-.019630	.089585	.089379	-.012553
		(1,3)	-.004454	.037401	.037099	.003539	-.025813	.092595	.092344	-.020017
		(1.8,6)	-.003874	.037881	.037598	.003947	-.026716	.094172	.093448	-.021432
		(3,25)	-.005430	.042480	.042179	.002244	-.036299	.116431	.115084	-.036463
153	10	(0,0)	-.006552	.015227	.015104	-.001045	-.017654	.061592	.061544	-.006370
		(1,3)	-.007023	.015361	.015234	-.001416	-.024650	.066959	.066828	-.014007
		(1.8,6)	-.006630	.016109	.015986	-.001080	-.026071	.069333	.068876	-.015805
		(3,25)	-.005940	.018269	.018112	-.000828	-.039287	.091827	.091088	-.032804
	15	(0,0)	-.007904	.024452	.024255	-.000362	-.020207	.076926	.076864	-.010307
		(1,3)	-.008693	.024608	.024398	-.001026	-.025955	.081724	.081608	.026066
		(1.8,6)	-.008465	.025439	.025229	-.000875	.084090	.083573	-.017916	-.035630
		(3,25)	-.008599	.028423	.028191	-.001100	.105499	.104583	-.032104	
	20	(0,0)	-.007575	.033347	.033133	.000565	-.021086	.087511	.087449	-.013943
		(1,3)	-.008272	.033511	.033312	-.000050	-.025291	-.091779	.091682	.019459
		(1.8,6)	-.008249	.034237	.034051	-.000098	-.024773	.094038	.093512	-.019415
		(3,25)	-.008856	.038024	.037844	-.000414	-.032505	.113444	.112478	-.031413
265	10	(0,0)	-.006541	.015030	.014940	-.000822	-.006541	.015030	.014940	-.000822
		(1,3)	-.006653	.015191	.015100	-.000898	-.006653	.015191	.015100	-.000898
		(1.8,6)	-.006538	.015429	.015344	-.000805	-.006538	.015429	.015344	-.000805
		(3,25)	-.005659	.017387	.017280	.000014	-.005659	.017387	.017280	.000014
	15	(0,0)	-.007400	.023945	.023794	-.000024	-.007400	.023945	.023794	.000024
		(1,3)	-.007536	.024159	.024016	-.000128	-.007536	-.024159	.024016	-.000128
		(1.8,6)	-.007468	.024466	.024326	-.000097	-.007468	-.024466	.024326	.000097
		(3,25)	-.006843	.026869	.026731	.000674	-.006843	-.026869	.026731	.000674
	20	(0,0)	-.008453	.032562	.032388	.000090	-.008453	.032562	.032388	.000090
		(1,3)	-.008719	.032793	.032638	-.000087	-.008719	.032793	.032638	-.000087
		(1.8,6)	-.008777	.033140	.032993	-.000147	-.008777	.033140	.032993	-.000147
		(3,25)	-.008820	.035666	.035538	.000063	-.008820	.035666	.035538	.000063

Table A.3: Bias in $R^2_{estimate}$: 4 predictors and high R^2 condition

			Monotonic Missing Pattern				Non-monotonic Missing Pattern			
N	Percent Missing	Normality (sk., kurt.)	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation
94	10	(0,0)	-.043071	-.032549	.006910	-.035688	-.095194	-.037159	-.001864	-.073098
		(1,3)	-.038364	-.026508	.012208	-.030401	-.107712	-.036615	.000183	-.085818
		(1.8,6)	-.032025	-.019128	.018052	-.023750	-.107743	-.032931	.004372	-.086825
		(3,25)	-.006893	.011155	.039464	.002736	-.145121	-.036613	.010853	-.131758
	15	(0,0)	-.072550	-.055330	.003996	-.061662	-.138762	-.057195	-.009321	-.116293
		(1,3)	-.068037	-.049376	.008624	-.056655	-.153030	-.060504	-.008062	-.133318
		(1.8,6)	-.061726	-.041904	.013712	-.050136	-.153357	-.058508	-.004159	-.135452
		(3,25)	-.037447	-.011666	.032307	-.024891	-.190371	-.069734	-.001821	-.182862
	20	(0,0)	-.102963	-.078183	.000198	-.089139	-.179676	-.077683	-.016991	-.159654
		(1,3)	-.098870	-.072418	.003744	-.084586	-.192358	-.082682	-.015700	-.176377
		(1.8,6)	-.093023	-.065212	.007473	-.078548	-.192811	-.081983	-.012635	-.178327
		(3,25)	-.070800	-.035847	.020160	-.055429	-.226364	-.098519	-.012941	-.222622
153	10	(0,0)	-.048001	-.038459	.003937	-.041188	-.101753	-.042653	-.004898	-.079005
		(1,3)	-.044315	-.033656	.007585	-.036928	-.117484	-.045083	-.004322	-.095044
		(1.8,6)	-.039674	-.028173	.011506	-.031941	-.120161	-.043950	-.001821	-.098610
		(3,25)	-.021813	-.006575	.029001	-.012702	-.162644	-.055057	.002803	-.149232
	15	(0,0)	-.077549	-.061168	.001876	-.066905	-.146712	-.063369	-.011451	-.122958
		(1,3)	-.074276	-.056434	.005151	-.063061	-.162431	-.067912	-.011314	-.141598
		(1.8,6)	-.070005	-.051070	.008595	-.058485	-.165459	-.068711	-.009097	-.146305
		(3,25)	-.052585	-.028829	.023926	-.039816	-.208631	-.089476	-.007394	-.200648
	20	(0,0)	-.106965	-.083823	-.000534	-.093800	-.183666	-.082938	-.017718	-.164268
		(1,3)	-.103710	-.078908	.002210	-.090011	-.198169	-.089170	-.017852	-.182995
		(1.8,6)	-.099457	-.073376	.005050	-.085493	-.200403	-.090644	-.015960	-.186945
		(3,25)	-.083699	-.051473	.017084	-.068645	-.240913	-.116290	-.016857	-.238137
265	10	(0,0)	-.051019	-.042208	.001108	-.044514	-.106710	-.047061	-.007823	-.083144
		(1,3)	-.048718	-.039454	.003460	-.042001	-.124264	-.051691	-.008146	-.101166
		(1.8,6)	-.046054	-.036391	.006005	-.039181	-.129720	-.053893	-.006644	-.107582
		(3,25)	-.032769	-.020787	.018571	-.024861	-.178230	-.074161	-.005797	-.164564
	15	(0,0)	-.080257	-.065187	-.000349	-.070342	-.149003	-.066694	-.013685	-.125639
		(1,3)	-.078058	-.062408	.001715	-.067940	-.166520	-.073567	-.014329	-.146217
		(1.8,6)	-.075346	-.059222	.003932	-.065106	-.171866	-.076811	-.013067	-.153113
		(3,25)	-.062479	-.043555	.015243	-.051156	-.219971	-.105540	-.014288	-.212453
	20	(0,0)	-.109551	-.087571	-.002185	-.096749	-.187856	-.086303	-.019511	-.167753
		(1,3)	-.107484	-.084790	-.000373	-.094517	-.203705	-.094695	-.020210	-.187895
		(1.8,6)	-.104830	-.081495	.001530	-.091743	-.208198	-.098943	-.018821	-.194088
		(3,25)	-.092830	-.065960	.011134	-.078886	-.250940	-.131747	-.021038	-.247438

Table A.4: Bias in b_1: 4 predictors and high R^2 condition

			Monotonic Missing Pattern				Non-monotonic Missing Pattern			
N	Percent Missing	Normality (sk., kurt.)	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation
94	10	(0,0)	-.044364	.026069	.025725	-.021430	-.132160	.100736	.100575	-.088402
		(1,3)	-.044122	.026840	.026485	-.021743	-.148362	.110025	.109789	-.108142
		(1.8,6)	-.043598	.027679	.027321	-.021862	-.150516	.114023	.113351	-.111796
		(3,25)	-.035885	.035400	.035054	-.018999	-.168047	.158559	.157811	-.145154
	15	(0,0)	-.058228	.042136	.041695	-.029849	-.152025	.125779	.125605	-.114983
		(1,3)	-.057623	.043638	.043176	-.029758	-.162449	.134388	.134122	-.131328
		(1.8,6)	-.056943	.044909	.044435	-.029670	-.161765	.138342	.137455	-.132865
		(3,25)	-.050522	.054138	.053716	-.027550	-.164706	.181136	.180042	-.153287
	20	(0,0)	-.073890	.056536	.055978	-.041796	-.168786	.142991	.142765	-.139666
		(1,3)	-.073106	.058172	.057663	-.041416	-.173436	.151206	.150919	-.150438
		(1.8,6)	-.071935	.059611	.059151	-040749	-.171206	.155421	.154344	-.149302
		(3,25)	-.066650	.069425	.069051	-.039302	-.161881	.195152	.193755	-.155670
153	10	(0,0)	-.046467	.024243	.024096	-.022734	-.136399	.100432	.100355	-.091043
		(1,3)	-.046337	.024550	.024394	-.023342	-.151974	.110715	.110517	-.110951
		(1.8,6)	-.045778	.025053	.024904	-.023499	-.153833	.114722	.114057	-.114661
		(3,25)	-.039826	.029514	.029379	-.021443	-.171656	.153399	.152608	-.149686
	15	(0,0)	-.062653	.039902	.039632	-.031998	-.158198	.125311	.125201	-.118249
		(1,3)	-.063226	.040630	.040343	-.033101	-.167713	.134493	.134260	-.134406
		(1.8,6)	-.063030	.041503	.041223	-.033529	-.166458	.138504	.137603	-.135336
		(3,25)	-.058190	.047533	.047246	-.032212	-.167809	.174731	.173572	-.155699
	20	(0,0)	-.075770	.051262	.050928	-.043572	-.169523	.141382	.141242	-.141259
		(1,3)	-.076477	.051749	.051429	-.044743	-.174145	.149399	.149108	-.152651
		(1.8,6)	-.076460	.052451	.052159	-.045344	-.171165	.153037	.151970	-.151277
		(3,25)	-.073103	.058327	.058094	-.044910	-.162976	.185987	.184636	-.159007
265	10	(0,0)	-.046282	.024345	.024173	-.021832	-.137794	.101371	.101335	-.090241
		(1,3)	-.045909	.024531	.024356	-.021873	-.152678	.111219	.111081	-.110171
		(1.8,6)	-.045558	.024702	.024540	-.021889	-.154090	.115099	.114524	-.113651
		(3,25)	-.041959	.026978	.026785	-.020685	-.169893	.149754	.149025	-.147836
	15	(0,0)	-.061748	.037609	.037369	-.031900	-.158278	.123975	.123940	-.118681
		(1,3)	-.061452	.038125	.037896	-.031964	-.167210	.132569	.132390	-.134581
		(1.8,6)	-.061205	.038441	.038228	-.032089	-.166410	.136006	.135228	-.135922
		(3,25)	-.058529	.041565	.041341	-.031495	-.166441	.167875	.166842	-.155049
	20	(0,0)	-.076647	.049888	.049602	-.043747	-.172854	.140906	.140858	-.142859
		(1,3)	-.076710	.050160	.049903	-.044036	-.176512	.148401	.148209	-.153502
		(1.8,6)	-.076655	.050354	-.050126	-.044308	-.173384	.151357	.150425	-.152026
		(3,25)	-.074796	.053468	.053238	-.044188	-.161446	.179177	.177942	-.156020

Table A.5: Bias in $R^2_{estimate}$: 9 predictors and low R^2 condition

			Monotonic Missing Pattern				Non-monotonic Missing Pattern			
N	Percent Missing	Normality (sk., kurt.)	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation
94	10	(0,0)	.052831	.059312	.050712	.045430	.053235	.056189	.043350	.034993
		(1,3)	.057554	.061613	.055008	.049185	.057184	.057450	.043191	.034861
		(1.8,6)	.068053	.069384	.065000	.058658	.065366	.064697	.048981	.040809
		(3,25)	.093996	.088722	.088315	.080302	.088336	.078952	.054953	.049918
	15	(0,0)	.042348	.051091	.038944	.032150	.045819	.046922	.030176	.021457
		(1,3)	.048042	.052648	.043937	.036539	.049473	.047485	.029153	.021458
		(1.8,6)	.059131	.059179	.054209	.046359	.058149	.052881	.034892	.027905
		(3,25)	.084774	.075178	.076173	.066702	.078671	.061385	.036271	.034405
	20	(0,0)	.033224	.039240	.027276	.020160	.036377	.034118	.014620	.007609
		(1,3)	.039237	.039235	.032185	.024548	.040572	.033331	.013703	.008001
		(1.8,6)	.050370	.043741	.042017	.034070	.048783	.037227	.018858	.013979
		(3,25)	.075960	.054680	.062433	.053163	.066895	.040351	.018218	.018735
153	10	(0,0)	.024081	.037727	.022876	.018704	.023693	.035168	.015583	.007499
		(1,3)	.027003	.039441	.025580	.021066	.026480	.036124	.014916	.006825
		(1.8,6)	.033505	.044340	.031796	.027030	.032271	.040693	.019070	.011250
		(3,25)	.055026	.062047	.052002	.045564	.049620	.053392	.022516	.017264
	15	(0,0)	.013263	.032734	.010793	.005506	.014975	.028709	.001444	-.006379
		(1,3)	.016091	.033953	.013235	.007658	.017794	.028972	.000218	-.006835
		(1.8,6)	.022644	.038081	.019311	.013459	.023191	.032933	.003764	-.002739
		(3,25)	.043891	.053315	.038164	.030784	.038522	.042412	.003946	.001574
	20	(0,0)	.002158	.026110	-.002039	-.007942	.005806	.020213	-.013383	-.020132
		(1,3)	.005529	.026781	.000791	-.005350	.008749	.019928	-.014600	-.020100
		(1.8,6)	.012498	.030075	.007104	.000792	.014139	.022724	-.011206	-.015933
		(3,25)	.033418	.041906	.024703	.017326	.026310	.029411	-.013422	-.013770
265	10	(0,0)	.004579	.020329	.003728	.000501	.003144	.018141	-.003967	-.011415
		(1,3)	.006521	.021854	.005526	.002107	.005299	.018829	-.004977	-.012436
		(1.8,6)	.010616	.025318	.009463	.005910	.008633	.021799	-.002985	-.010177
		(3,25)	.025436	.038053	.023430	.018729	.019797	.030759	-.003701	-.008634
	15	(0,0)	-.005299	.017434	-.007048	-.011483	-.005627	.014069	-.017567	-.024917
		(1,3)	-.003380	.018667	-.005359	-.009935	-.003845	.014442	-.019538	-.026099
		(1.8,6)	.000668	.021685	-.001566	-.006268	-.000708	.017014	-.017849	-.023887
		(3,25)	.015257	.032797	.011519	.005518	.008221	.023300	-.020797	-.023545
	20	(0,0)	-.015384	.013401	-.018555	-.023676	-.014619	.008777	-.031385	-.037650
		(1,3)	-.013325	.014295	-.016832	-.022071	-.013399	.008754	-.033816	-.038862
		(1.8,6)	-.009165	.016764	-.013045	-.018375	-.010731	.010739	-.032510	-.036887
		(3,25)	.005661	.026275	-.000244	-.006610	-.034483	.015989	-.036778	-.037600

Table A.6: Bias in b_1: 9 predictors and low R^2 condition

			Monotonic Missing Pattern				Non-monotonic Missing Pattern			
N	Percent Missing	Normality (sk., kurt.)	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation
94	10	(0,0)	.009833	.009915	-.006128	-.012206	.046551	.046828	-.030872	-.046161
		(1,3)	.012007	.012088	-.004135	-.009414	.051793	.051944	-.040177	-.052405
		(1.8,6)	.013377	.013477	-.002674	-.007032	.054007	.053769	-.042492	-.053081
		(3,25)	.017202	.017423	-.000692	-.002672	.071107	.070396	-.069777	-.067768
	15	(0,0)	.015750	.015961	-.008358	-.019312	.058620	.059061	-.041426	-.055151
		(1,3)	.017836	.018095	-.006379	-.016617	.063432	.063626	-.051031	-.060677
		(1.8,6)	.019042	.019393	-.004937	-.013997	.065537	.065216	-.053646	-.061220
		(3,25)	.023199	.023591	-.002933	-.009596	.083412	.082471	-.081168	-.072703
	20	(0,0)	.020966	.021316	-.010513	-.022902	.066610	.067359	-.049185	-.056977
		(1,3)	.022899	.023283	-.008691	-.020235	.071151	.071629	-.058033	.061282
		(1.8,6)	.024120	.024609	-.007232	-.017759	.073683	.073595	-.059195	-.060863
		(3,25)	.028878	.029354	-.005511	-.014754	.091164	.090313	-.082623	-.070609
153	10	(0,0)	.009867	.009773	-.005260	-.015038	.045468	.045503	-.031524	-.054054
		(1,3)	.011752	.011684	-.003581	-.012758	.050471	.050409	-.041568	-.061210
		(1.8,6)	.013083	.013041	-.002348	-.010711	.051916	.051574	-.044378	-.061964
		(3,25)	.016113	.016123	.000048	-.006159	.069493	.068918	-.071630	-.076080
	15	(0,0)	.015659	.015460	-.007083	-.019716	.056634	.056676	-.042634	-.059257
		(1,3)	.017369	.017194	-.005917	-.018070	.061270	.061194	-.052341	-.065164
		(1.8,6)	.018684	.018536	-.004998	-.016479	.062941	.062529	-.054350	-.065314
		(3,25)	.021626	.021543	-.003385	-.012894	.080118	.079382	-.079808	-.075771
	20	(0,0)	.020790	.020510	-.009701	-.025059	.065799	.065861	-.051334	-.063507
		(1,3)	.022727	.022481	-.008400	-.023304	.070028	.069976	-.059848	-.067817
		(1.8,6)	.024085	.023885	-.007472	-.021714	.071950	.071494	-.061095	-.067414
		(3,25)	.027615	.027512	-.005445	-.018000	.088507	.087753	-.081332	-.074605
265	10	(0,0)	.010831	.010839	-.004703	-.016976	.045312	.045329	-.031007	-.056489
		(1,3)	.011962	.011955	-.003757	-.015561	.050204	.050133	-.041230	-.063826
		(1.8,6)	.012407	.012397	-.003389	-.014727	.051504	.051205	-.043945	-.064762
		(3,25)	.015673	.015649	-.000770	-.010625	.068176	.067690	-.068839	-.076688
	15	(0,0)	.016277	.016234	-.007415	-.023142	.056651	.056693	-.042119	-.061757
		(1,3)	.017399	.017356	-.006549	-.021787	.061140	.061079	-.051694	-.067169
		(1.8,6)	.017859	.017824	-.006119	-.020807	.062256	.061927	-.053800	-.067625
		(3,25)	.021071	.021028	-.003894	-.017447	.078447	.077838	-.075765	-.076355
	20	(0,0)	.020799	.020700	-.010693	-.027943	.064524	.064590	-.050922	-.064597
		(1,3)	.021857	.021772	-.009978	-.026779	.068494	.068452	-.058832	-.068192
		(1.8,6)	.022322	.022261	-.009663	-.025880	.069789	.069426	-.059756	-.067639
		(3,25)	.025640	.025576	-.007823	-.023140	.085074	.084377	-.077983	-.073861

Table A.7: Bias in $R^2_{estimate}$: 9 predictors and high R^2 condition

			Monotonic Missing Pattern				Non-monotonic Missing Pattern			
N	Percent Missing	Normality (sk., kurt.)	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation
94	10	(0,0)	-.017995	-.009956	.025089	-.013499	-.028137	-.010155	.022230	-.019576
		(1,3)	-.011074	-.001459	.030886	-.005921	-.026974	-.003873	.026239	-.017799
		(1.8,6)	-.001342	.009484	.039016	.004122	-.021177	.004594	.032855	-.012141
		(3,25)	.026631	.043122	.065413	.032964	-.005060	.037215	.048783	.000882
	15	(0,0)	-.051404	-.039536	.019039	-.043741	-.065173	-.038214	.015042	-.052603
		(1,3)	-.044442	-.030618	.024243	-.036097	-.064187	-.031898	.018343	-.052256
		(1.8,6)	-.034444	-.019071	.031202	-.025821	-.057584	-.022675	.023618	-.046445
		(3,25)	-.009133	.012538	.053658	-.000012	-.046609	.004988	.033044	-.041798
	20	(0,0)	-.081299	-.064520	.009862	-.072634	-.096991	-.062968	.005055	-.085631
		(1,3)	-.074832	-.055274	.014020	-.065539	-.095711	-.056517	.007326	-.085894
		(1.8,6)	-.065674	-.043592	.019326	-.056088	-.089486	-.047601	.010928	-.080946
		(3,25)	-.041498	-.010947	.037520	-.032317	-.084568	-.025069	.011960	-.083590
153	10	(0,0)	-.036562	-.029355	.015445	-.031404	-.048530	-.029235	.013361	-.037181
		(1,3)	-.032284	-.023989	.019739	-.026709	-.048734	-.025062	.016540	-.037021
		(1.8,6)	-.026089	-.016858	.025384	-.020246	-.044695	-.019270	.021409	-.033354
		(3,25)	-.003966	.009549	.046896	.003173	-.038882	.000362	.035683	-.030228
	15	(0,0)	-.063702	-.053764	.011637	-.057242	-.077228	-.053296	.008996	-.065527
		(1,3)	-.059797	-.048354	.015178	-.052861	-.077823	-.049514	.011582	-.066571
		(1.8,6)	-.053840	-.041053	.019909	-.046613	-.073998	-.043663	.015786	-.063397
		(3,25)	-.032704	-.014338	.037531	-.024599	-.071683	-.028357	.025573	-.065662
	20	(0,0)	-.093607	-.081002	.006842	-.085963	-.108987	-.079821	.003031	-.096971
		(1,3)	-.090231	-.075987	.009479	-.082191	-.109711	-.076410	.004875	-.099037
		(1.8,6)	-.084793	-.069016	.013109	-.076503	-.106715	-.071468	.007742	-.097020
		(3,25)	-.065806	-.043205	.026572	-.056964	-.108682	-.060447	.014062	-.104516
265	10	(0,0)	-.043399	-.037496	.009098	-.038952	-.056210	-.039073	.007455	-.045490
		(1,3)	-.041188	-.034554	.011784	-.036420	-.058064	-.036977	.009473	-.046868
		(1.8,6)	-.037866	-.030653	.015164	-.032911	-.056347	-.033968	.012447	-.045487
		(3,25)	-.020277	-.010110	.031287	-.014378	-.057637	-.024000	.024606	-.049238
	15	(0,0)	-.072425	-.063715	.006886	-.066362	-.086826	-.064416	.004503	-.075533
		(1,3)	-.070719	-.061155	.009166	-.064405	-.089589	-.063159	.005985	-.078775
		(1.8,6)	-.067832	-.057591	.012047	-.061379	-.088402	-.060668	.008435	-.078224
		(3,25)	-.050912	-.036962	.026188	-.043600	-.094420	-.055867	.017096	-.088534
	20	(0,0)	-.100549	-.089379	.004104	-.093538	-.115445	-.088847	.000707	-.104511
		(1,3)	-.099016	-.086914	.005969	-.091833	-.118679	-.088481	.001677	-.109063
		(1.8,6)	-.096217	-.083300	.008242	-.088963	-.118034	-.086515	.003559	-.109263
		(3,25)	-.080272	-.062953	.020309	-.072286	-.127192	-.084163	.009566	-.122996

Table A.8: Bias in b_1: 9 predictors and high R^2 condition

			Monotonic Missing Pattern				Non-monotonic Missing Pattern			
N	Percent Missing	Normality (sk., kurt.)	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation	EM Imputation	Mean Substitution	Pairwise Deletion	Regress. Imputation
94	10	(0,0)	-.011287	.001953	.001858	-.012416	-.057029	.026934	.026994	-.040914
		(1,3)	-.007595	.005145	.004984	-.009720	-.065066	.031838	.031892	-.049455
		(1.8,6)	-.004372	.007152	.006943	-.007985	-.065198	.034185	.034207	-.050841
		(3,25)	.004262	.014062	.014011	-.003194	-.077835	.047597	.047385	-.074398
	15	(0,0)	-.026645	.007626	.007540	-.014560	-.080429	.041427	.041620	-.051136
		(1,3)	-.023158	.010240	.010120	-.012398	-.085836	.047062	.047222	-.061345
		(1.8,6)	-.020280	.011878	.011760	-.011009	-.084729	.049653	.049714	-.062781
		(3,25)	-.010855	.018699	.018673	-.006710	-.093152	.066573	.066442	-.088411
	20	(0,0)	-.034395	.009464	.009587	-.020381	-.084472	.048176	.048583	-.063265
		(1,3)	-.031407	.011364	.011453	-.018546	-.089198	.053214	.053615	-.072875
		(1.8,6)	-.029057	.012553	.012714	-.017468	-.087593	.056135	.056588	-.073021
		(3,25)	-.020205	.019898	.020006	-.013120	-.093497	.074165	.074469	-.094165
153	10	(0,0)	-.019732	.003859	.003686	-.009888	-.077828	.031150	.031165	-.041626
		(1,3)	-.017092	.006027	.005900	-.008104	-.086020	.035997	.036049	-.052341
		(1.8,6)	-.015059	.007503	.007415	-.006893	-.086306	.037559	.037587	-.054869
		(3,25)	-.006938	.012914	.012878	-.002294	-.096188	.053765	.053806	-.080782
	15	(0,0)	-.028588	.006187	.005901	-.013086	-.084891	.037233	.037215	-.052358
		(1,3)	-.026289	.008288	.008040	-.011926	-.091186	.042616	.042658	-.062658
		(1.8,6)	-.024436	.009665	.009457	-.011232	-.090582	.044540	.044591	-.064355
		(3,25)	-.016528	.015051	.014888	-.007673	-.097244	.062523	.062738	-.087630
	20	(0,0)	-.039642	.008587	.008156	-.016926	-.091750	.045712	.045734	-.060522
		(1,3)	-.037745	.010272	.009892	-.016041	-.096106	.050882	.050965	-.070347
		(1.8,6)	-.035971	.011498	.011195	-.015570	-.095059	.053045	.053113	-.071748
		(3,25)	-.029493	.016256	.016091	-.013042	-.097851	.072051	.072274	-.091367
265	10	(0,0)	-.021870	.004424	.004406	-.008603	-.078951	.030216	.030201	-.039014
		(1,3)	-.019896	.005821	.005806	-.007473	-.088048	.034897	.034864	-.050516
		(1.8,6)	-.018535	.006777	.006772	-.006806	-.088706	.036440	.036395	-.053299
		(3,25)	-.012584	.010609	.010603	-.003608	-.099820	.052223	.052186	-.079756
	15	(0,0)	-.032584	.007716	.007656	-.012830	-.086684	.039610	.039605	-.051024
		(1,3)	-.030834	.009050	.008999	-.011945	-.092939	.044699	.044642	-.062034
		(1.8,6)	-.029581	.009929	.009902	-.011423	-.092888	.046124	.046108	-.064270
		(3,25)	-.024657	.013471	.013453	-.009254	-.099939	.062823	.062794	-.088026
	20	(0,0)	-.041984	.009583	.009470	-.017201	-.091298	.045849	.045860	-.059978
		(1,3)	-.040564	.010794	.010703	-.016522	-.095787	.050734	.050730	-.070199
		(1.8,6)	-.039316	.011600	.011551	-.016131	-.095224	.052342	.052321	-.071788
		(3, 25)	-.035205	.015281	.015263	-.014199	-.020604	.023871	.023883	-.021451

APPENDIX B
GRAPHS FOR BIAS IN PARAMETER ESTIMATES

MONOTONIC PATTERN | **NON-MONOTONIC PATTERN**

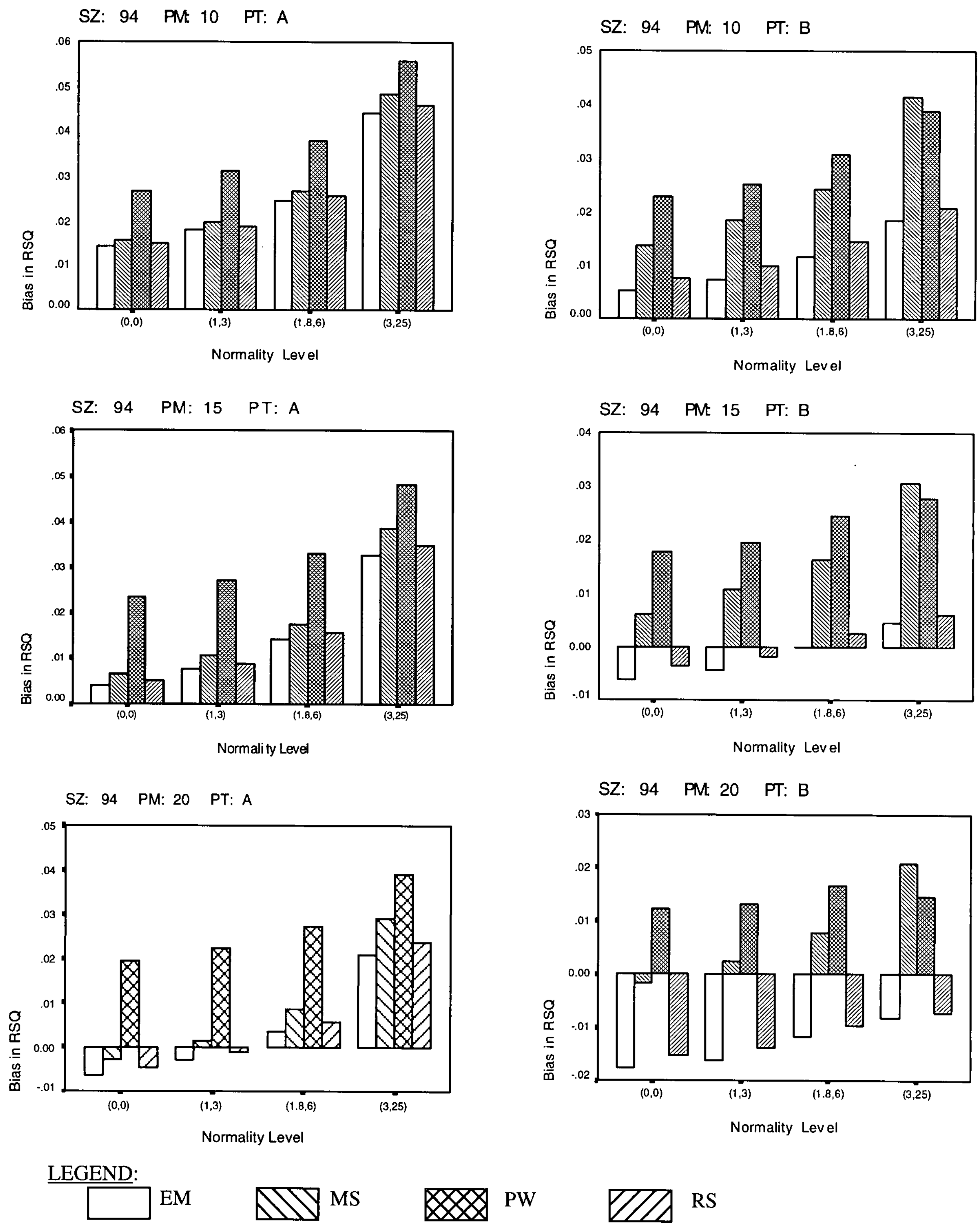


Figure B.1: Bias in $R^2_{estimate}$ under four predictors and low R^2 condition

Note: SZ = Sample size, PM = Percent missing. Values in parentheses under Normality Level represent skew and kurtosis, respectively.

MONOTONIC PATTERN | **NON-MONOTONIC PATTERN**

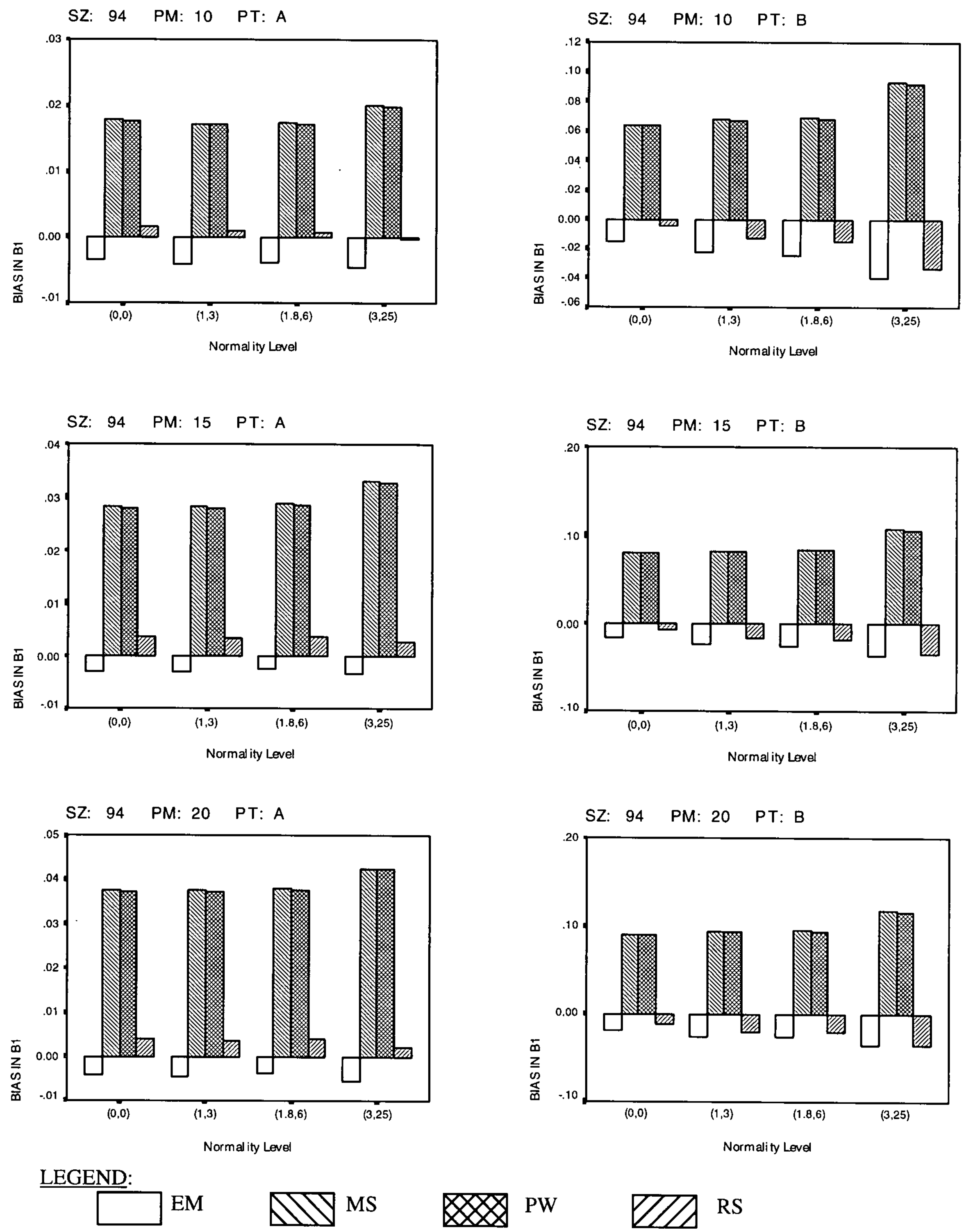


Figure B.2: Bias in b_1 under four predictors and low R^2 condition
Note: SZ = Sample size, PM = Percent missing, A = monotonic pattern, B = Non-monotonic pattern. Values in parentheses under Normality Level represent skew and kurtosis, respectively.

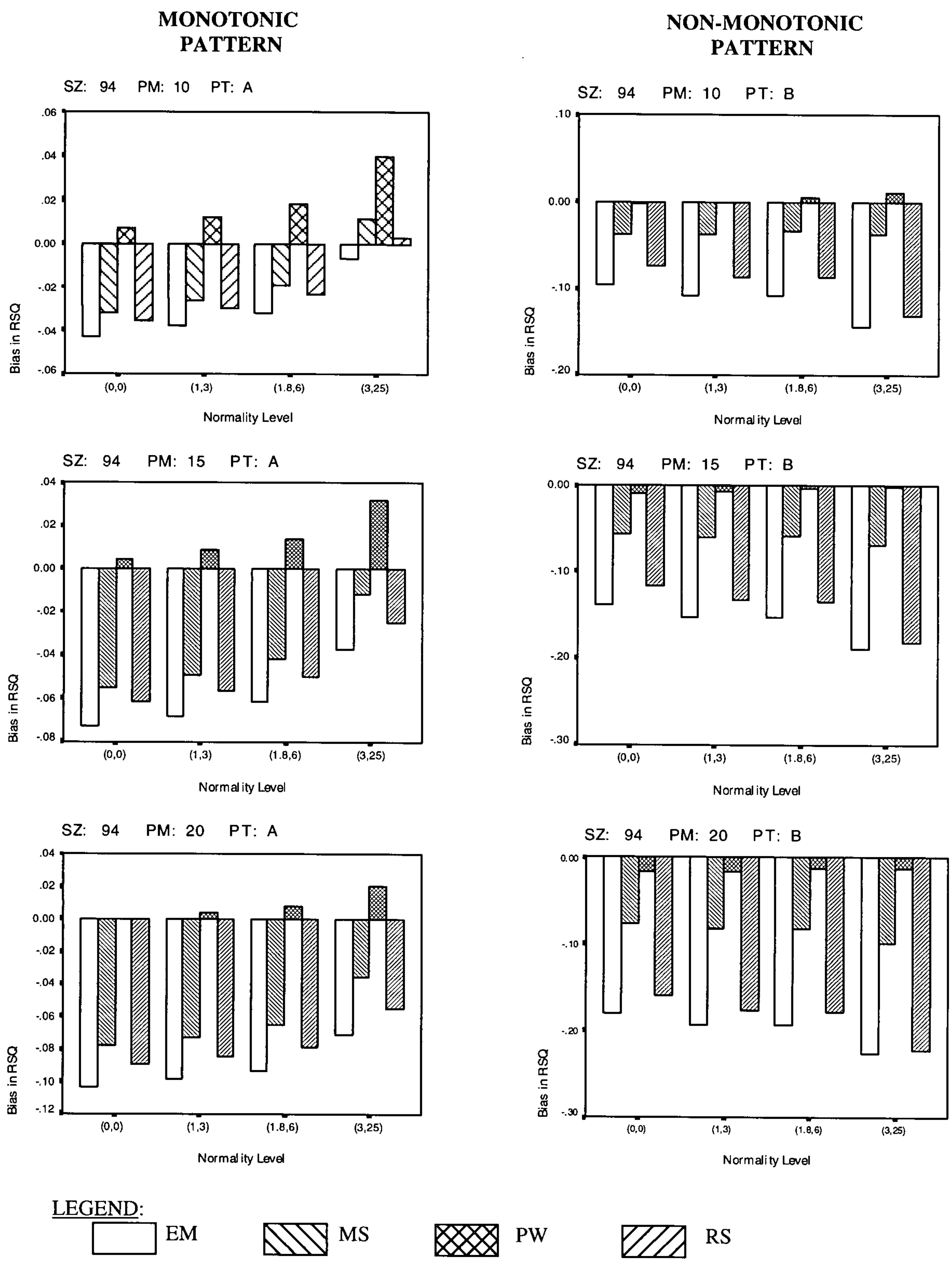


Figure B.3: Bias in $R^2_{estimate}$ under four predictors and low R^2 condition

Note: SZ = Sample size, PM = Percent missing. Values in parentheses under Normality Level represent skew and kurtosis, respectively.

MONOTONIC PATTERN | **NON-MONOTONIC PATTERN**

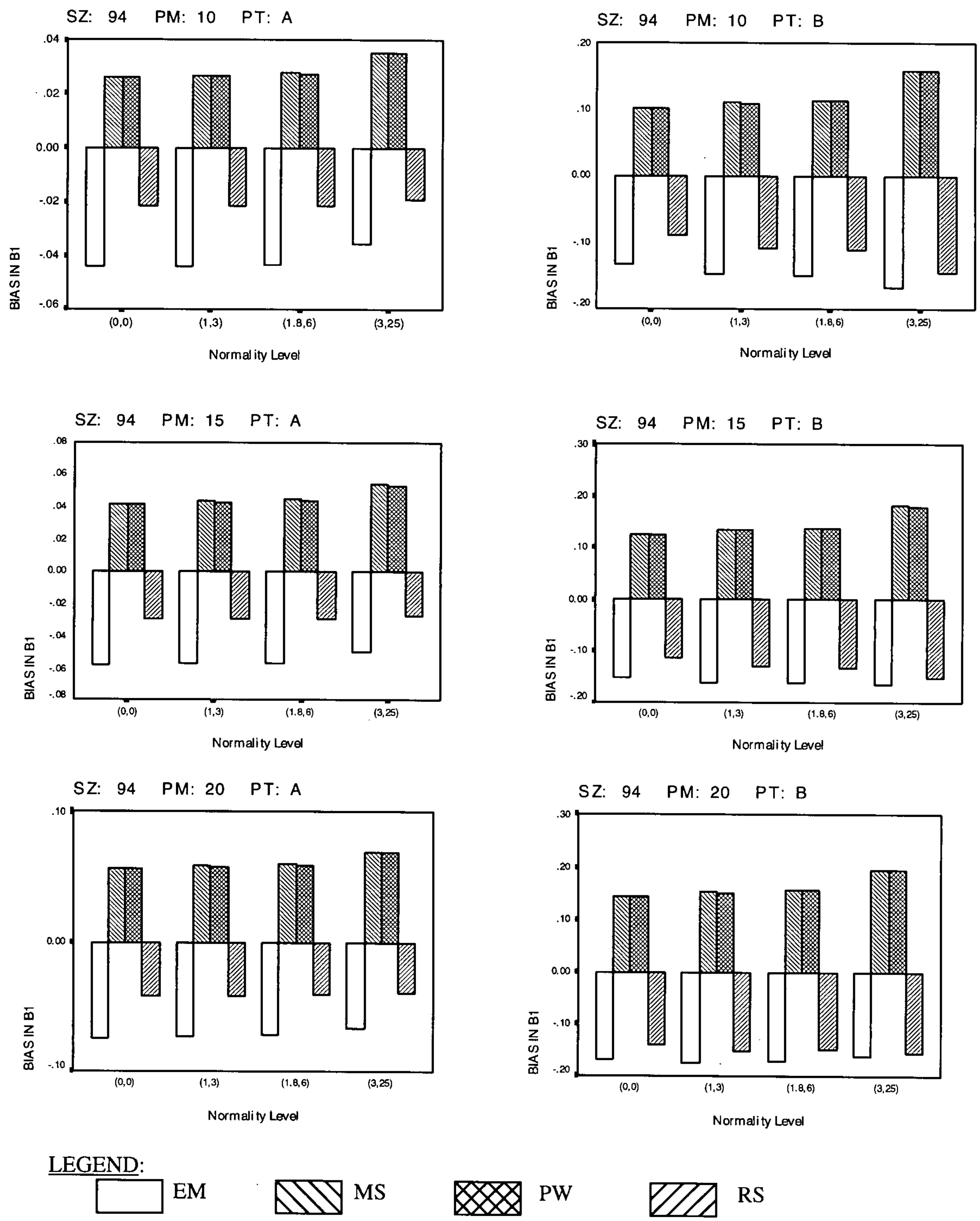


Figure B.4: Bias in b_1 under four predictors and high R^2 condition

Note: SZ = Sample size, PM = Percent missing. Values in parentheses under Normality Level represent skew and kurtosis, respectively.

MONOTONIC PATTERN **NON-MONOTONIC PATTERN**

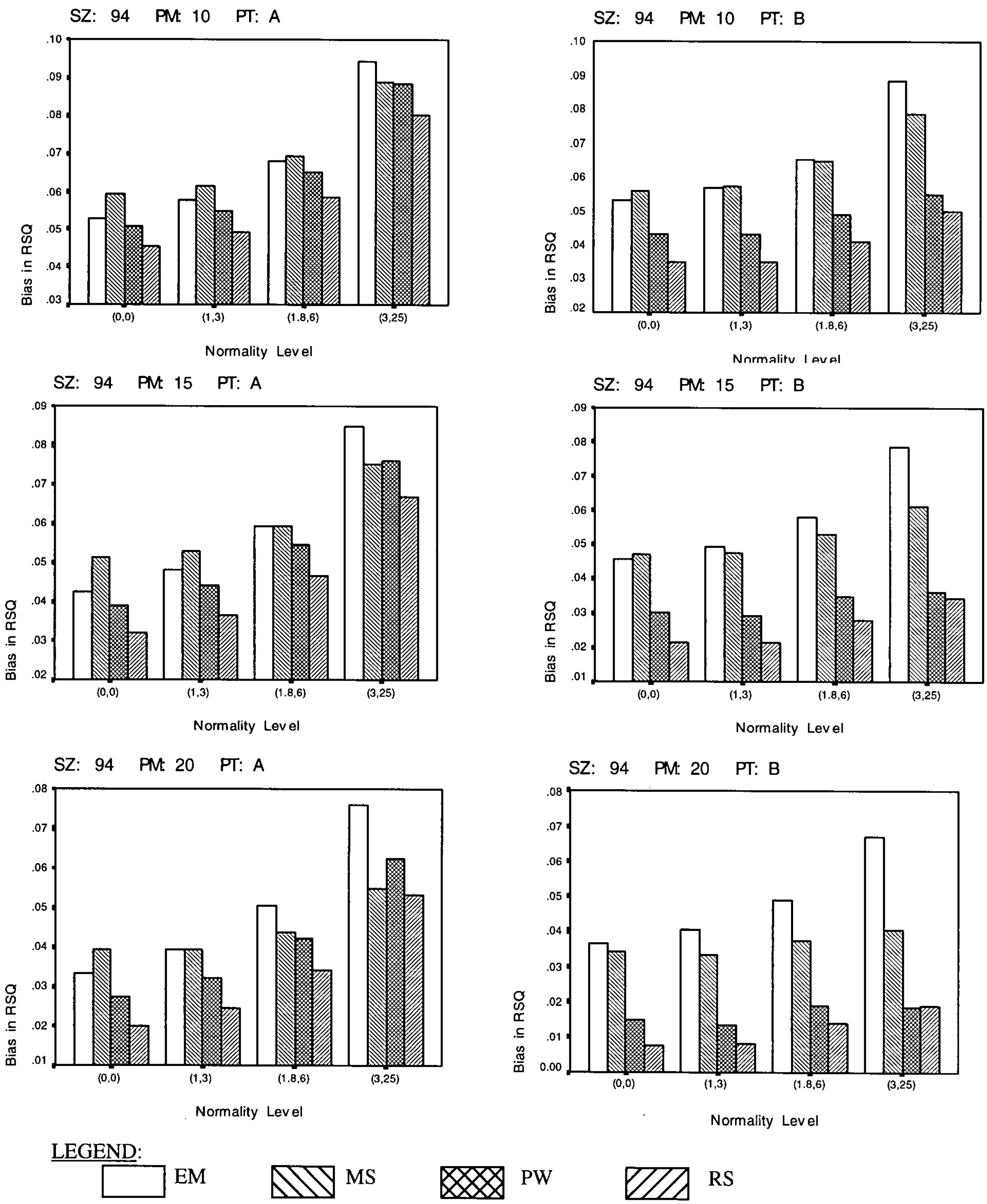


Figure B.5: Bias in $R^2_{estimate}$ under nine predictors and low R^2 condition

Note: SZ = Sample size, PM = Percent missing. Values in parentheses under Normality Level represent skew and kurtosis, respectively.

MONOTONIC PATTERN | **NON-MONOTONIC PATTERN**

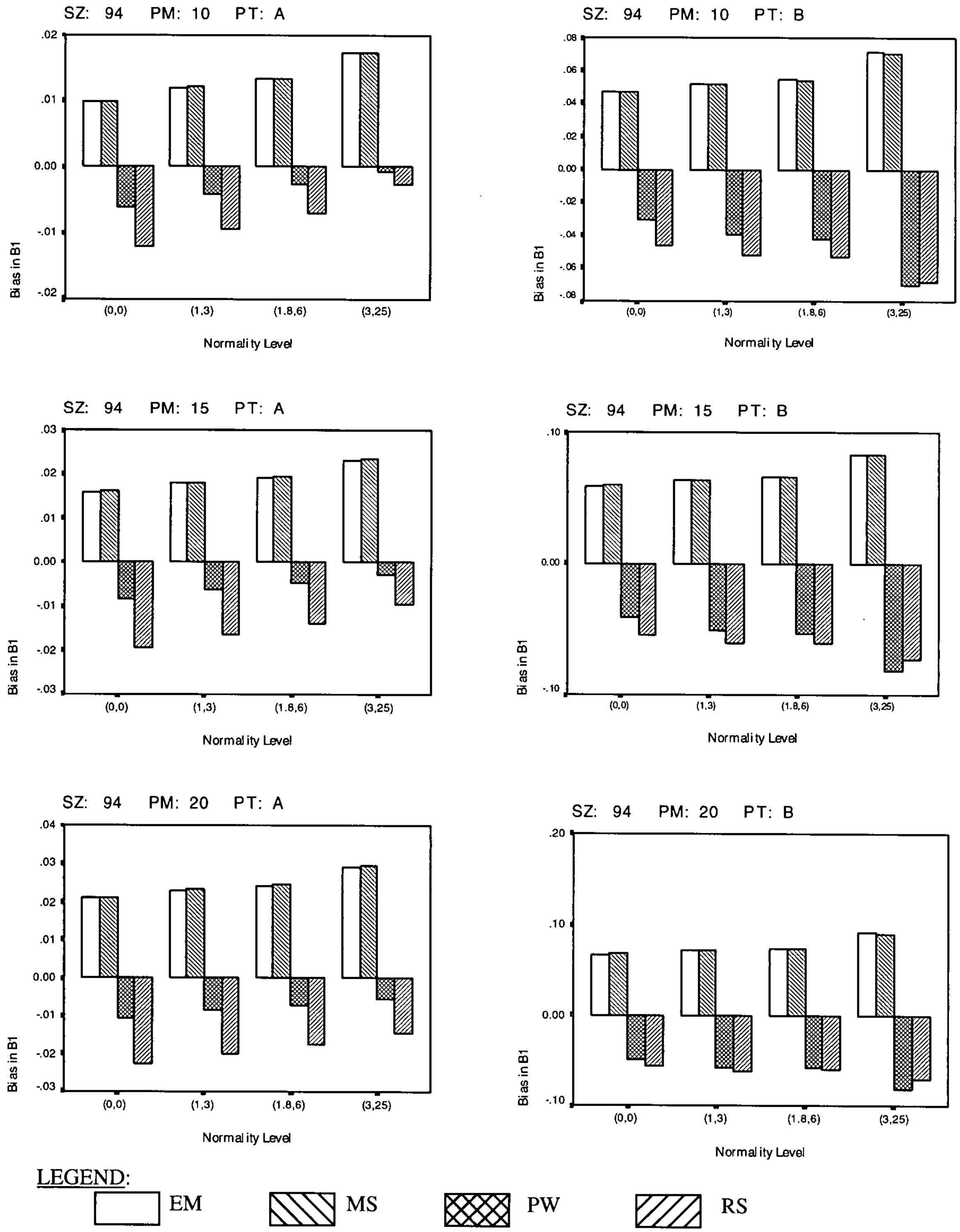


Figure B.6: Bias in b_1 under nine predictors and low R^2 condition

Note: SZ = Sample size, PM = Percent missing. Values in parentheses under Normality Level represent skew and kurtosis, respectively.

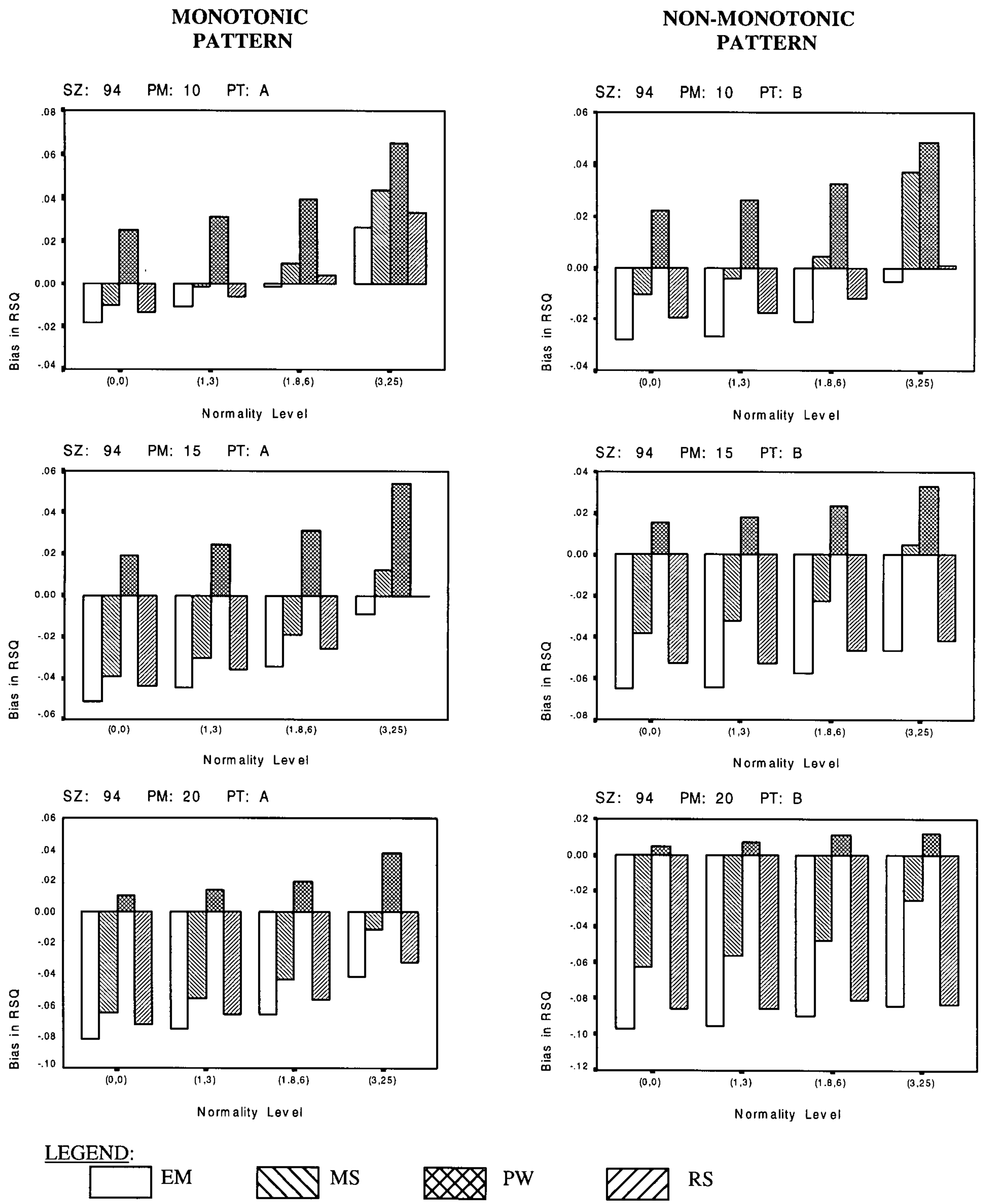


Figure B.7: Bias in $R^2_{estimate}$ under nine predictors and high R^2 condition

Note: SZ = Sample size, PM = Percent missing. Values in parentheses under Normality Level represent skew and kurtosis, respectively.

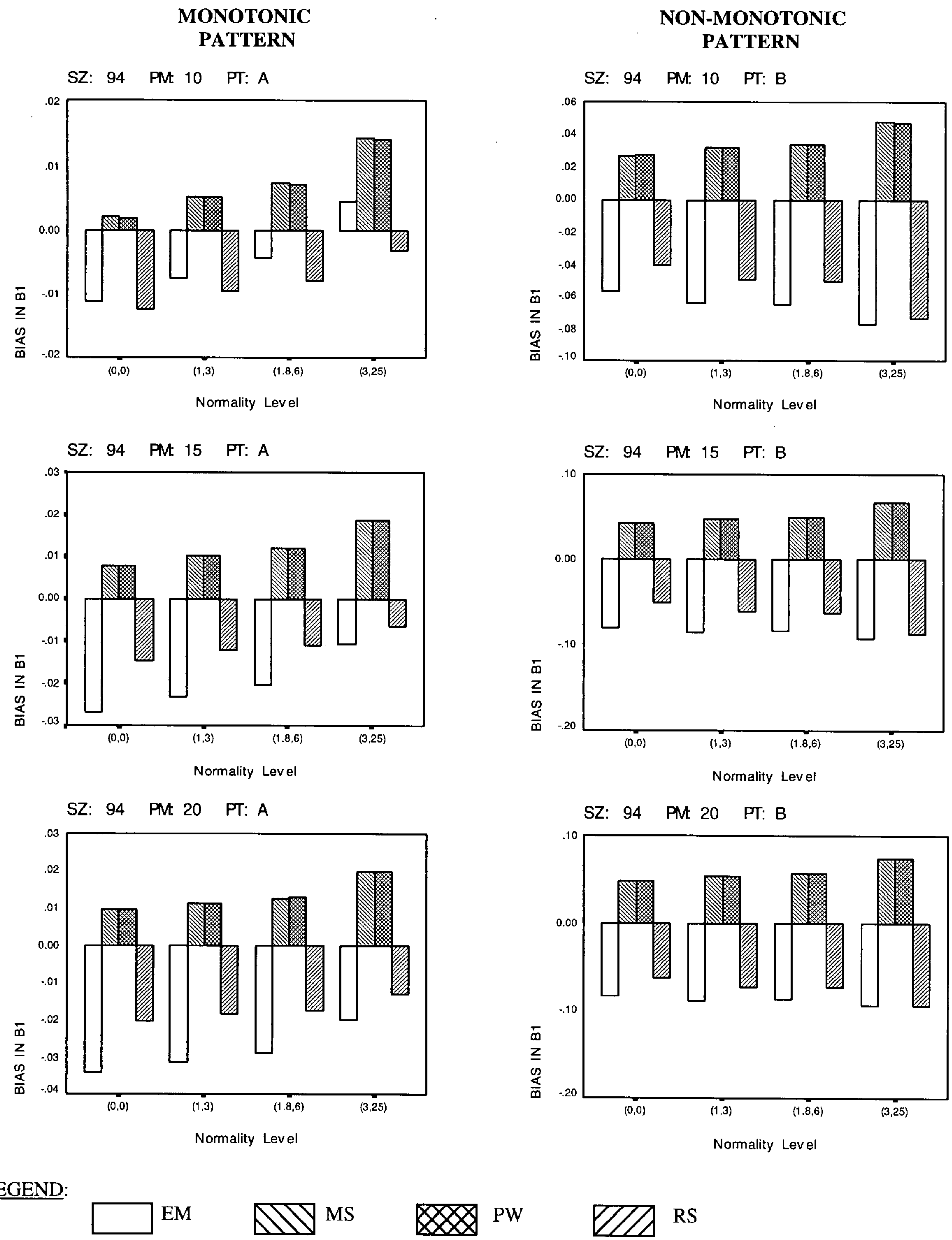


Figure B.8: Bias in b_1 under nine predictors and high R^2 condition

Note: SZ = Sample size, PM = Percent missing. Values in parentheses under Normality Level represent skew and kurtosis, respectively.

APPENDIX C

ARTICLES USED IN THE REVIEW

Aiello, J. R., & Kolb, K. J. (1995). Electronic performance monitoring and social context: Impact on productivity and stress. Journal of Applied Psychology, 80(3), 339-353.

Ainley, J., Foreman, J., & Sheret, M. (1991). High school factors that influence students to remain in school. Journal of Educational Research, 85(2), 69-80.

Ainley, M. D. (1993). Styles of engagement with learning: Multidimensional assessment of their relationship with strategy use and school achievement. Journal of Educational Psychology, 885(3), 395-405.

Alexander, P. A., Kulikowich, J. M., & Schulze, S. K. (1994). How subject-matter knowledge affects recall and interest. American Educational Research Journal, 31(2), 313-337.

Allen, L., Cipielewski, J., & Stanovich, K. E. (1992). Multiple indicators of children's reading habits and attitudes: Construct validity and cognitive correlates. Journal of Educational Psychology, 84(4), 489-503.

Arthur, W. (Jr)., Barrett, G. V., & Doverspike, D. (1990). Validation of an information-processing-based test battery for the prediction of handling accidents among petroleum product transport drivers. Journal of Applied Psychology, 75(6), 621-628.

Arvey, R. T., Landon, T. E., Nutting, S. M., & Maxwell, S. E. (1992). Development of physical ability tests for police officers: A construct validation approach. Journal of Applied Psychology, 7(6), 996-1009.

Avolio, B. J., & Waldman, D. A. (1990). An examination of age and cognitive test performance across job complexity and occupational types. Journal of Applied Psychology, 75(1), 43-50.

Baba, V. V. (1990). Methodological issues in modeling absence: A comparison of least squares and tobit analyses. Journal of Applied Psychology, 75(4), 428-486.

Barber, A. E., Hollenbeck, J. R., Tower, S. L., & Phillips, J. M. (1994). The effects of interview focus on recruitment effectiveness: A field experiment. Journal of Applied Psychology, 79(6), 886-896.

Barrick, M. R., & Mount, M. K. (1993). Autonomy as a moderator of the relationships between the big five personality dimensions and job performance. Journal of Applied Psychology, 78(1), 111-118.

Benbow, C. P., Arjmand, O., & Walberg, H. J. (1991). Educational productivity predictors among mathematically talented students. Journal of Educational Research, 84(4), 215-223.

Bennett, R. E., Gottesman, R. L., Rock, D., A., & Cerullo, F. (1993). Influence of behavior perceptions and gender on teachers' judgments of students' academic skill. Journal of Educational Psychology, 85(2), 347-356.

Benton, S. L., Sharp, J. M., Corkill, A. J., Downey, R. G., & Khramtsova, I. (1995). Knowledge, interest, and narrative writing. Journal of Educational Psychology, 87(1), 66-79.

Biddle, B. J., Bank, B. J., & Slavings, R. L. (1990). Modality of thought, campus experiences and the development of values. Journal of Educational Psychology, 82(4), 671-682.

Bisanz, G. L., Das, J. P., Varnhagen, C. K., & Henderson, H, R. (1992). Structural components of reading time and recall for sentences in narratives: Exploring changes with age and reading ability. Journal of Educational Psychology, 84(1), 103-114.

Bretz, R. D., & Thompsett, R. E. (1992). Comparing traditional and integrative learning methods in organizational training programs. Journal of Applied Psychology, 77(6), 941-951.

Bridgeman, B., & Wendler, C. (1991). Gender differences in predictors of college mathematics performance and in college mathematics course grades. Journal of Educational Psychology, 83(2), 275-284.

Britton, B. K., & Tesser, A. (1991). Effects of time-management practices on college grades. Journal of Educational Psychology, 83(3), 405-410.

Britton, B. K., Lani, D. V., Gulgoz, S., Sharp, L., & Glynn, S. M. (1991). Accuracy of learnability judgments for instructional texts. Journal of Educational Psychology, 83(1), 43-47.

Byrne, B., & Fielding-Barnsley, R. (1993). Evaluation of a program to teach phonemic awareness to young children: A 1-Year follow-up. Journal of Educational Psychology, 85(1), 104-111.

Caldas, S. J. (1993). Reexamination of input and process factor effects on public school achievement. Journal of Educational Research, 86(4), 206-214.

Chao, G. T., O'Leary-Kelly, A. M., Wolf, S., Klein, H. J., & Gardner, P. D. (1994). Organizational socialization: Its content and consequences. Journal of Applied Psychology, 79(5), 730-743.

Cleveland, J. N., & Shore, L. M. (1992). Self and supervisory perspectives on age and work attitudes and performance. Journal of Applied Psychology, 77(4), 469-484.

Cohen, E. G., & Lotan, R. A. (1995). Producing equal-status interaction in the heterogeneous classroom. American Educational Research Journal, 32(1), 99-120.

Conway, J. M., Jako, R. A., & Goodman, D. F. (1995). A meta-analysis of interrater and internal consistency reliability of selection interviews. Journal of Applied Psychology, 80(5), 565-579.

Cornell, D. G., Pelton, G. M., Bassin, L. E., Landrum, M., Ramsay, S. G., Cooley, M. R., Lynch, K. A., & Hamrick, E. (1990). Self-concept and peer status among gifted program youth. Journal of Educational Psychology, 82(3), 456-463.

Coulter, R. H., & Puito, M. B. (1995). Guilt appeals in advertising: What are their effects? Journal of Applied Psychology, 80(6), 697-705.

Crant, J. M. (1995). The proactive personality scale and objective job performance among real estate agents. Journal of Applied Psychology, 80(4), 532-537.

Cunningham, A. E., & Stanovich, K. E. (1990). Assessing print exposure and orthographic processing skill in children: A quick measure of reading experience. Journal of Educational Psychology, 82(4), 733-740.

Cunningham, A. E., & Stanovich, K. E. (1991). Tracking the unique effects of print exposure in children: Associations with vocabulary, general knowledge, and spelling. Journal of

Educational Psychology, 83(2), 264-274.

Dalton, D. R., & Mesch, D. J. (1991). On the extent and reduction of avoidable absenteeism: An assessment of absence policy provisions. Journal of Applied Psychology, 76(6), 810-817.

Day, J. D., & Cordon, L. A. (1993). Static and dynamic measures of ability: An experimental comparison. Journal of Educational Psychology, 85(1), 75-82.

Dee-Lucas, D., & Larkin, J. H. (1990). Organization and comprehensibility in scientific proofs, or "consider a particle p..". Journal of Educational Psychology, 82(4), 701-714.

Doran, L. I., Stone, V. K., Brief, A. P., & George, J. M. (1991). Behavioral intentions as predictors of job attitudes: The role of economic choice. Journal of Applied Psychology, 76(1), 40-45.

Dreher, G. F., & Bretz, R. D. (Jr.). (1991). Cognitive ability and career attainment: Moderating effects of early career success. Journal of Applied Psychology, 76(3), 392-397.

Dreher, M. J., & Zenge, S. D. (1990). Using metalinguistic awareness on first grade to predict reading achievement in third and fifth grade. Journal of Educational Research, 84(1), 13-21.

Duda, J. L., & Nicholls, J. G. (1992). Dimensions of achievement motivation in schoolwork and sport. Journal of Educational Psychology, 84(3), 290-299.

Dunegan, K. J. (1993). Framing, cognitive modes, and image theory: Toward an understanding of a glass half full. Journal of Applied Psychology, 78(3), 491-503.

Duran, B. J., & Weffer, R. E. (1992). Immigrants' aspirations, high school processes, and academic outcomes. American Educational Research Journal, 29(1), 163-181.

Durgunoglu, A. Y., Nagi, W. E., & Hancin-Bhatt, B. J. (1993). Cross-language transfer of phonological awareness. Journal of Educational Psychology, 85(3), 453-465.

Durrant, J. E., Cunningham, C. E., & Voelker, S. (1990). Academic, social, and general self-concepts of behavioral subgroups of learning disabled children. Journal of Educational Psychology, 82(4), 657-663.

Earley, P. C., & Erez, M. (1991). Time-dependency effects of goals and norms: The role of cognitive processing on motivational models. Journal of Applied Psychology, 76(5), 717-724.

Earley, P. C., & Lituchy, T. R. (1991). Delineating goal and efficacy effects: A test of three models. Journal of Applied Psychology, 76(1), 81-98.

Eden, D., & Zuk, Y. (1995). Sea-sickness as a self-fulfilling prophecy: Raising self-efficacy to boost performance at sea. Journal of Applied Psychology, 80(5), 628-635.

Edwards, J. R., & Baglioni, A. J. (Jr.). (1991). Relationship between Type A behavior pattern and mental and physical symptoms: A comparison of global and component measures. Journal of Applied Psychology, 76(2), 276-290.

Ethington, C. A. (1991). A test of a model of achievement behaviors. American Educational Research Journal, 28(1), 155-172.

Evans, G. W., & Carrere, S. (1991). Traffic congestion, perceived control, and psychophysiological stress among urban bus drivers. Journal of Applied Psychology, 76(5), 658-63.

Feldman, S. S., & Wentzel, K. R. (1990). Relations among family interaction patterns, classroom self-restraint, and academic achievement in preadolescent boys. Journal of Educational Psychology, 82(4), 813-819.

Fried, Y., & Tiegs, R. B. (1995). Supervisors' role conflict and role ambiguity differential relations with performance ratings of subordinates and the moderating effect of screening ability. Journal of Applied Psychology, 80(2), 282-291.

Friedman, I. A. (1995). Student behavior patterns contributing to teacher burnout. Journal of Educational Research, 88(5), 281-289.

Fuchs, L. S., Fuchs, D., & Bishop, N. (1992). Instructional adaptation for students at risk. Journal of Educational Research, 86(2), 70-84.

Furlong, P. R. (1993). Personal factors influencing informal reasoning of economic issues and the effect of specific instructions. Journal of Educational Psychology, 85(1), 171-181.

Ganster, D. C., Williams, S., & Poppler, P. (1991). Does training in problem solving improve the quality of group discussion? Journal of Applied Psychology, 76(3), 479-483.

Gellatly, I. R., & Meyer, J. P. (1992). The effects of goal difficulty on physiological arousal, cognitive and task performance. Journal of Applied psychology, 77(5), 694-704.

George, J. M. (1990). Personality, affect, and behavior in groups. Journal of Applied Psychology, 75(2), 107-116.

George, J. M. (1991). State or trait: Effects of positive mood on prosocial behaviors at work. Journal of Applied Psychology, 76(2), 299-307.

Gerhart, B., & Rynes, S. (1991). Determinants and consequences of salary negotiations by male and female MBA graduates. Journal of Applied Psychology, 76(2), 256-262.

Gibbons, F. X., Larsen-Helweg, M., & Gerrard, M. (1995). Prevalence estimates and adolescent risk behavior: Cross-cultural differences in social influence. Journal of Applied Psychology, 80(1), 107-121.

Gottfried, A. E. (1990). Academic intrinsic motivation in young elementary school children. Journal of Educational Psychology, 82(3), 525-538.

Graham, J. W., Marks, G., & Hansen, W. B. (1991). Social influence processes affecting adolescent substance use. Journal of Applied Psychology, 76(2), 291-298.

Grover, S. L. (1991). Predicting the perceived fairness of parental leave policies. Journal of Applied Psychology, 76(2), 247-255.

Haenggi, D., & Perfetti, C. A. (1992). Individual differences in reprocessing of text. Journal of Educational Psychology, 84(2), 182-192.

Hafner, A. L. (1993). Teaching-method scales and mathematics-class achievement: What works with different outcomes? American Educational Research Journal, 30(1), 71-94.

Hancock, D. R. (1994). Motivating adults to learn academic course content. Journal of Educational Research, 88(2), 102-108.

Harder, J. W. (1991). Equity theory versus expectancy theory: The case of major league baseball free agents. Journal of Applied Psychology, 76(3), 458-464.

Harrison, D. A. (1995). Volunteer motivation and attendance decisions: Competitive theory testing in multiple samples from a homeless shelter. Journal of Applied Psychology, 80(3), 371-385.

Hoge, D. R., Smit, E. K., & Hanson, S. L. (1990). School experiences predicting changes in Self-esteem of sixth and seventh-grade students. Journal of Educational Psychology, 82(1), 117-127.

Hollenbeck, J. R., Ilgen, D. R., Sego, D. J., Hedlund, J., Major, D. A., & Phillips, J. (1995). Multilevel theory of team decision making: Decision performance in teams incorporating distributed expertise. Journal of Applied Psychology, 80(2), 292-316.

Hossler, D., & Stage, F. K. (1992). Family and high school experience influences on the postsecondary educational plans of ninth-grade students. American Educational Research Journal, 29(2), 425-451.

Jackson, L. A., Gardner, P. D., & Sullivan, L. A. (1992). Explaining gender difference in self-pay expectations: Social comparison standards and perceptions of fair play. Journal of Applied Psychology, 77(5), 651-663.

Jackson, S. E., Brett, J. F., Sessa, V. I., Cooper D. M., Julin, J. A., & Peyronnin, K. (1991). Some differences make a difference: Individual dissimilarity and group heterogeneity as correlates of recruitment promotions and turnover. Journal of Applied Psychology, 76(5), 675-689.

Jacobs, J. E. (1991). Influence of gender stereotypes on parent and child mathematics attitudes. Journal of Educational Psychology, 83(4),518-527.

Jurden, F. H. (1995). Individual differences in working memory and complex cognitions. Journal of Educational Psychology, 87(1), 93-102.

Kennedy, E. (1992). A multilevel study of elementary male black students and white students.

Journal of Educational Research, 86(2), 105-118.

Kinicki, A. J., Hom, P. W., Lockwood, C. A., & Griffeth, R. W.W. (1990). Interviewer predictions of applicant qualifications and interviewer validity: Aggregate and individual analyses. Journal of Applied Psychology, 75(5), 477-488.

Kistner, J., Metzler, A., Gatlin, D., & Risi, S. (1993). Classroom racial proportions and children's peer relations: Race and gender effects. Journal of Educational Psychology, 85(3), 446-452.

Klein, J. D., & Keller, J. M. (1990). Influence of student ability, locus of control, and type of instructional control on performance and confidence. Journal of Educational Research, 83(3), 140-150.

Konovsky, M. A., & Cropanzano, R. (1991). Perceived fairness of employee drug testing as a predictor of employee attitudes and job performance. Journal of Applied Psychology, 76(5), 698-707.

Kupermintz, H., Ennis, M. M., Hamilton, L. S., Talbert, J. E., & Snow, R. E. (1995). Enhancing the validity and usefulness of large-scale educational assessments: I.NELS:88 Mathematics Achievement. American Educational Research Journal, 32(2), 525-581.

Lee, V. E., & Marks, H. M. (1990). Sustained effects of the single-sex secondary school experience on attitudes, behaviors, and values in college. Journal of Educational Psychology, 82(3), 578-592.

Lehman, W. E. K., & Simpson, D. D. (1992). Employee substance use and on-the-job behaviors. Journal of Applied Psychology, 77(3), 309-321.

Lehrer, R., & Littlefield, J. (1993). Relationships among cognitive components in logo learning and transfer. Journal of Educational Psychology, 85(2), 317-330.

Licata, J. W., & Teddlie, C. B. & Greenfield, W. D. (1990). Principal vision, teacher sense of autonomy, and environmental robustness. Journal of Educational Research, 84(2), 93-99.

Lowe, N., & Bickel, R. (1993). Computer-assisted instruction in Appalachia's postsecondary schools. Journal of Educational Research, 87(1), 46-52.

Mael, F. A. (1995). Staying afloat: Within-group swimming proficiency for whites and blacks. Journal of Applied Psychology, 80(4), 479-490.

Major, D. A., Kozlowski, S. W. J., Chao, G. T., & Gardner, D. (1995). A longitudinal investigation of newcomer expectations, early socialization outcomes, and the moderating effects of role development factors. Journal of Applied Psychology, 80(3), 418-431.

Maple, S. A., & Stage, F. K. (1991). Influences on the choice of math/science major by gender and ethnicity. American Educational Research Journal, 28(1), 37-60.

Marsh, H. W. (1990). Two-parent, stepparent, and single-parent families: Changes in achievement attitudes, and behaviors during the last two years of high school. Journal of Educational Psychology, 82(2), 327-340.

Marsh, H. W. (1991). Failure of high-ability high schools to deliver academic benefits commensurate with their students' ability levels. American Educational Research Journal, 28(2), 445-480.

Marsh, H. W. (1992). Extracurricular activities: Beneficial extension of the traditional curriculum or subversion of academic goals? Journal of Educational Psychology, 84(4), 553-562.

Martocchio, J. J. (1994). Effects of conceptions of ability on anxiety, self-efficacy, and learning in training. Journal of Applied Psychology, 79(6), 819-825.

Mathews, G., Jones, D. M., & Chamberlain, A. G. (1992). Predictors of individual differences in mail-coding skills and their variation with ability level. Journal of Applied Psychology, 77(4), 406-418.

Maurer, T. J., & Tarulli, B. A. (1995). Investigation of perceived environment, perceived outcome, and person variables in relationship to voluntary development activity by employees. Journal of Applied Psychology, 79(1), 3-14.

McGregor, J. (1993). Effectiveness of role playing and antiracist teaching in reducing student prejudice. Journal of Educational Research, 86(4), 215-226.

McIntyre, L. L. (1990). Teacher standards and gender: Factors in special education referral? Journal of Educational Research, 83(3), 166-172.

McLain, K. V. M., Gridley, B. E., & McIntosh, D. (1991). Value of a scale used to measure metacognitive reading awareness. Journal of Educational Research, 85(2), 81-87.

McNeely, B. L., & Meglino, B. M. (1994). The role of dispositional and situational antecedents in prosocial organizational behavior: An examination of the intended beneficiaries of prosocial behavior. Journal of Applied Psychology, 79(6), 836-844.

Meisels, S. J., & Liaw, F (1993). Failure in grade: Do retained students catch up? Journal of Educational Research, 87(1), 69-77.

Melamed, S., Ben-Avi, I., Luz, J., & Green, M. S. (1995). Objective and subjective work monotony: Effects on job satisfaction, psychological distress, and absenteeism in blue-collar workers. Journal of Applied Psychology, 80(1), 29-42.

Mellor, S. (1995). Gender composition and gender representation in local unions: Relationships between women's participation in local office and women's participation in local activities. Journal of Applied Psychology, 80(6), 706-720.

Mevarech, Z. R. (1991). Learning mathematics in different mastery environments. Journal of Educational Research, 84(4). 225-231.

Mills, J. R., & Jackson, N. E. (1990). Predictive significance of early giftedness: The case of precocious reading. Journal of Educational Psychology, 82(3), 410-419

Morrison, E. W. (1993). Longitudinal study of the effects of information seeking on newcomer socialization. Journal of Applied Psychology, 78(2), 173-183.

Motowildo, S. J., Carter, G. W., Dunnette, M. D., Tippins, N., Werner, S., Burnett, J. R., & Vaughan, M. J. (1992). Studies of the structured behavioral interview. Journal of Applied Psychology, 77(5), 571-587.

Mount, M. K., Barrick, M. R., & Strauss, J. P. (1994). Validity of observer ratings of the big five personality factors. Journal of Applied Psychology, 79(2), 272-280.

Newby, T. J. (1991). Classroom motivation: Strategies of first-year teachers. Journal of Educational Psychology, 83(2), 195-200.

Newman, R. S., & Schwager, M. T. (1995). Students' help seeking during problem solving: Effects of grade, goal, and prior achievement. American Educational Research Journal, 32(2), 352-376.

Newton, T., & Keenan, T. (1991). Further analyses of the dispositional argument in organizational behavior. Journal of Applied Psychology, 76(6), 781-787.

Nist, S. L., Simpson, M. L., Olejnik, S., & Mealey, D. L. (1991). The relation between self-selected study processes and test performance. American Educational Research Journal, 28(4), 849-874.

Noe, R. A., & Wilk, S. L. (1993). Investigation of the factors that influence employees' participation in development activities. Journal of Applied Psychology, 78(2), 291-302.

Oldham, G. R., Cummings, A., Mischel, L. J., Schmdtke, J. M., & Zhou, J. (1995). Listen while you work? Quasi-experimental relations between personal-stereo headset use and employee work responses. Journal of Applied Psychology, 80(5), 547-564.

Ostroff, C. (1992). The relationship between satisfaction, attitudes, and performance: An organizational level analysis. Journal of Applied Psychology, 77(6), 963-974.

Paiker, D., Manstead, S. R., Stradling, S. G., Reason, J. T., & Baxter, J. S. (1992). Intention to commit driving violations: An application of the theory of planned behavior. Journal of Applied Psychology, 77(1), 94-101.

Peng, S. S., & Wright, D. (1994). Explanation of academic achievement of Asian American students. Journal of Educational Research, 87(6), 346-352.

Piel, J. A. (1990). Unmasking sex and social class differences in childhood aggression: The case for language maturity. Journal of Educational Research, 84(2), 100-106.

Pintrich, P. R., & De Groot, V. D. (1990). Motivational and self-regulated learning components of classroom academic performance. Journal of Educational Psychology, 82(1), 33-40.

Pittman, R. B. (1991). Social factors, enrollment in vocational/technical courses, and high school dropout rates. Journal of Educational Research, 84(5), 288-295.

Podell, D. M., & Soodak, L. C. (1993). Teacher efficacy and bias in special education referrals. Journal of Educational Research, 86(4), 247-253.

Quinones, M. A. (1995). Pretraining context effects: Training assignment as feedback. Journal of Applied Psychology, 80(2), 226-238.

Reynolds, A. J. (1991). Early schooling of children at risk. American Educational Research Journal, 28(2), 392-422.

Rice, R. W., Gentile, D. A., & McFarlin, D. B. (1991). Facet importance and job satisfaction. Journal of Applied Psychology, 76(1), 31-39.

Royer, J. M., Marchant, G. M., Lovejoy, D. A., & Sinatra, G. M. (1990). The prediction of college course performance from reading comprehension performance: Evidence for general and specific prediction factors. American Educational Research Journal, 27(1), 158-179.

Rumberger, R. S. (1995). Dropping out of middle school: A multilevel analysis of students and schools. American Educational Research Journal, 32(3), 583-625.

Sacket, P. R., DuBois, C. L. Z., & Noe, A. W. (1991). Tokenism in performance evaluation: The effects of workgroup representation on male-female and white-black differences in performance ratings. Journal of Applied Psychology, 76(2), 263-267.

Saks, A. M. (1995). Longitudinal field investigation of the moderating and mediating effects of self-efficacy on the relationship between training and newcomer adjustment. Journal of Applied Psychology, 80(2), 211-225.

Schaubroeck, J., & Ganster, D. C. (1993). Chronic demands and responsivity to challenge. Journal of Applied Psychology, 78(1), 73-85.

Schaubroeck, J., May, D., & Brown, F. (1994). Procedural justice explanations and employee reactions to economic hardship: A field experiment. Journal of Applied Psychology, 79(3), 455-460.

Schneider, B., Hanges, P. J., Goldstein, H. W., & Braverma, E. P. (1994). Do customer service perceptions generalize? The case of student and chair ratings of faculty effectiveness. Journal of Applied Psychology, 79(5), 685-690.

Schommer, M., Crouse, A., & Rhodes, N. (1992). Epistemological beliefs and mathematical text comprehension: Believing it is simple does not make it so. Journal of Educational Psychology, 84(4), 435-443.

Shaw, J., & Reyes, P. (1992). School cultures: Organizational value orientation and commitment. Journal of Educational Research, 86(2), 295-302.

Shore, T. H., Shore, L. M., & Thornton, G. C. (III). (1992). Construct validity of self and peer evaluations of performance dimensions in an assessment center. Journal of Applied Psychology, 77(1), 42-54.

Short, E. J., Schatschneider, C. W., & Friebert, S. E. (1993). Relationship between memory and metamemory performance: A comparison of specific and general strategy. Journal of Educational Psychology, 85(3), 412-423.

Skaalvik, E. M. (1990). Gender differences in general academic self-esteem and in success expectations on defined academic problems. Journal of Educational Psychology, 82(3), 593-598.

Skinner, E. A., Wellborn, J. G., & Connel, J. P. (1990). What it takes to do well in school and whether I've got it: A process model of perceived control and children's engagement and achievement in school. Journal of Educational Psychology, 82(1), 22-32.

Smith, J. B., Maehr, M. L., & Midgley, C. (1992). Relationship between personal and contextual characteristics and principal' administrative behaviors. Journal of Educational Research, 8(2), 111-118.

Smith, M. C. (1990). A longitudinal investigation of reading attitude development from childhood to adulthood. Journal of Educational Research, 83(4), 215-219.

Solomon, I. (1994). Analogical transfer and "functional fixedness" in the science classroom. Journal of Educational Research, 87(6), 371-377.

Sparks, R. L., Ganschow, L., & Palton, J. (1995). Prediction of performance in first-year foreign language courses: Connections between native and foreign language learning. Journal of Educational Psychology, 87(4), 638-655.

Spector, J. E. (1992). Predicting progress in beginning reading: Dynamic assessment of phonemic awareness. Journal of Educational Psychology, 84(3), 353-363.

Stanovich, K. E., & Cunningham, A. E. (1993). Where does knowledge come from? Specific associations between print exposure and information acquisition. Journal of Educational Psychology, 85(2), 211-229.

Stanovich, K. E., & Siegel, L. S. (1994). Phenotypic performance profile of children with reading disabilities: A regression-based test of the phonological-core variable-difference model. Journal of Educational Psychology, 86(1), 24-53.

Stevens, C. K., & Kristof, A. L. (1995). Making the right impression: A field study of applicant impression management during interviews. Journal of Applied Psychology, 80(5), 587-606.

Stipek, D. J., & Gralinski, J. H. (1991). Gender differences in children's achievement-related beliefs and emotional responses to success and failure in mathematics. Journal of Educational Psychology, 83(3), 361-371.

Stricker, L. J., Rock, D. A., & Burton, N. W. (1993). Sex differences in predictions of college grades from scholastic aptitude test scores. Journal of Educational Psychology, 85(4), 710-718.

Stroh, L. K., Brett, J. M., & Reilly, A. H. (1992). All the right stuff: A comparison of female and male managers' career progression. Journal of Applied Psychology, 77(3), 251-260.

Talaga, J. A., & Beehr, T. A. (1995). Are there gender differences in predicting retirement decisions? Journal of Applied Psychology, 80(1), 16-28.

Tannenbaum, S. I., Mathieu, J. E., Salas, E., & Cannon-Bowers, J. A. (1991). Meeting trainees' expectations: The influence of training fulfillment on the development of commitment, self-efficacy, and motivation. Journal of Applied Psychology, 76,759-769.

Taylor, B. M., Frye, B. J., & Maruyama, G. M. (1990). Time spent reading and reading growth. American Educational Research Journal, 27(2), 351-362.

Tepper, B. J. (1994). Investigation of general and program-specific attitudes toward corporate drug-testing policies. Journal of Applied Psychology, 79(3), 392-401.

Thomas, D. C., & Ravlin, E. C. (1995). Responses of employees to cultural adaptation by a foreign manager. Journal of Applied Psychology, 80(1), 133-146.

Thomas, L. T., & Ganster, D. C. (1995). Impact of family-supportive work variables on work-family conflict and strain: A control perspective. Journal of Applied Psychology, 80(1), 6-15.

Thompson, M. S., Entwisle, D. R., Alexander, K. L., & Sundias, M. J. (1992). The influence of family composition on children's' conformity to the student role. American Educational Research Journal, 29(2), 405-424.

Thomson, W. C., & Wendt, J. C. (1995). Contribution of hardiness and school climate to alienation experienced by student teachers. Journal of Educational Research, 88(5), 269-274.

Tirre, W. C., & Pena, C. M. (1992). Investigation of functional working memory in the reading span test. Journal of Educational Psychology, 84(4), 462-472.

Trevino, L. K., & Youngblood, S. A. (1990). Bad apples in bad barrels: A causal analysis of ethical decision-making behavior. Journal of Applied Psychology, 75(4), 378-385.

Tubbs, M. E. (1993). Commitment as a moderator of the goal-performance relation: A case for clearer construct definition. Journal of Applied Psychology, 78(1), 86-97.

Turban, D. B., & Keon, T. L. (1993). Organizational attractiveness: An interactionist perspective.

Journal of Applied Psychology, 78(2), 184-193.

Vance, R. J., & Colella, A. (1990). Effects of two types of feedback on goal acceptance and personal goals. Journal of Applied Psychology, 75(1), 68-76.

Vancouver, J. B., Millsap, R. E., & Peters, P. A. (1994). Multilevel analysis of organizational goal congruence. Journal of Applied Psychology, 79(5), 666-679.

Vliert, E., Euwema, M. C., & Huismans, S. E. (1995). Managing conflict with a surbodinate or a superior: Effectiveness of conglomerated behavior. Journal of Applied Psychology, 80(2), 271-281.

Walton, P. D. (1995). Rhyming ability, phoneme identity, letter-sound knowledge, and the use of orthographic analogy by prereaders. Journal of Educational Psychology 87(4), 587-597.$

Wang, A. Y., & Thomas, M. H. (1995). Effects of keywords on long-term retention: Help or hindrance? Journal of Educational Psychology, 87(3), 468-475.

Ward, T. H., & Clark, H. T. (III). (1991). A reexamination of public- versus private school achievement: The case for missing data. Journal of Educational Research, 84(3), 153-163.

Webb, N. M., Troper, J. D., & Fall, R. (1995). Constructive activity and learning in collaborative small groups. Journal of Educational Psychology, 87(3), 406-423.

Weed, K., Ryan, E. B. & Day, J. (1990). Metamemory and attributions as mediators of strategy use and recall. Journal of Educational Psychology, 82(4), 849-855.

Weingart, L. R. (1992). Impact of group goals. task component complexity, effort, and planning on group performance. Journal of Applied Psychology, 77(5), 682-693.

Weishew, N., & Peng, S. S. (1993). Variables predicting students' problem behaviors. Journal of Educational Research, 87(1), 5-17.

Wentzel, K. R. (1993). Does being good make the grade? Social behavior and academic competence in middle school. Journal of Educational Psychology, 85(2), 357-364.

Wentzel, K. R. (1994). Relations of social goal pursuit to social acceptance, classroom behavior, and perceived social support. Journal of Educational Psychology, 86(2), 173-182.

Whitney, P., Kellas, G., & Ferraro, F. R. (1990). Reading ability and semantic processing: Isolating speed of access. Journal of Educational Psychology, 82(3), 479-485.

Wilhite, S. C. (1990). Self-efficacy, locus of control, self-assessment of memory ability and study activities as predictors of college course achievement. Journal of Educational Psychology, 82(4), 696-700.

Williams, K. J., Suls, J., Alliger, G. M., Learner, S. M., & Wan, C. K. (1991). Multiple role juggling and daily mood states in working mothers: An experience sampling study. Journal of Applied Psychology, 76(5), 664-674.

Wong, C., & Campion, M. A. (1991). Development and test of a task level model of motivational job design. Journal of Applied Psychology, 76(6), 825-837.

Woolfolk, A. E., & Hoy, W. K. (1990). Prospective teachers' sense of efficacy and beliefs about control. Journal of Educational Psychology, 82(1), 81-91.

Wright, P. M., George, J. M., Farnsworth, S. R., & McMahan, G. C. (1993). Productivity and extra-role behavior: The effects of goals and incentives on spontaneous helping. Journal of Applied Psychology, 78(3), 374-381.

Zalesny, M. D. (1990). Rater confidence and social influence in performance appraisals. Journal of Applied Psychology, 75(3), 274-289. 117-127.