1994

Topics In Forward Stepwise Logistic Regression

Kang-in Lee

TOPICS IN FORWARD STEPWISE LOGISTIC REGRESSION

by

Kang-In Lee

Department of Epidemiology and Biostatistics

Submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Faculty of Graduate studies

The University of Western Ontario

London, Ontario

November 1993

ISBN 0-315-90546-8

ABSTRACT

In this dissertation, five topics related to the process and prediction of forward stepwise logistic regression are investigated.

Forward stepwise logistic regression is involved with selection and stopping criteria. Seven selection criteria are used: the likelihood ratio statistic, Lawless and Singhal (1978)'s statistic, the Wald statistic, the score statistic, Peduzzi, Hardy, and Holford (1980)'s statistic, Lee and Koval's statistic (LK), and a sweep operator's statistic (SW). Five stopping criteria are used: χ^2 test based on a fixed α level, minimum value of $\hat{ERR}$, minimum value of the C_p statistic (Hosmer, 1989), minimum value of the Akaike information criterion (Akaike, 1974), and minimum value of Schwarz's criterion (Schwarz, 1978).

Apparent error rate (ARR) tends to underestimate true error rate (ERR). In our study, estimated true error rate ($\hat{ERR}$) is obtained by $\hat{ERR} = ARR + \hat{\omega}$, where $\hat{\omega}$ is Efron (1986)'s parametric estimate of bias for ARR.

We use Monte Carlo simulation with both multivariate normal and multivariate binary independent variables; we implement the simulation with SAS/IML programs. We then analyze the experimental design to see which factors of the distribution of independent variables affect various outcomes.

As a result, we recommend the best α level for the $\chi^2_{(\alpha)}$ stopping criterion. Second, we compare the order of variables selected by different selection criteria. Third, we investigate the effects of different structures of predictor variables on ARR, $\hat{\omega}$, and $\hat{ERR}$. Fourth, we compare the sizes of subset models determined by different stopping criteria. Finally, we compare the performances of selection and stopping criteria in terms of $\hat{ERR}$.

This thesis is dedicated to my wife, Keum-Joo,

and to my two daughters, Eun-Hae and Eun-Joo.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor, Dr. John J. Koval, for his insightful guidance, consistent encouragement, and invaluable help throughout the course of this thesis. I consider myself very fortunate to have been his student. I would also like to thank my advisor and chair of the Department of Epidemiology and Biostatistics, Dr. Allan Donner, for his generous help with a variety of situations during my studies in the department.

I am very grateful to Dr. Yang Mao at the Laboratory Centre for Disease Control in Ottawa, who has done so much for me over the past several years.

I would like to thank Drs. Serge B. Provost and Thomas H. Wonnacott in the Department of Statistical and Actuarial Sciences for their friendship and moral support during my studies at U.W.O..

I would like to thank Dr. David W. Hosmer, Jr. at University of Massachusetts, U.S.A., who supplied the valuable references for my thesis.

I wish to express my sincere thanks to my brothers (Kang-Suk, Kang-Bok, and Kang-Woo), sisters (Kang-Myung, Kang-Sook, and Kang-Ae), brothers in law (Wang-Suk and Wang-Il), for their financial and moral support.

I am deeply indebted to my wife, Keum-Joo, and my two daughters, Eun-Hae and Eun-Joo, for their love that helped me through trying times during my studies.

Finally, my heartfelt love and thanks go to my mother and father and my mother in law and father in law for their unselfish sacrifice that have been a source of strength during my studies.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION AND REVIEW OF THE LITERATURE

1.1 Introduction of the logistic regression model

Regression methods have played a major role in data analysis concerned with describing the relationship between a response (outcome or dependent) variable and one or more predictor (explanatory or independent) variables. It is often the case that the response variable is dichotomous, taking on one of two possible values. The logistic regression model is a standard method of analysis in this situation. For example, in the Framingham heart study Truett, Cornfield, and Kannel (1967) used the logistic regression model to provide a multivariate analysis of the risk of coronary heart disease in Framingham. The traditional analytic method of a multiple cross-classification quickly becomes impracticable as the number of variables to be investigated increases. The logistic regression model provides a more powerful form of analysis than inspection of the results of the multiple cross-classification.

Let Y be a dichotomous response variable and let **X** be a p×1 vector of predictor variables. The logistic model specifies that the probability of Y=1 on a set of variables $\mathbf{X}' = (x_1, x_2, \ldots, x_p)$ in the following way:

$$\pi(\mathbf{X}) = \Pr(Y=1 \mid \mathbf{X}) \qquad (1.1.1)$$

$$= \frac{1}{1 + e^{-(\beta_0 + \sum_{i=1}^{p} \beta_i X_i)}}$$

$$= \frac{e^{(\beta_0 + \sum_{i=1}^{p} \beta_i X_i)}}{1 + e^{(\beta_0 + \sum_{i=1}^{p} \beta_i X_i)}} .$$

Then the logit transformation of $\Pr(Y=1 \mid \mathbf{X})$ is defined as

$$\log\left[\frac{\Pr(Y=1 \mid \mathbf{X})}{\Pr(Y=0 \mid \mathbf{X})}\right] = \log\left[\frac{\Pr(Y=1 \mid \mathbf{X})}{1-\Pr(Y=1 \mid \mathbf{X})}\right] \tag{1.1.2}$$

$$= \beta_0 + \sum_{i=1}^{p} \beta_i X_i .$$

This transformation has the desirable property of transforming the (0,1) interval for $\pi(\mathbf{X})$ to $(-\infty,+\infty)$.

Although this model is formulated in terms of the analysis of cohort studies, it may be applied directly to the analysis of case-control studies. Furthermore, the interpretation of the model parameters is the same with exception of the intercept for both study designs (see e.g. Cornfield, 1951; Farewell, 1979; Prentice and Pyke, 1979).

The general shape of the logistic function, which may be considered as a basic model for dose-response relationships, is the S-shaped curve. The higher the cholesterol for instance, the greater the incidence of coronary heart disease. In addition to simply postulating a logistic dose-response relationship between a set of variables and the probability of disease, several different assumptions about the variables $x_1, \ldots, x_p$ in diseased and nondiseased populations have been shown to lead to the logistic model. These include

1) Assumption of multivariate normality with equal covariance matrices (Cornfield, Gordon, and Smith, 1961).

2) Multivariate independent dichotomous variables (Anderson, 1972).

3) Discrete variables following a loglinear model with second and higher order effects the same in each population (Birch, 1963).

4) A combination of 1) and 3).

There is, in general, no theoretical justification for the form (1.1.2). Many other choices for a transformation of $\pi(X)$ are possible, the most common being the inverse normal cumulative distribution function, leading to the probit model. The probit model, based on the probability integral transformation, has been commonly used to represent dose-response relationships in biological assay (Finney, 1971). There is, however, little practical difference between the logistic model and the probit model. The principal appeal to the logistic model and its widespread application to the analysis of epidemiologic studies relative to the probit and other choices of transformation can only be supported on grounds of mathematical convenience. Preference for the logistic model as opposed to the probit model is mainly based on the existence of simpler computational methods for parameter estimation (Grizzle, 1971; Cox and Snell, 1989). Also, the interpretation of logistic regression coefficients is easy and the odds ratio as a measure of association is usually the parameter of interest in a logistic regression. This easy interpretability of the coefficients is the fundamental reason why logistic regression has proven such a powerful analytic tool for epidemiologic research.

1.1.1 Methods of parameter estimation

There are two basic approaches to estimating the logistic parameters: maximum likelihood estimation and discriminant function estimation.

The maximum likelihood estimates (MLE's) are based on large-sample theory and are asymptotically unbiased. The MLE approach requires a function, called the likelihood function. This function expresses the probability of the observed data as a function of the unknown parameters. The MLE's of these parameter vector $\boldsymbol{\beta}$ in (1.1.1) are chosen to be those values which maximize this function. Thus, the resulting estimators are those which agree most closely with the observed data. The likelihood function of logistic regression is constructed as follows:

Suppose that we have a sample of N independent observations (Y_i, X_i), i=1,2,...,N. The contribution to the likelihood function for an observation (Y_i, X_i) is

$$\pi(X_i)^{y_i}[1-\pi(X_i)]^{1-y_i} . \qquad (1.1.1.1)$$

Since the observations are assumed to be independent, the likelihood function is obtained as the product of the terms given in (1.1.1.1). That is

$$l(\boldsymbol{\beta}) = \prod_{i=1}^{n} \pi(X_i)^{y_i}[1-\pi(X_i)]^{1-y_i} . \qquad (1.1.1.2)$$

The log-likelihood function is then defined as

$$L(\boldsymbol{\beta}) = \log[l(\boldsymbol{\beta})] \qquad (1.1.1.3)$$

$$= \sum_{i=1}^{N} \{ y_i \log[\pi(X_i)] + (1-y_i)\log[1-\pi(X_i)] \} .$$

To find the value of $\boldsymbol{\beta}$ that maximizes $L(\boldsymbol{\beta})$ we differentiate $L(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ and set the resulting expressions equal to zero. These equations are called the likelihood equations and expressed as follows:

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_0} = \sum_{i=1}^{N} [y_i - \pi(X_i)] = 0 \qquad (1.1.1.4)$$

and

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_j} = \sum_{i=1}^{N} X_{ij}[y_i - \pi(X_i)] = 0 \ \text{ for } j=1,2,\ldots,p . \qquad (1.1.1.5)$$

The likelihood equations in (1.1.1.4) and (1.1.1.5) are nonlinear in β's; thus the solution of them requires special methods. Most computer packages such as GLIM, BMDP, SAS, and SPSS use a method of iteratively reweighted least squares to obtain this solution.

We now discuss how the estimates of the standard errors of the estimated parameters are obtained. The method of estimating the variances and covariances of the estimated coefficients follows from the theory of maximum likelihood estimation (Rao, 1973). This theory states that the estimators are obtained from the matrix of second partial derivatives of the log likelihood function. These partial derivatives have the following general form

$$\frac{\partial^2 L(\beta)}{\partial \beta_j^2} = -\sum_{i=1}^{N} X_{ij}^2 \pi_i (1-\pi_i) \qquad (1.1.1.6)$$

and

$$\frac{\partial^2 L(\beta)}{\partial \beta_j \partial \beta_u} = -\sum_{i=1}^{N} X_{ij} X_{iu} \pi_i (1-\pi_i) \qquad (1.1.1.7)$$

for $j,u=0,1,2,\ldots,p$ where π_i denotes $\pi(X_i)$.

Let the (p+1) by (p+1) matrix containing the negative of the terms given in (1.1.1.6) and (1.1.1.7) be denoted as $I(\beta)$. This matrix is called Fisher's information matrix. The variances and covariances of the estimated coefficients are obtained from the inverse of this matrix which we will denote as $H(\beta) = I^{-1}(\beta)$.

The maximum likelihood estimator, $\hat{\beta}$, of the parameter vector β in the logistic regression model may be expressed in the iterative scheme as

$$\beta^{t+1} = \beta^t + (X'HX)^{-1}X's \text{, for } t=0,1,2,\ldots,^* \qquad (1.1.1.8)$$

where X is the design matrix, H is the diagonal matrix with general element $h_i = \hat{\pi}_i(1-\hat{\pi}_i)$, $\hat{\pi}_i$ is the estimated logistic probability at x_i, s is $(y - \hat{\pi})$, and both H and s are evaluated at β^t (see e.g. Pregibon, 1981; McCullagh and Nelder, 1983).

A most useful way to view the iterative process is by the method of iteratively reweighted least squares. This is obtained by employing the pseudo-observation vector $z^t = X\beta^t + H^{-1}s$, upon which the above equation (1.1.1.8) becomes

$$\beta^{t+1} = (X'HX)^{-1}X'Hz^t. \tag{1.1.1.9}$$

At convergence (t = *), we have $z = X\hat{\beta} + H^{-1}s$. Thus we may write the MLE of β as

$$\hat{\beta} = (X'HX)^{-1}X'Hz \tag{1.1.1.10}$$

The discriminant function estimators of the logistic regression coefficients are based on the normality assumption: $X \mid Y{=}j \sim N_p(\mu_j, \Sigma)$ for j = 0,1. Under this assumption the discriminant function estimates of β can be obtained as

$$\beta_0 = -\frac{1}{2}(\overline{X}_1 - \overline{X}_0)'S^{-1}(\overline{X}_1 + \overline{X}_0) + \log\left[\frac{n_0}{n_1}\right] \tag{1.1.1.11}$$

and

$$\beta = (\overline{X}_1 - \overline{X}_0)'S^{-1} \tag{1.1.1.12}$$

where n_0 and n_1 are the number of observations with Y equal to 0 and 1, respectively, and $\overline{X}_j$ and S are the maximum likelihood estimates of μ_j and Σ, respectively (see e.g. Hosmer and Lemeshow, 1989, p35).

Halperin et al. (1971) and Seigel and Greenhouse (1973) discussed the consequences of using the linear discriminant approach when the normality assumption is violated, and found that the discriminant estimates would remain biased in large samples. In terms of bias, maximum likelihood approach is then preferable to the linear discriminant function approach unless the normality assumption is satisfied. On the other hand, when the normality assumption is appropriate, Efron (1975) has shown that the discriminant estimates will be more efficient than the maximum likelihood estimates.

The assumption of multivariate normality will rarely if ever be satisfied because of the frequent occurrence of dichotomous independent variables in many situations. Press and Wilson (1978) recommended the logistic regression model with the maximum likelihood estimates in the case where the predictor variables consist of a mixture of continuous and discrete variables. Hosmer et al. (1983) have compared the two methods when the model contains a mixture of continuous and discrete variables, with the general conclusion that the discriminant function estimators are sensitive to the assumption of normality. The discriminant function estimates of the parameters for nonnormally distributed independent variables, especially dichotomous variables, are biased away from zero when the true value of parameter is nonzero. The practical implication of this is that for dichotomous independent variables the discriminant function estimators overestimate the magnitude of the association.

For these reasons, although estimating the logistic parameters based on the discriminant model is much easier than the maximum likelihood approach, maximum likelihood is preferable to discriminant analysis.

In our study, the maximum likelihood estimator, $\hat{\boldsymbol{\beta}}$, was obtained by the method of iteratively reweighted least squares for the parameter vector $\boldsymbol{\beta}$ in the logistic model (1.1.1.1).

1.1.2 Subset selection procedures

One of the main goals in logistic regression is to find the best fitting and most parsimonious, yet scientifically relevant model to describe the relationship between a response variable and a set of predictor variables; this is called model-building.

The most common strategy of model-building is a subset selection procedure which chooses a subset of predictor variables according to specific selection and stopping criteria. There are basically two approaches in selecting subset variables, namely 'best subsets' approach and 'stepwise' approach.

With best subsets approach a number of models containing one, two, three, and so on, variables are examined which are considered the 'best' according to some specified criteria (Furnival and Wilson, 1974; Lawless and Singhal, 1978).

The stepwise approach is a method in which variables are selected either for inclusion or exclusion from the model in a sequential fashion (see e.g. Draper and Smith, 1981). There are many variations, but the three main versions of the stepwise procedures are:

1) Forward selection (FS)

2) Backward elimination (BE)

3) Efroymson's procedure (Combination of FS and BE) (Efroymson, 1960)

These procedures have been widely used in epidemiologic studies in building a subset model from a large number of variables.

In any case, we should bear in mind that procedure for selection or deletion of variables from a model in any statistical package is solely based on a 'statistical' importance of variables.

1.2 Review of the literature

The literature will be separately reviewed in linear regression, discriminant analysis, and logistic regression. Some additional references will be reviewed in the context of the thesis.

1.2.1 Linear regression

The problem of selecting variables in linear regression has received considerable attention in the statistical literature. Among the more common procedures are the forward selection method, the backward elimination method, and Efroymson's (1960) stepwise regression. These procedures are discussed in detail in Draper and Smith (1981).

Garside (1965) and Schatzoff, Feinberg, and Tsao (1968) propose efficient methods of enumerating all possible regression equations. Hocking and Leslie (1967) and LaMontte and Hocking (1970) give a branch-and-bound procedure for determining the subset of each size with minimum residual sum of squares(RSS) without evaluating all possible regressions. Furnival and Wilson's algorithm (1974) appears to be the fastest program, which combines the best features of the branch-and-bound concepts and clever computing into a highly efficient algorithm.

The main property for so called 'subset modeling' is probably the one that subset model has generally smaller variance and larger bias than full model. (see e.g. Walls and Weeks, 1969; Rao, 1971; Narula and Ramberg, 1972; Rosenberg and Levy, 1972; and Hocking, 1974 for more details). In other words, there is a trade-off between the variance and the bias.

A selection criterion determines the order of inclusion or exclusion of the predictor variables. The 'standard' selection criterion in stepwise regression, namely the F-statistic, is based on the residual sum of squares (RSS).

Allen (1971) proposed the mean square error of prediction (MSEP) as a selection criterion. He claimed that when prediction is the main objective, the MSEP is a more meaningful criterion than the commonly used criterion, the residual sum of squares (RSS). This criterion utilizes the values of the predictor variables associated with the future observation and the magnitude of the estimated variance. It means that the use of the mean square error (squared bias + variance) takes into account bias and variability simultaneously.

The determination of the final subset size of model is generally referred to as a stopping criterion. A number of stopping criteria have been proposed in linear regression. Some of the more common ones are:

1) Sequential F test based on a fixed α level (F_α)
2) Minimum value of residual mean square (RMS)
3) Maximum value of squared multiple correlation coefficient (R^2)
4) Maximum value of adjusted R^2
5) Minimum value of average prediction variance (Mallows, 1967; Rothman, 1968; and Hocking, 1972)
6) Minimum value of Mallows' C_p (Mallows, 1973)
7) Minimum value of average prediction mean squared error (Tukey, 1967)
8) Minimum value of standardized residual sum of squares (Schmidt, 1973)
9) Minimum value of prediction sum of squares (PRESS) (Allen, 1971a; Schmidt, 1973; and Stone, 1974).
10) Minimum value of Akaike information criterion (Akaike, 1973)

Many of these stopping criteria are simple functions of the residual sum of squares (RSS).

Forsythe et al. (1973) developed a stopping criterion in the forward stepwise regression. At each step in the regression, the squared partial correlation is computed and compared with a random sample of squared 'permuted' partial correlations from N possible permutations of the Y vector. Stopping occurs whenever 5%, say, of the squared 'permuted' partial correlations exceed the computed squared partial correlation.

Kennedy and Bancroft (1971) sought to determine the best α level of significance to be used for repeated sequential F tests made in model building. They recommended, based upon the results of the simulation study, $\alpha = 0.25$ for the forward selection and $\alpha = 0.10$ for the backward elimination procedures.

Bendel and Afifi (1977) compared the unconditional mean square error of prediction (UMSE) with eight different stopping criteria including the 'standard' criterion in forward stepwise linear regression. The selection criterion used, at a given step, was to select the candidate variable that maximizes the squared partial correlation coefficient with the dependent variable, given the independent variables selected at previous steps. The eight stopping criteria were:

1) Minimum value of sample UMSE
2) Sequential Z test of equality of UMSE
3) Minimum value of conditional MSE
4) Minimum value of C_p statistic
5) Sequential F test based on a fixed α level
6) Lack-of-fit F test (Morgan and Tatar, 1973)
7) Combination of the sequential F test and the lack-of-fit F test (Summerfield and Lubin, 1951)

8) Maximum value of unbiased R^2 (Olkin and Pratt, 1958).

All of these stopping criteria are a simple functions of the partial correlation coefficient which requires only the correlation matrix for computation. They did not consider other stopping criteria such as PRESS and Akaike information criterion which require the actual values of the dependent and predictor variables for their computation. They showed that the Mallow's C_p statistic and the minimum value of sample UMSE are preferred to other stopping rules for higher degrees of freedom (40 or more), but the sequential Z test of equality of UMSE with $0.25 \le \alpha \le 0.35$ is preferred for smaller degrees of freedom (20 or less). They also found that the best overall test is the sequential F test with $\alpha = 0.15$. These results on the significant level are consistent with those reported by Kennedy and Bancroft (1971).

The question of how these many selection and stopping criteria suggested in the literature should be used in view of the intended use of the final subset model remains largely to be answered. The availability of good computer algorithms for computing subset regressions and the use of high speed computers would allow this area to be investigated.

1.2.2 Discriminant analysis

Discriminant analysis is a statistical technique for classifying individuals or objects into mutually exclusive and exhaustive groups on the basis of a set of independent variables. It involves the derivation of linear combinations of the independent variables, called the linear discriminant function (LDF), that discriminates between a priori defined groups in such a way that the misclassification error rates are minimized. This is accomplished by maximizing the between-group variance relative to the within-group variance. The LDF developed by Fisher (1936) is 'best' in that the misclassification error rates obtained with Fisher's LDF are smaller than those obtained with any other linear combination. However, the optimality of Fisher's LDF is conditional upon certain assumptions being met. In particular, the p independent variables must be a multivariate normal distribution with a common variance-covariance matrix in each of the two groups.

Fisher's LDF has become very popular in the field of multivariate analysis, and has consequently attracted a large amount of methodological research. The main areas of research have included the distribution of classification statistic, estimation of probabilities of misclassification, and the performance under non-optimal conditions.

The selection of the most useful variables in discriminant analysis is an important and difficult problem in practice. There are a number of various approaches which have been used for selecting a subset of variables from a larger set of variables in discriminant analysis. These approaches range from the univariate methods, which largely ignore the correlations among variables (Kendall and Stuart, 1968, p329; Lachenbruch, 1975, p73), to methods that consider all possible subsets (McCabe, 1975). Intermediate to these extremes are the popular stepwise procedures. Stepwise procedures such as forward selection, backward elimination, and combination of forward selection and backward elimination in discriminant analysis are available in program packages such as SPSS, BMDP, and SAS.

Discriminant analysis generally has two goals, namely description and prediction. The sense in which variable subsets are considered important should depend on the intended use of discriminant analysis. This point is often overlooked, although it has been stressed by writers such as Huberty (1975), Habbema and Hermans (1977), and Schaafsma and van Vark (1979).

In descriptive discriminant analysis, the primary aim may be to obtain a more parsimonious description, and hence a clearer understanding, of the nature of the differences among the populations in question. This may involve the construction and interpretation of discriminant functions. The importance of a variable subset should then be assessed in terms of the extent to which the populations are separated by that subset, a subset being considered adequate if it provides the same separation as the original set of variables.

Predictive discriminant analysis is concerned with the problem of assigning future observations to the appropriate populations. The importance of variables and the adequacy of subsets should then be assessed in terms of appropriate probabilities of misallocation.

Subset selection procedures of discriminant analysis in SPSS, BMDP, and SAS are associated with the descriptive rather than the predictive case. However, when prediction is the main goal, the 'standard' selection procedure based on Wilks' Λ (or the associated F-statistic for two groups) or related criteria do not necessarily yield a subset with maximum rate of correct classification (Habbema and Hermans, 1977).

McLachlan (1976) presented a different concept of selecting variables for prediction problems in stepwise discriminant analysis. His selection procedure is based on the conditional risk of misclassification and provides a confidence level that the conditional risk is not increased by deleting a given variable or subset of variables.

Costanza and Afifi (1979) used a simulation study to compare seven different stopping criteria in forward stepwise discriminant analysis. Four of their seven stopping criteria are based on $\Delta^2_{(q)}$, the q-variate Mahalanobis distance:

1) Sequential tail F test of $\Delta^2_{(q)} = \Delta^2_{(p)}$ based on a fixed α level

2) Sequential F test of $\Delta^2_{(q)} = \Delta^2_{(q+1)}$ based on a fixed α level

3) Combination of 1) and 2) based on a fixed α level

4) Maximum value of unbiased estimate of $\Delta^2_{(q)}$.

The other three stopping criteria are based on $P_u(q)$,the unconditional probability of correct classification:

5) Maximum value of estimate of $P_u(q)$ based on 'usual' estimate of $\Delta^2_{(q)}$

6) Maximum value of estimate of $P_u(q)$ based on 'almost unbiased' estimate of $\Delta^2_{(q)}$

7) Maximum value of estimate of $P_u(q)$ based on unbiased estimate of $\Delta^2_{(q)}$.

The forward selection method used in their study determines the candidate variable to be included in the discriminant function, based on maximizing the Mahalanobis distance given the variables selected at previous steps. They employed conditional and estimated unconditional probabilities of correct classification to compare the stopping rules that can be used in the two-group multivariate normal classification problem. Based on Monte Carlo simulation, they generally recommended using $0.10 \le \alpha \le 0.25$ as optimal α levels for the sequential F test which is the most commonly used stopping rule. The sequential F test uses a sequence of standard F tests to determine the significance of the additional distance contributed by each forward stepwise entry.

Costanza and Ashikaga (1986) explored choice of significance level for forward stepwise discrimination on small samples ($10 < N < 65$). $0.10 \le \alpha \le 0.25$ for F_α test were generally recommended.

1.2.3 Logistic regression

The literature associated with subset procedures in logistic regression is far less extensive than that of linear regression or discriminant analysis.

Lawless and Singhal (1978) proposed a statistic that approximates the likelihood ratio statistic for tests of submodels against the full model in nonnormal regression models such as exponential, Poisson, and logistic models. They compared the likelihood ratio statistic, Wald statistic, and their proposed statistic in 'best subsets' procedure with application to some epidemiologic data. They found that use of the three different statistics produced exactly the same best models in logistic model.

Peduzzi, Hardy, and Holford (1980) proposed a statistic that is similar to Rao's score statistic (Rao, 1973). Both can be used in stepwise selection procedure in nonlinear regression models including survival models and logistic model. They applied their statistic to data on survival from multiple myeloma (Krall et al., 1975) in an exponential regression model. They found that their statistic selected the same variables as the likelihood ratio statistic in forward stepwise procedure.

Lawless and Singhal (1987a, 1987b) developed an all-subsets regression program ISMOD for generalized linear models. They adopted the Furnival and Wilson's algorithm (1974) which is considered to be the most efficient computational algorithm.

Hosmer et al. (1989a) showed that best subsets procedure in logistic regression may be performed in a straightforward manner with any best subsets linear regression program that allows for case weights.

Over the last several years the logistic model derived from a subset procedure with application to a large number of predictor variables has been used with increasing frequency for the purpose of prediction in medical and epidemiologic studies. Although discriminant analyses have been used for the purpose of prediction (e.g., Afifi et al., 1971; Hans et al., 1986; Gardlund, 1986; Schaefer et al., 1991; and Ashutosh et al., 1992), these may not be appropriate when many of the predictor variables are binary or ordinal scaled. Ordinary least-squares multiple regression analyses have also been used (e.g., Snyder et al., 1981), which may not be appropriate when the response variable is binary. Since most of the practical data in medical and epidemiologic studies have the binary response variable and the mixture of discrete and continuous predictor variables, the logistic model is more appropriate than discriminant analysis and linear regression.

Here are several examples which used a subset procedure to develop the logistic model for the purpose of prediction in their studies.

Pozen et al. (1984) used a stepwise logistic regression to develop a predictive model for acute ischemic heart disease with patients admitted to coronary care units (CCUs). Lemeshow et al. (1985 and 1987) used a forward stepwise logistic regression to predict mortality of intensive care unit (ICU) patients. Tierney et al. (1985) used a backward elimination logistic regression to develop a predictive model for myocardial infarction in emergency room patients. Viviand et al. (1991) used a forward stepwise logistic regression to develop a predictive model for mortality of multidisciplinary intensive care unit (ICU) patients. Ferraris et al. (1992) used a stepwise logistic regression to develop a predictive model for an unimproved outcome in critical care patients. Horbar et al. (1993) used a backward elimination logistic regression to develop a predictive model for mortality of infants weighing 501 to 1500 grams at birth.

As can be seen in the above examples, computer-based stepwise procedures based on testing hypotheses have been the dominant approach in data analyses. All of the above examples used a procedure either in SAS or BMDP statistical software with $\alpha = 0.05$ as a cutoff level of significance. There are two logical drawbacks to this approach. First, the selection of significance level is necessarily subjective. Second, the test statistic used is the likelihood ratio or an asymptotically equivalent the score statistic with critical values that are appropriate only in the comparison of nested models.

It was not until Akaike (1973) introduced a model selection criterion, called Akaike information criterion (AIC), that issues such as the philosophical justifications and the asymptotic properties of the AIC and its variants were investigated (see e.g., Shibata, 1976; Stone, 1977 and 1979; Bhansali and Downham, 1977; Sawa, 1978; Schwarz, 1978; Kitagawa, 1979; Leamer, 1979; Atkinson, 1980; and Chow, 1981). Although these model selection criteria have been proposed in linear regression and time series contexts, they can also be applied to other areas. For example, AIC criterion and Schwarz criterion (Schwarz, 1978) are used as model selection criteria in the LOGISTIC procedure in SAS; lower values of the statistic indicate a more desirable model.

Where the logistic model is concerned, there has been no study to compare these nonparametric approach with the standard parametric approach that is, $\chi^2_{(\alpha)}$ based on a fixed α level in selecting a model for the purpose of prediction.

We note that a model selection criterion is different from a variable selection criterion; a model selection criterion may be considered as a stopping criterion which determines subset variables in a model, whereas a variable selection criterion determines the order of variables in a subset procedure.

CHAPTER 2

BACKGROUND AND OBJECTIVES OF STUDY

2.1 Motivation for using subset variables for the prediction

There are a variety of practical and economical reasons for reducing the number of predictor variables in the final model. Reasons for preferring a subset model to the full model include:

a. The subset model is more likely to be numerically stable than the full model. When some of the predictor variables are highly correlated (multicollinearity), they give rise to unstable parameter estimates (see e.g. Hoerl and Kennard, 1970).

b. The subset model is more easily interpretable than the full model. It describes a multivariate data set parsimoniously.

c. The subset model provides a reduced number of variables on which data are collected in a future study, thus lowering the cost of that study.

In addition to the above reasons, the subset model may be desirable in terms of predictive ability. The subset model may predict more accurately than the full model by eliminating uninformative variables.

In chapter 3 the true error rate of prediction (ERR) will be defined as

$$ERR = ARR + Bias \tag{2.1.1}$$

where ARR is the apparent error rate of prediction. The apparent error rate, originally named by Hills (1966), is estimated by the 'resubstitution' method. The apparent error rate tends to underestimate the true error rate because the data is used twice, both to fit the model and to evaluate its accuracy (Glick, 1972 and 1973). In the case of logistic regression the apparent error rate can be measured by the proportion of misclassified cases. Let $\hat{\eta}_i$ be the prediction rule defined as

$$\hat{\eta}_i = \begin{cases} 1 & \text{if } \hat{\pi}_i > C_0 \\ 0 & \text{if } \hat{\pi}_i \le C_0 \end{cases} \tag{2.1.2}$$

where $\hat{\pi}_i$ is the MLE of π_i and C_0 is the cutoff point. The choice $C_0 = 0.5$ is common. Then the apparent error rate is

$$ARR = \frac{\#\{y_i \neq \hat{\eta}_i\}}{N} \tag{2.1.3}$$

which is the proportion of cases in the original data set **Y** incorrectly predicted by $\hat{\eta}$.

The apparent error rate is a monotonically decreasing function of the number of variables in the model and is an optimistically biased estimate of the true error rate. Its bias becomes larger as the number of variables in the model increases. Bias is an important measure of how optimistic the apparent error rate is to the size of the model. A large value of bias suggests retreating to a more parsimonious model.

It may be possible that the rate of decrease for the apparent error rate may be small compared to that of increase for its bias after some number of variables, say q, have been included in the model. In other words, the true error rate depends on the 'trade-offs' between these two opposite rates and we may obtain smaller true error rate with a subset model than with the full model. This provides the motivation for using subset variables for the prediction in the logistic model.

2.2 Objectives of study

The principal aim in this study is to address the problem of building a model with a subset of predictor variables in forward stepwise logistic regression. Special attention will be given to the prediction problem. Prediction is one of the most important objectives of the logistic regression. In epidemiologic studies, for instance, given a new individual whose medical indicators are measured on a number of predictor variables, we want to predict whether or not the individual has cancer.

Since the logistic model requires less restriction on the distribution of predictor variables than discriminant analysis, the logistic model can be used in many practical situations in medical and epidemiologic studies. However, little attention has been given to issues surrounding the problem of selecting subset variables and assessing the performance of the logistic model with subset variables.

For prediction purposes, the valid estimate of an individual regression coefficient in the logistic model is not of the primary interest. The selection of predictor variables which yield a good prediction is of the primary interest. It is, therefore, of vital importance to understand the nature of how the prediction is changed when a different subset of predictor variables is selected.

The processes of building a subset model can be separated into two steps:

Step 1. A selection criterion which defines the order in which the predictor variables enter the logistic model.

Step 2. A stopping criterion which determines how many predictor variables are included in the final logistic regression model.

In this dissertation, we explore five topics which are related to the process and prediction of forward stepwise logistic regression:

1. Investigation of the best α level of significance for the $\chi^2_{(\alpha)}$ stopping criterion.
2. Comparison of the order of variables selected by different selection criteria.
3. Investigation of the effects of different structures of predictor variables on the apparent error rate of prediction, its bias, and the estimated error rate of prediction.
4. Comparison of the sizes of subset models determined by different stopping criteria.
5. Comparison of the performances of the selection and stopping criteria in terms of the estimated true error rate of prediction.

Monte Carlo simulations with both multivariate normal and multivariate binary distributions for predictor variables will be used to investigate them. In this dissertation we shall learn more about modeling logistic regression for the use of prediction.

CHAPTER 3

PERFORMANCE, SELECTION, AND STOPPING CRITERIA

3.1 Performance criterion

3.1.1 Bias of the apparent error rate of prediction

A common goal in medical studies is prediction. Suppose we observe n patients, $(y_1,x_1), \ldots, (y_n,x_n)$, where y_i is a binary response variable and x_i is a vector of predictors. For example, x_i might describe a medical patient's age, weight, sex, race, previous disease history, and so on, and y_i might indicate whether the patient survived a certain operation. These n patients are called the sample set. On the basis of the sample set, a prediction rule $\eta(y,\mathbf{x})$ is constructed. The intention is to use $\eta(y_0,\mathbf{x}_0)$ to predict a future unobserved response y_0 on the basis of its predictor vector $\mathbf{x}_0$.

Specifically, we will consider a prediction rule based on forward stepwise logistic regression. The true error rate of prediction (ERR) is the probability of predicting a future observation incorrectly. Our goal is to estimate ERR on the basis of the sample set $\mathbf{x}$. The most obvious estimate is the apparent error rate of prediction (ARR), which is the proportion of misclassified observations made by the prediction rule on its own sample set. Usually the apparent error rate tends to be smaller than the true error rate, because the same data have been used both to construct and to evaluate the prediction rule. Denoting ω as the bias of the apparent error rate, the true error rate is

$$\mathrm{ERR} = \mathrm{ARR} + \omega . \tag{3.1.1.1}$$

There are several nonparametric methods of estimating ω such as cross-validation, the jackknife, and the bootstrap (see e.g. Efron, 1982). Cross-validation is a traditional method of estimating ω. The jackknife method relates cross-validation to the bootstrap method and its estimate is a quadratic approximation to the bootstrap estimate. The bootstrap method gives the nonparametric maximum likelihood estimate (MLE) of ω. Gong (1986) compared these three nonparametric methods of estimating ω in logistic regression. Her conclusion, based on Monte Carlo simulations, is that whereas the jackknife and cross-validation do not offer significant improvement over the apparent error rate, the bootstrap shows substantial gain.

All three nonparametric estimates require substantial computing time and cost. Although computers are becoming faster and cheaper, these nonparametric methods of estimating ω are not feasible in our situation because estimates of ω must be computed at each step in forward stepwise logistic regression.

3.1.2 Efron's parametric estimate for bias of ARR

Efron (1986) derived the parametric MLE's for bias of the apparent error rate in the general exponential family linear models including the logistic regression. The bias estimate for the logistic regression is

$$\omega(\hat{\pi}) = \frac{2}{n}\sum_{i=1}^{n}\hat{\pi}_i(1-\hat{\pi}_i)Z\left[\frac{\hat{c}_i}{\sqrt{\hat{d}_i}}\right]\sqrt{\hat{d}_i} \qquad (3.1.2.1)$$

where

$$\begin{aligned}
Z(a) &= (2\pi)^{-1/2}\exp(-\frac{1}{2}a^2)\,,\\
\hat{c}_i &= \log\left[\frac{C_0}{1-C_0}\right]-x_i'\hat{\beta}\,,\\
\hat{d}_i &= x_i'\hat{\Sigma}^{-1}x_i\,,\\
\hat{\Sigma} &= \sum_{j=1}^{n}\hat{\pi}_j(1-\hat{\pi}_j)x_jx_j'\,.
\end{aligned}$$

The matrix $\hat{\Sigma}^{-1}$ is the usual estimate for the covariance matrix of $\hat{\beta}$, so $\hat{d}_i = \hat{Var}(x_i'\hat{\beta})$.

For a subset the basic form of the bias estimate $\omega(\hat{\pi})$ is not altered. Suppose that $\boldsymbol{\beta}$ is partitioned into $(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1)$, and likewise $\mathbf{x} = (\mathbf{x}_0, \mathbf{x}_1)$. The bias estimate in the subset is

$$\omega(\hat{\pi}) = \frac{2}{n}\sum_{i=1}^{n}\hat{\pi}_i(1-\hat{\pi}_i)Z\left[\frac{\hat{c}_i}{\sqrt{\hat{D}_i}}\right]\frac{\hat{d}_i^{\,0}}{\sqrt{\hat{D}_i}} \qquad (3.1.2.2)$$

Here

$$\hat{c}_i = \log\left[\frac{C_0}{1-C_0}\right] - \mathbf{x}_i'\hat{\boldsymbol{\beta}},$$

$$\hat{d}_i^{\,0} = \mathbf{x}_{0i}'\hat{\Sigma}^{0-1}\mathbf{x}_{0i},$$

$$\hat{\Sigma}^0 = \sum_{j=1}^{n}\hat{\pi}_j^{\,0}(1-\hat{\pi}_j^{\,0})\mathbf{x}_{0j}\mathbf{x}_{0j}',$$

$$\hat{D}_i = \mathbf{x}_{0i}'\hat{\Sigma}^{0-1}\hat{\Sigma}\hat{\Sigma}^{0-1}\mathbf{x}_{0i},$$

$$\hat{\Sigma} = \sum_{j=1}^{n}\hat{\pi}_i(1-\hat{\pi}_i)\mathbf{x}_{0j}\mathbf{x}_{0j}'.$$

In both situations the estimated true error rate of prediction $\hat{ERR}$ is given by

$$\hat{ERR} = ARR + \omega(\hat{\pi}). \qquad (3.1.2.3)$$

Efron (1986) compared $\omega(\hat{\pi})$ with the cross-validation estimate and showed that $\omega(\hat{\pi})$ performs better than the cross-validation estimate in terms of the mean squared error (MSE) of an estimate for ERR. In addition to the better performance of ω, we can obtain the estimates of ω at each step in forward stepwise logistic regression with simple calculations relative to those required for the nonparametric estimators. Hence $\hat{ERR}$ is used as the performance criterion in this study.

3.2 Selection criteria

In this section we discuss the statistics which will be used for selecting predictor variables for the model. Suppose k-1 variables have been previously selected and a k-th variable is considered for inclusion into the model. Then the components of β can be partioned as:

$$\beta = \begin{bmatrix} \beta_{k-1} \\ \beta_k \end{bmatrix} \tag{3.2.1}$$

and the hypotheses of interest are

$$H_0 : \beta_k = 0 \tag{3.2.2}$$

$$H_1 : \beta_k \neq 0$$

Note that β_{k-1} is a vector of k-1 parameters and β_k is a scalar of the k-th parameter. Since β can be estimated under both the null and alternative hypotheses, we need additional notation: let $\hat{\beta}^0$ denote the maximum likelihood estimation under the null hypothesis (restricted MLE), and $\hat{\beta}^1$ denote the maximum likelihood estimation under the alternative hypothesis (unrestricted MLE). That is,

$$\hat{\beta}^0 = \begin{bmatrix} \hat{\beta}_{k-1} \\ 0 \end{bmatrix} \tag{3.2.3}$$

$$\hat{\beta}^1 = \begin{bmatrix} \hat{\beta}_{k-1} \\ \hat{\beta}_k \end{bmatrix} \tag{3.2.4}$$

There are several large-sample statistics used to test the null hypothesis that the parameter value of the k-th variable is zero.

We describe four asymptotically equivalent statistics in distribution under the null hypothesis which will be defined as selection criteria of predictor variables in sections 3.2.1 - 3.2.4: the likelihood ratio statistic (LR), the Lawless and Singhal statistic (LS), Wald's statistic (WD), and Rao's score statistic (SC). Peduzzi, Hardy, and Holford's statistic (PH) will be described in section 3.2.5. We propose a statistic (LK) similar in form to Rao's score statistic but requiring less computation in section 3.2.6. We develop a stepwise procedure which adapts the SWEEP operator in section 3.2.7. This procedure is based on a statistic (SW) which is similar to the residual sum of squares in the linear regression.

3.2.1 The likelihood ratio statistic (LR)

The likelihood ratio statistic for testing (3.2.2) is

$$LR = 2[L(\boldsymbol{\beta}^1) - L(\boldsymbol{\beta}^0)] \qquad (3.2.1.1)$$

$$= 2[L(\hat{\boldsymbol{\beta}}_{k-1}, \hat{\beta}_k) - L(\hat{\boldsymbol{\beta}}_{k-1}, 0)]$$

We need to compute MLE's for $\boldsymbol{\beta}$ under both the null and alternative hypotheses.

We compute LR statistics for each of q non-selected variables, $LR(x_j)$, $j=1,\ldots,q$ where $LR(x_j)$ is the LR statistic when variable x_j is added to the k-1 variables previously selected. We then select variable x such that

$$LR(x) = \mathrm{Max}\{LR(x_j)\}, \ 1 \le j \le q .$$

3.2.2 The Lawless and Singhal statistic (LS)

Lawless and Singhal (1978) proposed a statistic that approximates the likelihood ratio statistic. Let $H(\boldsymbol{\beta})$ be partitioned as

$$H = \begin{bmatrix} H_{(1,\ldots,k-1)\times(1,\ldots,k-1)} & H_{(1,\ldots,k-1)\times k} \\ H_{k\times(1,\ldots,k-1)} & H_{k\times k} \end{bmatrix} . \qquad (3.2.2.1)$$

They showed that $\boldsymbol{\beta}^0$ can be approximated by

$$\hat{\beta}_{k-1} - H_{(1,\ldots,k-1)\times k}(\hat{\beta}^1)H^{-1}_{k\times k}(\hat{\beta}^1)\hat{\beta}_k . \qquad (3.2.2.2)$$

That is,

$$LS = 2[L(\hat{\beta}^1) - L(\hat{\beta}^0)] \qquad (3.2.2.3)$$

$$= 2[L(\hat{\beta}_{k-1},\hat{\beta}_k) - L(\hat{\beta}_{k-1} - H_{(1,\ldots,k-1)\times k}(\hat{\beta}^1)H^{-1}_{k\times k}(\hat{\beta}^1)\hat{\beta}_k)]$$

We need compute MLE's for β only under the alternative hypothesis.

We compute statistics for each of q non-selected variables, $LS(x_j)$, $j=1,\ldots,q$ where $LS(x_j)$ is the value of the LS statistic when variable x_j is added to the k-1 variables previously selected. We then select variable x such that

$$LS(x) = \text{Max}\{LS(x_j)\}, \quad 1 \le j \le q .$$

3.2.3 Wald's statistic (WD)

Wald's statistic depends only on MLE's for β under the alternative hypothesis, but does require the inverse $H^{-1}_{k\times k}(\hat{\beta}^1)$ for each step. For the partitioned matrix H described in (3.2.2.1), Wald's statistic is defined as

$$WD = \hat{\beta}_k H^{-1}_{k\times k}(\hat{\beta}^1)\hat{\beta}_k \qquad (3.2.3.1)$$

We compute statistics for each of q non-selected variables, $WD(x_j)$, $j=1,\ldots,q$ where $WD(x_j)$ is the value of the WD statistic when variable x_j is added to the k-1 variables previously selected. We then select variable x such that

$$WD(x) = \text{Max}\{WD(x_j)\}, \quad 1 \le j \le q .$$

3.2.4 Score statistic (SC)

The score statistic, also known as the efficient score statistic is due to Rao (1947); it requires a knowledge of MLE's only under the null hypothesis.

Let $U(\beta)=\frac{\partial}{\partial\beta}L(\beta)$ denote the efficient score vector, and $I(\beta) = -E\left[\frac{\partial}{\partial\beta}U(\beta)\right]$

denote the Fisher's information matrix. The score statistic is

$$SC = U'(\beta^0)\Gamma^{-1}(\beta^0)U(\beta^0) \tag{3.2.4.1}$$

$$= U'(\hat{\beta}_{k-1},0)\Gamma^{-1}(\hat{\beta}_{k-1},0)U(\hat{\beta}_{k-1},0)$$

We compute statistics for each of q non-selected variables, $SC(\mathbf{x}_j)$, j=1,...,q where $SC(\mathbf{x}_j)$ is the value of the SC statistic when variable $\mathbf{x}_j$ is added to the k-1 variables previously selected. We then select variable **x** such that

$$SC(\mathbf{x}) = \text{Max}\{SC(\mathbf{x}_j)\}, \ \ 1 \le j \le q \ .$$

3.2.5 Peduzzi, Hardy, and Holford statistic (PH)

Peduzzi, Hardy, and Holford (1980) proposed a statistic which is a modification of the Rao's score statistic. The statistic, which we call the PH, requires MLE's for β only under the null hypothesis and is given by

$$PH = \{\mathbf{0}_{(1,\ldots,k-1)}, U_k(\hat{\beta}_{k-1},0)\}'\Gamma^{-1}(\hat{\beta}_{k-1},0)\{\mathbf{0}_{(1,\ldots,k-1)}, U_k(\hat{\beta}_{k-1},0)\}\ , \tag{3.2.5.1}$$

where $\mathbf{0}_{(1,\ldots,k-1)}$ is a k-1 dimensional vector of zeros, and $U_k(\hat{\beta}_{k-1},0)$ is a scalar that is the efficient score of the k-th variable.

We compute statistics for each of q non-selected variables, $PH(\mathbf{x}_j)$, j=1,...,q where $PH(\mathbf{x}_j)$ is the value of the PH statistic when variable $\mathbf{x}_j$ is added to the k-1 variables previously selected. We then select variable **x** such that

$$PH(\mathbf{x}) = \text{Max}\{PH(\mathbf{x}_j)\}, \ \ 1 \le j \le q \ .$$

3.2.6 Lee and Koval statistic (LK)

We propose a statistic similar in form to the Rao's score statistic, but one requiring less computation. The statistic, which we call the LK, is given by

$$LK = U(\hat{\beta}_{k-1},0)'\text{Diag}\{\Gamma^{-1}(\hat{\beta}_{k-1},0)\}U(\hat{\beta}_{k-1},0)\ , \tag{3.2.6.1}$$

where $\text{Diag}\{I^{-1}(\boldsymbol{\beta}_{k-1},0)\}$ is the diagonal matrix of the inverse of the Fisher's information matrix I defined in the section 3.2.4 and evaluated under the null hypothesis. This is an effort to approximate the score statistic but ignoring the covariance among the elements of $\boldsymbol{\beta}$.

We compute statistics for each of q non-selected variables, $LK(\mathbf{x}_j)$, $j=1,\dots,q$ where $LK(\mathbf{x}_j)$ is the value of the LK statistic when variable $\mathbf{x}_j$ is added to the k-1 variables previously selected. We then select variable $\mathbf{x}$ such that

$$LK(\mathbf{x}) = \text{Max}\{LK(\mathbf{x}_j)\},\ 1 \le j \le q\ .$$

3.2.7 Sweep operator statistic (SW)

The SWEEP operator is perhaps the most versatile of all statistical operators. It has been extensively adapted for use in statistical computations such as all possible regressions, regression by leaps and bounds, stepwise regression, and etc. One of the first references to SWEEP operations may be found in Ralston (1960), but the term "SWEEP" operator was coined by Beaton (1964). We do not explain the actual processes of the SWEEP operator in detail; an excellent tutorial on the SWEEP operator may be found in Goodnight (1979).

Basically the SWEEP operator is a technique for the inversion of matrices. For example, let

$$A = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} \qquad (3.2.7.1)$$

Sweeping A on the columns associated with A_{11} yields

$$B = \begin{bmatrix} A_{11}^{-1} & A_{11}^{-1}A_{12} & A_{11}^{-1}A_{13} \\ -A_{21}A_{11}^{-1} & A_{22} - A_{21}A_{11}^{-1}A_{12} & A_{23} - A_{21}A_{11}^{-1}A_{13} \\ -A_{31}A_{11}^{-1} & A_{32} - A_{31}A_{11}^{-1}A_{12} & A_{33} - A_{31}A_{11}^{-1}A_{13} \end{bmatrix} \qquad (3.2.7.2)$$

Now we describe our variable selection procedure which adapts the SWEEP operator. Let **X** be partitioned as

$$\mathbf{X} = [\mathbf{X}_1 \mid \mathbf{X}_2] \tag{3.2.7.3}$$

$$= [\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{k-1}, \mathbf{x}_k \mid \mathbf{x}_{k+1}, \ldots, \mathbf{x}_p]$$

where k-1 variables are previously selected into the model and $\mathbf{x}_k$ is considered for inclusion into the model. Let G denote $\hat{g}(\mathbf{X}) = \log \frac{\hat{\pi}(\mathbf{X})}{1-\hat{\pi}(\mathbf{X})}$, and let W denote $\hat{\pi}(\mathbf{X}) \times (1-\hat{\pi}(\mathbf{X}))$. Note that G and W are obtained from the full model with p variables. Then we form the augmented matrix

$$A = \begin{bmatrix} X_1'WX_1 & X_1'WX_2 & X_1'WG \\ X_2'WX_1 & X_2'WX_2 & X_2'WG \\ G'WX_1 & G'WX_2 & G'G \end{bmatrix} \tag{3.2.7.4}$$

We sweep A on the columns of $X_1'WX_1$, and it gives

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} & \mathbf{B}_{13} \\ \mathbf{B}_{21} & \mathbf{B}_{22} & \mathbf{B}_{23} \\ \mathbf{B}_{31} & \mathbf{B}_{32} & \mathbf{B}_{33} \end{bmatrix} \tag{3.2.7.5}$$

where

$$\mathbf{B}_{11} = (\mathbf{X}_1'\mathbf{W}\mathbf{X}_1)^{-1}$$

$$\mathbf{B}_{21} = -\mathbf{X}_2'\mathbf{W}\mathbf{X}_1(\mathbf{X}_1'\mathbf{W}\mathbf{X}_1)^{-1}$$

$$\mathbf{B}_{31} = -\mathbf{G}'\mathbf{W}\mathbf{X}_1(\mathbf{X}_1'\mathbf{W}\mathbf{X}_1)^{-1}$$

$$\mathbf{B}_{12} = (\mathbf{X}_1'\mathbf{W}\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{W}\mathbf{X}_2$$

$$\mathbf{B}_{22} = \mathbf{X}_2'\mathbf{W}\mathbf{X}_2 - \mathbf{X}_2'\mathbf{W}\mathbf{X}_1(\mathbf{X}_1'\mathbf{W}\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{W}\mathbf{X}_2$$

$$B_{32} = G'WX_2 - G'WX_1(X_1'WX_1)^{-1}X_1'WX_2$$

$$B_{13} = (X_1'WX_1)^{-1}X_1'WG$$

$$B_{23} = X_2'WG - X_2'WX_1(X_1'WX_1)^{-1}X_1'WG$$

$$B_{33} = G'G - G'WX_1(X_1'WX_1)^{-1}X_1'WG$$

Let B_{33} be called SW, that is,

$$SW = G'G - G'WX_1(X_1'WX_1)^{-1}X_1'WG \qquad (3.2.7.6)$$

We compute statistics for each of q non-selected variables, $SW(\mathbf{x}_j)$, $j=1,\ldots,q$ where $SW(\mathbf{x}_j)$ is the value of the SW statistic when variable $\mathbf{x}_j$ is added to the k-1 variables previously selected. We then select variable $\mathbf{x}$ such that

$$SW(\mathbf{x}) = \mathrm{Min}\{SW(\mathbf{x}_j)\}, \quad 1 \le j \le q .$$

Note that in the linear regression situation SW is a kind of weighted residual sum of squares (RSS) for the model

$$G = (\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_k) \qquad (3.2.7.7)$$

The amount of time it takes to compute $(X'X)^{-1}$, consequently RSS, using the SWEEP operator is only a fraction of the time it takes to form the sum of squares and cross-product matrix. This can be verified by counting the number of multiplications and additions that are performed. Thus, the approximate computing time ratio of doing an inversion to building the $X'X$ is approximately $\frac{2p}{N}$, where p is the total number of independent variables and N is the number of observations. Because N is usually much larger than p, inversion represents only a fraction of the cost of regression analysis (Goodnight, 1979).

Our procedure requires only a small additional computation to fit the full model with p variables in order to obtain G and W. The remaining computations needed to perform the stepwise procedure in the logistic situation are the same as in the linear regression situation. Given that the 'conventional' selection procedures defined in sections 3.2.1 - 3.2.6 require $\boldsymbol{\beta}^0$ or/and $\boldsymbol{\beta}^1$ which must be obtained by iterative calculations, our procedure based on the SW statistic is much faster than the 'conventional' selection procedures.

3.3 Stopping criteria

In this section five stopping criteria are defined. Mnemonics for each of the stopping criteria are defined as follows:

1) $\chi^2_{(\alpha)}$: χ^2 test based on a fixed α level

2) E_m : Minimum value of $\hat{ERR}$

3) C_{pm} : Minimum value of C_p

4) AIC_m : Minimum value of Akaike information criterion

5) SCH_m : Minimum value of Schwarz's criterion

As indicated by the subscript, the stopping criterion $\chi^2_{(\alpha)}$ depends on a choice of the significant level α for its application. We use the best α whose computation is described in section 5.1. The stopping criterion $\chi^2_{(\alpha)}$ is the one used by most computer programs; the selection process is stopped at any step whenever the corresponding p-value exceeds the pre-specified α level.

The application of the other four stopping criteria E_m, C_{pm}, AIC_m, and SCH_m is slightly more general than $\chi^2_{(\alpha)}$. In their application, variable selection is continued until all variables are in the model, then the best model is determined by looking at values of the criterion for each subset selected.

3.3.1 χ^2 test based on a fixed α level, $\chi^2_{(\alpha)}$

In sections 3.2.1 - 3.2.7, seven selection criteria were defined. Four of the seven selection criteria in sections 3.2.1 - 3.2.4 are asymptotically equivalent statistics, namely the likelihood ratio statistic LR, the Lawless and Singhal statistic(LS), the Wald's statistic(WD), and the Rao's score statistic(SC), are asymptotically equivalent statistics. Under the null hypothesis, $H_0 : \beta_k = 0$, their asymptotic distribution is a chi-square on 1 degree of freedom.

The distribution of the PH statistic described in section 3.2.5 is not known. Peduzzi, Hardy, and Holford (1980) assumed that because of its similarity to the Rao's score statistic it had follow a chi-square distribution on 1 degree of freedom.

The distribution of the LK statistic described in section 3.2.6 is unknown; in this study because of its similarity to the Rao's score statistic it assumed to follow a chi-square distribution on 1 degree of freedom.

The SW statistic described in section 3.2.7 is a very different one. Its distribution is unknown. Therefore, the SW statistic is solely used as a tool for selecting a variable, and the corresponding likelihood ratio statistic Λ is used to test the null hypothesis.

The general application of the χ^2 test to all seven selection criteria is as follows: if χ_{max} denotes the value of a selection criterion associated with the predictor variable selected and the associated p-value is

$$p = \Pr[\chi_1^2 > \chi_{max}] \ , \tag{3.3.1.1}$$

then model building is stopped whenever p exceeds the pre-specified α level and the model is that with the previously selected k-1 variables.

3.3.2 Minimum value of $\hat{ERR}$, E_m

The expression for the estimated true error rate of prediction, $\hat{ERR}$, was given in (3.1.2.3).

We compute $\hat{ERR}(q)$ for q = 0, 1,...,p where $\hat{ERR}(q)$ is the value of the $\hat{ERR}$ with q variables. The E_m stopping criterion chooses a model with $\bar{q}$ variables such that

$$\hat{ERR}(\bar{q}) = \min_{0 \le q \le p} \{\hat{ERR}(q)\} \ . \tag{3.3.2.1}$$

3.3.3 Minimum value of C_p, C_{pm}

In the linear regression case Mallows' measure of predictive squared error, known as the C_p statistic, for a particular subset of q variables is given by

$$C_p(q) = \frac{SSE(q)}{SSE(p)/(n-p-1)} + 2(q+1) - n \qquad (3.3.3.1)$$

where SSE(q) is the residual sum of squares for the fitted linear regression model containing the subset of q variables (Mallows, 1973).

A C_p statistic for logistic regression was given by Hosmer (1989a). The expression for a particular subset of q variables is

$$C_p(q) = \frac{X^2 + \Lambda}{X^2/(n-p-1)} + 2(q+1) - n \qquad (3.3.3.2)$$

where X^2 is the Pearson chi-square goodness-of-fit statistic for the model with p variables that is,

$$X^2 = \sum_{i=1}^{n} \{(y_i - \hat{\pi}_i)^2/[\hat{\pi}_i(1-\hat{\pi}_i)]\}\ , \qquad (3.3.3.3)$$

and Λ is the Wald's statistic for the hypothesis that the coefficients for the $(p-q)$ variables not in the model are equal to zero.

Under the null hypothesis that all $(p-q)$ coefficients in the vector $\boldsymbol{\beta}_2$ are equal to zero, the appropriate expected values of X^2 and Λ are $(n-p-1)$ and $(p-q)$ respectively. Substitution of these two quantities in (3.3.3.2) yields $C_p(q) = q+1$. If the subset of variables under consideration has excluded important variables the Λ will follow a noncentral chi-square distribution and we would expect $C_p(q)$ to be larger than $(q+1)$. Thus, good subsets of variables will be those with small values of $C_p(q)$.

We compute $C_p(q)$ for $q = 0, 1, \ldots, p$ where $C_p(q)$ is the value of the C_p with q variables. The C_{pm} stopping criterion chooses $\tilde{q}$ variables such that

$$C_p(\tilde{q}) = \underset{0 \le q \le p}{\mathrm{Min}} \{C_p(q)\} \ . \qquad (3.3.3.4)$$

3.3.4 Minimum value of Akaike information criterion, AIC_m

In this section, we introduce the Akaike information criteria (AIC) as a basis of comparison and selection among several models.

AIC was introduced by Akaike (1973). He showed the important role of the Kullback-Leibler information quantity in statistics and also derived AIC as its estimator.

Without loss of generality, let the true distribution be given a discrete distribution $\mathbf{p} = \{p_1, p_2, \ldots, p_k\}$ where p_i is the probability that the event w_i occurs and satisfies $p_i > 0$ and $p_1 + \cdots + p_k = 1$. Our problem is when there are many models that approximate this true distribution, how do we evaluate the goodness of the approximation of these models to the true distribution? To answer this problem, we need an objective criterion that measures the distance between the true distribution and the model.

Suppose $\mathbf{p} = \{p_1, p_2, \ldots, p_k\}$ is the true distribution and $\mathbf{q} = \{q_1, q_2, \ldots, q_k\}$ a discrete distribution model, $\log(\mathbf{p}/\mathbf{q})$ is a random variable that takes the value $\log(p_i/q_i)$ when the event w_i occurs. The expectation of $\log(\mathbf{p}/\mathbf{q})$,

$$I(\mathbf{p};\mathbf{q}) = E[\log(\mathbf{p}/\mathbf{q})] \qquad (3.3.4.1)$$

$$= \sum_{i=1}^{k} p_i \log(p_i/q_i)$$

$$= \sum_{i=1}^{k} p_i \log(p_i) - \sum_{i=1}^{k} p_i \log(q_i)$$

is called the Kullback-Leibler quantity of information (K-L information quantity) of the true distribution **p** with respect to the model **q**. Here the first term on the right hand side is a constant that depends on the true distribution **p** only. Therefore, the larger the second term, the smaller the K-L information quantity becomes. This means that for the comparison of models by the K-L information quantity, we just need to estimate the value of $\sum_{i=1}^{k} p_i \log(q_i)$.

Suppose n independent observations $\{x_1, \ldots, x_n\}$ are obtained from the true distribution $\mathbf{p} = \{p_1, \ldots, p_k\}$. Each observation results in one of the events $w_1, \ldots, w_k$. If we define n_i as the number of occurrences of the events w_i, then we have $n_1 + \ldots + n_k = n$. Since p_i is the true probability, $\sum_{i=1}^{k} p_i \log(q_i)$ is the expectation of the random variable $\log(\mathbf{q})$ that takes the value $\log(q_i)$ when the event w_i occurs. This is called the expected log-likelihood. Given that the number of times that the random variable $\log(\mathbf{q})$ takes the value $\log(q_i)$ is n_i, the log-likelihood of the model **q** is

$$l(\mathbf{q}) = \frac{1}{n} \sum_{i=1}^{k} n_i \log(q_i) \qquad (3.3.4.2)$$

By the law of large numbers $l(\mathbf{q})$ converges to the expected log-likelihood as $n \rightarrow \infty$. It is considered that the larger the log-likelihood of the model **q** is, the closer the model **q** is to the true distribution **p**. Thus the comparison of the K-L information quantity is essentially equivalent to the comparison of the log-likelihood estimated from the data.

We use the mean expected log-likelihood as a measure for the goodness-of-fit of a model. This quantity is defined as the mean, with respect to the data **X**, of the expected log-likelihood of the maximum likelihood model. The larger the mean expected log-likelihood the better the fit of the model. It would seem that the mean expected log-likelihood can be estimated by the maximum log-likelihood, but the maximum log-likelihood can be shown to be a biased estimator of the mean expected log-likelihood. The maximum log-likelihood has a general tendency to over estimate the true value of the mean expected log-likelihood. This tendency is more prominent for models with a larger number of parameters. This means that if we choose the model with the largest maximum log-likelihood, a model with an unnecessarily large number of parameters is likely to be chosen. By a close examination of the relationship between the bias and the number of parameters of a model, Akaike found that (maximum log-likelihood of a model) - (number of parameters of the model) is an asymptotically unbiased estimator of the mean expected log-likelihood.

Akaike (1973) proposed the use of

$$\text{AIC} = -2\times(\text{maximum log-likelihood of the model}) + 2\times(\text{number of parameters of the model}) \qquad (3.3.4.3)$$

as the criterion for model selection. A model which minimizes the AIC is considered to be the most appropriate model. Definition (3.3.4.3) implies that when there are several models whose values of maximum likelihood are about the same level, we choose the one with the smallest number of parameters. In this sense AIC realizes the principle of parsimony.

We compute AIC(q) for q = 0, 1,...,p where AIC(q) is the value of the AIC with q variables. The AIC_m stopping criterion chooses a model with $\tilde{q}$ variables such that

$$\text{AIC}(\tilde{q}) = \min_{0 \le q \le p} \{\text{AIC}(q)\} . \qquad (3.3.4.4)$$

3.3.5 Minimum value of Schwarz's criterion, SCH_m

Schwarz (1978) proposed a different way of adjusting the log-likelihood statistic. If the log-likelihood maximized over q parameters is L_q, Schwartz's criterion is defined as

$$SCH(q) = -2L_q + q \times \log N \qquad (3.3.5.1)$$

where N is the number of observations in the data set.

We compute SCH(q) for q = 0, 1,...,p. The SCH_m stopping criterion chooses a model with $\tilde{q}$ variables such that

$$SCH(\tilde{q}) = \underset{0 \le q \le p}{\mathrm{Min}} \{SCH(q)\} \ . \qquad (3.3.5.2)$$

CHAPTER 4

MONTE CARLO EXPERIMENTAL DESIGN

Real data sets are limited in their usefulness in that they limit consideration of the effect of sample size, number of predictor variables, distribution of dependent and predictor variables, etc. The choice of data sets may also influence conclusions, hence the generalizability of the results may be limited. Simulation studies can be used to obtain results over a wide range of sampling situations.

In the general design of the simulation experiments there are four steps.

Step 1. Generation of predictor variables, $\mathbf{X}_i$, i=1,2,...,N.

$(\mathbf{X}_i \mid Y=0)$, i=1,2,...,n_0, and $(\mathbf{X}_i \mid Y=1)$, i=1,2,...,n_1 are generated from populations Π_0 and Π_1, respectively. Note that $N = n_0 + n_1$.

Step 2. Computation of $\hat{\boldsymbol{\beta}}$, the maximum likelihood estimates of $\boldsymbol{\beta}$.

$\hat{\beta}_i$, i = 0,1,2,...,p are obtained via iteratively reweighted least squares.

Step 3. Computation of estimated probabilities, $\Pr(Y=1 \mid \mathbf{X}_i) = \hat{\pi}_i(\mathbf{X})$.

$\hat{\pi}_i(\mathbf{X})$, i=1,2,...,N, are computed using $\mathbf{X}_i$ from Step 1 and $\hat{\boldsymbol{\beta}}$ from Step 2.

Step 4. Generation of predicted dependent variable, $\hat{Y}_i$.

$\hat{Y}_i$ is equal to 0 if $\hat{\pi}_i(\mathbf{X}) \le 1/2$ and $\hat{Y}_i$ is equal to 1 if $\hat{\pi}_i(\mathbf{X}) > 1/2$.

Repeat Steps 1-4 R=20 times.

We consider two multivariate distribution cases of predictor variables: 1) multivariate normal case and 2) multivariate binary case.

4.1 Multivariate normal case

In section 4.1.1 we describe how multivariate normal variables were generated; this follows the reparameterization method of Bendel and Afifi (1977) in which μ and Σ are reparameterized by several factors.

In section 4.1.2 we describe an experimental design in which a variety of values of the factors are specified to generate different sampling situations. Specifically, we use a second-order central response surface design (Cochran and Cox 1957, chapter 8A).

4.1.1 Generation of multivariate normal variables, X

We assume that $\mathbf{X} \sim N_p(0,\Sigma)$ in population Π_0 and $\mathbf{X} \sim N_p(\mu,\Sigma)$ in population Π_1. We reparameterize μ and Σ in terms of four factors P, V, Δ^2, and D.

The first factor is the number of predictor variables, P. Given a value of P, we then specify a value of the second factor, $V \in (0,1]$. It determines the eigenvalues λ_i of Σ by means of the expression

$$\lambda_i = aV^{i-1} + \delta, \text{ for } i=1,2,\ldots,P \qquad (4.1.1.1)$$

where

$$a = \begin{cases} 0.9P(1-V)/(1-V^P) & \text{if } 0<V<1 \\ 1-\delta & \text{if } V=1 \end{cases} \qquad (4.1.1.2)$$

A value of δ=0.1 is chosen as a lower bound on the smallest eigenvalue λ_p in order to avoid the numerical difficulties encountered with nearly singular matrices.

This model arises from the observed behavior of the eigenvalues in a principal component or factor analysis of a sample correlation matrix. When the predictor variables are highly intercorrelated, there are usually a few large eigenvalues which exceed the value one. On the other hand, if the variables are nearly independent, the eigenvalues do not extensively deviate from the value one. The eigenvalues therefore reflect the degree of interdependence among the predictor variables. The possible values of V represent a continuum such that the dependence among the variates increases as V decreases from 1; the variables are highly interdependent near V = 0 and mutually independent at V = 1 (in this case all eigenvalues are equal to one).

Since $\Sigma = L\Lambda L'$, where L is the matrix of eigenvectors of Σ and Λ is the diagonal matrix of eigenvalues λ_i, then once the λ_i have been specified a random orthogonal matrix L is generated and used to create Σ.

The third factor is the Mahalanobis distance between Π_0 and Π_1 defined by $\Delta^2 = \mu'\Sigma^{-1}\mu$; It describes the separation of the two populations.

The fourth factor D determines the elements μ_i of μ. It is an attempt to produce more realistic patterns than the canonical case which has zeros in all positions except the first.

Let

$$\mu_i^* = (bD^{i-1})^{1/2} \text{ for } i=1,2,\ldots,P \text{ and } 0<D\leq 1 \,. \qquad (4.1.1.3)$$

where

$$b = \begin{cases} \Delta^2(1-D)/(1-D^P) & \text{if } 0<D<1 \\ \Delta^2/P & \text{if } D=1 \end{cases} \,. \qquad (4.1.1.4)$$

Elements μ_i of μ are then obtained from $\mu = R\mu^*$ where $\Sigma = RR'$ is the Cholesky decomposition of Σ.

As D varies from 0 to 1, the rates of increase in Δ^2 decreases as the number of included variables increases from 1 to P.

To generate a P-variate observation **X** from $N_p \sim (\mu,\Sigma)$, P independent N(0,1) values Z's are first generated. The vector Z is then transformed to the required vector **X** by **X** = μ + RZ, with $\Sigma = RR'$ as above.

4.1.2 Second-order central response surface design

The levels of five factors P, V, Δ^2, D, and N must be specified. The simplest design is to specify each of 5 factors at 2 levels each; this is a 2^5 factorial design. With only two levels it is not possible to determine if the effect of a factor is linear.

Specifying a 3^5 factorial design leads to 243 combinations, although a one-third replicate would reduce the number to 81. A one-ninth replicate has only 27 combinations but, as explained in Davis (1956), "only two main effects are clear of two-factor interactions".

In any case, the fractional design approach was abandoned in favor of a response surface design, which has the advantage of specifying each factor at more than 2 levels while not requiring as many combinations as a full factorial design.

An introduction to response surfaces is given in Davis (1956) and various 'plans' are available in Cochran and Cox (1957, chapter 8A). In our study a second-order central response surface design was chosen because this design is typically used in practice (Cochran and Cox, 1957, p347). In this design a quadratic surface or second degree polynomial is fitted by the method of least squares. The fitted second degree polynomial may be expressed as:

$$\hat{y} = b_0 + \sum_{i=1}^{k} b_i x_i + \sum_{i=1}^{k}\sum_{j=i}^{k} b_{ij} x_i x_j \,, \qquad (4.1.2.1)$$

where k=number of factors. For five factors there are 1+5+5+10=21 coefficients to be estimated.

We define a sampling situation to be the particular **X** obtained by specifying levels for the factors P, V, Δ^2, D, and N. A sampling situation is often called a run in the experiment. The method used to specify the sampling situations is referred to as the experimental design.

The factors may be viewed as spanning a five dimensional factor space in which a 'point' represents the set of all sampling situations which can be characterized by the given levels of the five factors; thus specification of a number of sampling situations is equivalent to specifying a set of points in the factor space. The idea of the design is to specify the sampling situations in a symmetrical configuration about a reasonable 'centre point', such that the region defined provides adequate coverage of the possible factor levels.

To describe the experimental design, each point in the factor space is regarded as a quintuplet of factor levels, written symbolically as

$$(P, V, \Delta^2, D, N)\ . \qquad (4.1.2.2)$$

Each factor has five levels, which are taken to be equally spaced on a suitable scale and are coded as

$$-2, -1, 0, +1, +2\ . \qquad (4.1.2.3)$$

These levels are termed 'low star', 'low factorial', 'centre', 'high factorial', and 'high star', respectively.

The values of the 5 factors for this study are given in Table 4.1.2.1. The values of P from 5 through 25 represent a range from a small number of variables to a large number of variables; V from 0.2 through 1.0 represents a range of dependency from highly dependent to perfectly independent; Δ^2 from 1.0 through 3.0 represents a range of distances between populations from close to well-separated; D from 0.2 through 1.0 represents rates of increases of $\Delta^2_{(q)}$ to $\Delta^2_{(p)}$ from fast to slow; and N from 100 through 300 represents from a small sample to a moderate sample.

The design consists of 48 sampling situations or points of three types: a) 2^5=32 factorial points which are all possible combinations of the ±1 levels for each factor; b) 10 star points which have the -2 or +2 level; and c) 6 centre points which have the 0 levels (see Table 4.1.2.2). This is a second-order central response surface design.

The star and centre points are added to the basic full factorial points in order to give the design the characteristic of rotatability. If $\hat{y}_j$ is the estimated regression function in (4.1.2.1) for the j-th observation, then rotatability means that, for all points for which

$$\sum_{i=1}^{5} x_{ij}^2 = \text{constant}\,, \qquad (4.1.2.4)$$

the standard error of response,

$$\text{S.E.}(\hat{y}_j) = \sqrt{\text{var}(\hat{y}_j)} \qquad (4.1.2.5)$$

is also constant.

For five factors a value of 'high star' and 'low star' = ±2.378 would make the design rotatable if the basic factorial is a full factorial. Since the property of rotatability is desirable rather than critical to the analysis of data (Box and Behnken 1960), a full factorial with a convenient value of ±2 is chosen. The use of 6 centre points enables us to obtain a fairly stable picture of the results of the simulations in this 'intermediate' region. On the other hand, the 10 star points serve to extend the region covered by the factorial points.

Table 4.1.2.1

The values of the five factors in the response surface design

	Level (Code)				
Factor	low star (-2)	low factorial (-1)	centre (0)	high factorial (+1)	high star (+2)
P	5	10	15	20	25
V	0.2	0.4	0.6	0.8	1.0
Δ^2	1.0	1.5	2.0	2.5	3.0
D	0.2	0.4	0.6	0.8	1.0
N	100	150	200	250	300

Table 4.1.2.2

48 sampling situations in the second-order central response surface design

	Factor						Factor				
Run	P	V	Δ^2	D	N	Run	P	V	Δ^2	D	N
1	-1	-1	-1	-1	-1	17	-1	-1	-1	-1	1
2	1	-1	-1	-1	-1	18	1	-1	-1	-1	1
3	-1	1	-1	-1	-1	19	-1	1	-1	-1	1
4	1	1	-1	-1	-1	20	1	1	-1	-1	1
5	-1	-1	1	-1	-1	21	-1	-1	1	-1	1
6	1	-1	1	-1	-1	22	1	-1	1	-1	1
7	-1	1	1	-1	-1	23	-1	1	1	-1	1
8	1	1	1	-1	-1	24	1	1	1	-1	1
9	-1	-1	-1	1	-1	25	-1	-1	-1	1	1
10	1	-1	-1	1	-1	26	1	-1	-1	1	1
11	-1	1	-1	1	-1	27	-1	1	-1	1	1
12	1	1	-1	1	-1	28	1	1	-1	1	1
13	-1	-1	1	1	-1	29	-1	1	1	1	1
14	1	-1	1	1	-1	30	1	1	1	1	1
15	-1	1	1	1	-1	31	-1	1	1	1	1
16	1	1	1	1	-1	32	1	1	1	1	1
33	0	0	0	0	-2	38	0	0	2	0	0
34	0	0	0	0	2	39	0	-2	0	0	0
35	0	0	0	-2	0	40	0	2	0	0	0
36	0	0	0	2	0	41	-2	0	0	0	0
37	0	0	-2	0	0	42	2	0	0	0	0
43	0	0	0	0	0	46	0	0	0	0	0
44	0	0	0	0	0	47	0	0	0	0	0
45	0	0	0	0	0	48	0	0	0	0	0

4.2 Multivariate binary case

The generation of multivariate binary variables from the second-order Bahadur (1961) model is described in section 4.2.1. The full factorial design is specified in section 4.2.2.

4.2.1 Generation of multivariate binary variables

For data generated by discrete random variables $X_1, X_2, \ldots, X_p$, each assuming at most a finite number of distinct values $s_1, s_2, \ldots, s_p$ the sample space S consists of $S = \prod_{j=1}^{p} s_j$ states and may be assumed to be generated by the random vector $\mathbf{X}_i = (X_1, X_2, \ldots, X_p)$, $i=1,2,\ldots,N$, with a multinomial distribution. Assume also that the data $\mathbf{X}$ can arise from one of the two multinomial populations Π_1 and Π_0 with sample size n_1 and n_0, respectively.

The 'full' multinomial model has the problem of a rapid increase in number of states which is troublesome in practice because of small sample sizes. For example, five variables, each assuming only three distinct values, generates $3^5 = 243$ states. Obviously, a large number of observations relative to the number of variables is required if sufficient data in each state are to be available for the estimation of state probabilities. Because of the problem of none or few observations in some states researchers have been reluctant to apply 'full' multinomial procedures.

One of the most common variants of the full multinomial procedures is to assume that the data are dichotomous. In particular, suppose that X_j=0 or 1 for $j=1,\ldots,P$. The induced multinomial distributions have $S=2^P$ states, with 2^P-1 independent parameters in each population.

We denote f to represent a general multivariate dichotomous density, whereas f_1 and f_0 are the corresponding densities associated with populations Π_1 and Π_0, respectively. The Bahadur (1961) model uses a reparametrization of f(**X**), in terms of means and correlations. The number of new parameters utilized in the reparametrization is essentially equal to the number of parameters needed to characterize the full multinomial model, that is, 2^P-1. Hence, if the full model is used, the problem of a large number of parameters remains. However, in many instances, the value of a given state probability when expressed in the Bahadur form is mostly captured when only a few parameters are used.

Let $\mathbf{X} = (X_1, X_2, \cdots, X_p)$, where X_j is a Bernoulli random variable with $p_j = \Pr(X_j = 1)$ and $1-p_j = \Pr(X_j = 0)$, j=1,...,P. A particular value of **X**, denoted by $\mathbf{x} = (x_1, x_2, \ldots, x_p)$, is called a response pattern. The response pattern consists of a series of 0's and 1's, and the probability that response pattern **x** is observed will be denoted by f(**x**).

Now set

$$Z_j = \frac{X_j - p_j}{\sqrt{p_j(1-p_j)}} \tag{4.2.1.1}$$

and define

$$\rho_{jk} = E(Z_j Z_k) \tag{4.2.1.2}$$

$$\rho_{jkl} = E(Z_j Z_k Z_l)$$

$$\vdots$$

$$\rho_{jk\ldots p} = E(Z_j Z_k \ldots Z_p).$$

Bahadur has shown that for any response pattern $\mathbf{x} = (x_1, x_2, \ldots, x_p)$, $f(\mathbf{x})$ is given by

$$f(\mathbf{x}) = \Pr(\mathbf{X}=\mathbf{x}) \tag{4.2.1.3}$$

$$= \prod_{j=1}^{p} p_j^{x_j}(1-p_j)^{1-x_j} \times$$

$$\{1+\sum_{j<k}\rho_{jk}z_jz_k+\sum_{j<k<l}\rho_{jkl}z_jz_kz_l+\ldots+\rho_{12\ldots p}z_1z_2\cdots z_p\},$$

where z_j is the observed value of Z_j corresponding to x_j.

The appeal of this reparametrization lies in its ability to describe the multinomial probabilities $f(\mathbf{X})$ in terms of means, p_j, and correlations, ρ_{jk}. In many applications it makes sense to assume that higher-order correlations are zero, thus reducing the number of parameters required for estimation.

To distinguish parameters under Π_1 as opposed to those from Π_0 we write p_{ij} for $\Pr(X_j=1 \mid \Pi_i)$, i=0,1, and $\rho_i(jk)$ for the corresponding correlation terms.

The 'reduced' Bahadur model used in this study is a second-order approximation obtained by retaining the p_{ij} and $\rho_i(jk)$ and omitting all higher order terms. In this case $f_i(\mathbf{x})$ is given by

$$f_i(\mathbf{x}) = \Pr(\mathbf{X}=\mathbf{x} \mid \Pi_i) \tag{4.2.1.4}$$

$$= \prod_{j=1}^{p} p_{ij}^{x_{ij}}(1-p_{ij})^{1-x_{ij}}\{1+\sum_{j<k}\rho_i(jk)z_{ij}z_{ik}\},$$

for i=0,1, and j=1,...,P.

However, when the higher order correlations are taken to be zero, function (4.2.1.4) is not necessarily a probability distribution function; $f_i(\mathbf{x})$ may fail to be nonnegative for some response patterns (it is always true that $\sum_{\mathbf{x}\in\mathbf{X}} f_i(\mathbf{x}) = 1$). The problem with this function is that $\rho_1(jk)$ are not free to vary over $[-1,+1]$; they must satisfy certain linear inequalities defined by the marginal expectations.

By assuming that many of the higher correlations are zero, severe limitations must be placed on $\rho_1(jk)$ to ensure that $0 \le f_1(x) \le 1$. Letting $t = \sum_{j=1}^{P} x_j$, Bahadur (1961) has shown that the second-order Bahadur model (4.2.1.4) is a valid probability distribution if and only if

$$-\frac{2}{P(P-1)}\mathrm{Min}\left[\frac{p_j}{1-p_j},\frac{1-p_j}{p_j}\right] \le \rho \le \frac{2p_j(1-p_j)}{(P-1)p_j(1-p_j)+1/4-\gamma_0} \qquad (4.2.1.5)$$

where

$$\gamma_0 = \mathrm{Min}_t\{[t-(P-1)p_j-1/2]^2\} \le 1/4 \qquad (4.2.1.6)$$

Table 4.2.1.1 provides values of the lower and upper bounds in (4.2.1.5) for various choices of P and p_j; because of symmetry, only values for $p_j \le 0.5$ need to be tabulated.

A few comments may be made regarding the inequality (4.2.1.5) and table 4.2.1.1.

1. The upper bound is clearly less restrictive than the lower bound; this is desirable since in most (but not all) situations one would expect ρ to be positive.
2. Both bounds become closer to zero as P increases, so that, in practice, the largest P in a given data set is associated with the most restrictive and governing set of bounds.
3. As p_j becomes closer to 0.50, the width of bounds becomes larger, so that $p_j=0.50$ is associated with the least restriction on ρ.

In general, a sampling experiment is characterized by the values assigned to the parameters p_{0j}, p_{1j}, $\rho_0(jk)$, and $\rho_1(jk)$ for j,k=1,...,P. The general class of correlation structures used in the experiments is that the correlation between any two variables in Π_0 is identical to that in Π_1. This allows us to investigate situations in which the variance-covariance matrices in the two populations are identical. When $\rho_0(jk) = \rho_1(jk)$ for all j and k, then letting $p_{1j} = 1-p_{0j}$ for all j yields equal covariance structure.

Let ρ_{upp_i} denote the upper bound in (4.2.1.5). The numerical values used in the correlation structures are:

$$\rho_0(jk) = \rho_1(jk) = r \text{ if all } j \neq k \text{ for } r = \rho_{upp_i} \qquad (4.2.1.7)$$

Once the parameters have been specified according to the scheme the method for generating Monte Carlo samples can be outlined as follows:

Suppose that $x_1, x_2, \ldots, x_{k-1}$ have been generated and we are to generate the next x, x_k. The conditional probability $f(x_k \mid x_1, x_2, \ldots, x_{k-1})$ in the second-order Bahadur model (4.2.1.4) can be expressed as

$$f(x_k \mid x_1, \ldots, x_{k-1}) = \frac{f(x_1, \ldots, x_k)}{f(x_1, \ldots, x_{k-1})} \qquad (4.2.1.8)$$

$$= \frac{\prod_{j=1}^{k} p_i^{x_j}(1-p_i)^{1-x_j}\{1+\sum_{j<k}\rho(jk)z_j z_k\}}{\prod_{j=1}^{k-1} p_i^{x_j}(1-p_i)^{1-x_j}\{1+\sum_{j<(k-1)}\rho(jk)z_j z_k\}}$$

$$= \frac{p_i^{x_k}(1-p_i)^{1-x_k}\{1+\sum_{j<k}\rho(jk)z_j z_k\}}{\{1+\sum_{j<(k-1)}\rho(jk)z_j z_k\}}, \quad k=1,\ldots,p .$$

Generate u from a uniform random variable on [0,1]; then x_k is generated according to

$$x_k = \begin{cases} 0, & \text{if } u \le f(x_k = 0 \mid x_1, \ldots, x_{k-1}) \\ 1, & \text{otherwise} \end{cases} \quad (4.2.1.9)$$

4.2.2 The values of four factors in the full factorial design

We take P = 10, 15, and 20; N = 150, 200, and 250. These are the same levels as 'low factorial', 'centre', and 'high factorial' in the multivariate normal case (see table 4.1.2.1). We also take p_0 and p_1 such that $p_0 < p_1$; $p_0 = 0.2$ with $p_1 = 0.3, 0.4$, and 0.5; $p_0 = 0.4$ with $p_1 = 0.5, 0.6$, and 0.7; and $p_0 = 0.6$ with $p_1 = 0.7, 0.8$, and 0.9. These 9 pairs of (p_0, p_1) give rise to 3 levels of p_0 (0.2, 0.4, and 0.6) and 3 levels of $(p_1 - p_0)$ (0.1, 0.2, and 0.3) (see table 4.2.2.1). For convenience, let the symbols B and M denote the p_0 and $(p_1 - p_0)$, respectively. The experiment is then a $P \times N \times B \times M = 3^4$ full factorial design. The values of the four factors are summarized in table 4.2.2.2.

4.3 Number of replications

The number of replications of size R control the inherent sampling error of the Monte Carlo experiment. In our study we choose R=20.

Let the symbol $Y = \sin^{-1}(E\hat{R}R)^{1/2}$ denote the arc-sine of estimated true error rate of prediction. Y has variance $\frac{1}{4 \times N}$, where N is number of observations in one replicate. N is set by experimental design, say N = 200 which is the centre level in table 4.1.2.1. The required number of replications is then

$$R = 2 \times (Z_{\alpha/2} + Z_{\beta})^2 \times \frac{Var(Y)}{\delta^2} \quad (4.3.1)$$

where α is the level of significance test, β is the power of test, Var(Y) is the variance of the transformed observation Y, and δ is detectable difference.

$R = 20$, $\alpha = 0.05$, and $\beta = 0.20$ in (4.3.1) gives

$$20 = 2\times(1.96 + 0.84)^2 \times \frac{\frac{1}{4\times 200}}{\delta^2} \qquad (4.3.2)$$

Solving for δ in (4.3.2) gives $\delta = 0.0313$ on arc-sine scale. For example, say $\hat{ERR} = 0.20$ and then $\sin^{-1}(0.20)^{1/2} = 0.4636$. The difference of 0.0313 from 0.4636 is then 0.4323. $\sin^2(0.4323) = 0.176$ is the difference of 0.024 from 0.20. It means that $R = 20$ is enough to detect an 0.024 difference in $\hat{ERR}$.

Table 4.2.1.1

Permissible ranges of values for ρ based on (4.2.1.5)

for various choices of P and p_j

	p_j				
P	0.10	0.20	0.30	0.40	0.50
5	(-0.01, 0.30)	(-0.03, 0.40)	(-0.04, 0.42)	(-0.07, 0.40)	(-0.10, 0.50)
10	(-0.00, 0.20)	(-0.01, 0.20)	(-0.01, 0.20)	(-0.02, 0.20)	(-0.02, 0.20)
15	(-0.00, 0.12)	(-0.00, 0.13)	(-0.00, 0.14)	(-0.01, 0.13)	(-0.01, 0.14)
20	(-0.00, 0.10)	(-0.00, 0.10)	(-0.00, 0.10)	(-0.00, 0.10)	(-0.01, 0.10)
25	(-0.00, 0.08)	(-0.00, 0.08)	(-0.00, 0.08)	(-0.00, 0.08)	(-0.00, 0.08)

Table 4.2.2.1 9 pairs of (p_0, p_1) which give rise to 3 levels of B and 3 levels of M.			
	[a]B		
[b]M	0.2	0.4	0.6
0.1	(0.2, 0.3)	(0.4, 0.5)	(0.6, 0.7)
0.2	(0.2, 0.4)	(0.4, 0.6)	(0.6, 0.8)
0.3	(0.2, 0.5)	(0.4, 0.7)	(0.6, 0.9)

[a]B denotes p_0.

[b]M denotes $(p_1 - p_0)$.

Table 4.2.2.2 The values of the four factors in the full factorial design			
Factor	Values		
P	10	15	20
[a]B	0.2	0.4	0.6
[b]M	0.1	0.2	0.3
N	150	200	250

[a]B denotes p_0.

[b]M denotes $(p_1 - p_0)$.

CHAPTER 5

RESULTS OF THE SIMULATION EXPERIMENTS

This chapter presents the results of the sampling experiments described in chapter 4. There are five sections in this chapter; each contains both the results of the multivariate normal case and the multivariate binary case. In section 5.1 the best α level of significance for the $\chi^2_{(\alpha)}$ stopping criterion is recommended. In section 5.2 orders of variables selected by alternative selection criteria are compared with those of variables selected by the likelihood ratio criterion (LR). In section 5.3 effects of different structures of predictor variables on ARR, Bias, and $E\hat{R}R$ are investigated. In section 5.4 sizes of subset models determined by different stopping criteria are compared. Finally in section 5.5 performances of the seven selection criteria and the five stopping criteria are compared in terms of $E\hat{R}R$.

For convenience, Δ^2 is interchangeable with the symbol 'M' in the multivariate normal case and p_0 and $(p_1 - p_0)$ are interchangeable with the symbols 'B' and 'M', respectively, in the multivariate binary case.

5.1 Best α level for the $\chi^2_{(\alpha)}$ stopping criterion

As indicated in section 3.3, the $\chi^2_{(\alpha)}$ stopping criterion depend on a choice of 'significance level' for its application. To cover the possible values of α, the following 19 'significance levels' are considered in the study:

$$\alpha = 0.05 \text{ to } 0.95 \text{ in steps of } 0.05. \qquad (5.1.1)$$

The best α level is then defined to be that for which the grand mean of $E\hat{R}R$ over 48 sampling situations in the multivariate normal case and over 81 sampling situations in the multivariate binary case is minimum.

This section has two main purposes: 1) to recommend the best α level of significance for the $\chi^2_{(\alpha)}$ stopping criterion; and 2) to investigate the effects of the five factors P, V, Δ^2, D, and N on the best α level in the multivariate normal case, and of the four factors P, B, M, and N on the best α level in the multivariate binary case.

5.1.1 Multivariate normal case

Table 5.1.1.1 gives the mean of $\hat{ERR}$ over 48 sampling situations for each level of significance with the seven selection criteria (LR, LS, WD, SC, PH, LK, and SW) with application to the $\chi^2_{(\alpha)}$ stopping criterion. The best α level among 19 α levels was found to be 0.20. The minimum values of $\hat{ERR}$ occur at $\alpha = 0.20$ for all seven selection criteria.

Figure 5.1.1.1 is a representative graph of the mean of ARR and Bias over 48 sampling situations by α level of significance for the LR selection criterion. The word 'representative' is used to indicate that the graphs for other six selection criteria had this general appearance. This figure shows that ARR and Bias are monotonically decreasing and increasing functions of α level, respectively. In other words, ARR and Bias are monotonically decreasing and increasing functions of number of predictor variables in the model, respectively.

As defined in (3.1.2.3), $\hat{ERR}$ depends on these two opposite functions:

$$\hat{ERR} = ARR + \omega(\hat{\pi}) \qquad (5.1.1.1)$$

where $\omega(\hat{\pi})$ is the estimated bias of ARR. Figure 5.1.1.2 shows that, taking ARR and Bias functions into account simultaneously, $\hat{ERR}$ function is U-shaped with the minimum at $\alpha = 0.20$.

In order to assess the effects of the five factors P, V, Δ^2, D, and N on the best α level, we need the best α level for each of 48 sampling situations. Table 5.1.1.2

gives the best α levels for 48 sampling situations. The best α level for each sampling situation is defined to be that for which the mean of $\hat{ERR}$ over 20 replications is minimum. The best α levels are between 0.05 and 0.40.

Table 5.1.1.3 gives the means of the best α levels by level for the five factors P, V, Δ^2, D, and N. The results of table 5.1.1.3 are graphically presented in figures 5.1.1.3 through 5.1.1.7 for the five factors P, V, Δ^2, D, and N, respectively.

A response surface analysis (Cochran and Cox, 1959; chapter 8A) was employed to assess the effects of the five factors P, V, Δ^2, D, and N on the best α levels in table 5.1.1.2. Before presenting the principal findings, some remarks for the response surface analysis may be made. First, it uses the star (-2 and +2) and centre levels (0) in conjunction with the factorial levels (-1 and +1). Hence all 48 sampling situations are utilized in the data analysis. Second, there are 21 parameters to estimate when a quadratic surface is fitted to the data which has 48 observations. Third, since the centre level (0) has been independently replicated six times, it is possible to decompose the residual sum of squares (SS_r) into two components, typically called the sum of squares due to 'experimental error' (SS_e) and the sum of squares due to 'lack-of-fit' (SS_{lof}); that is,

$$SS_r = SS_e + SS_{lof} \,. \qquad (5.1.1.2)$$

The lack-of-fit test is the ratio of mean squares for lack-of-fit and experimental error; that is,

$$F_{lof} = \frac{SS_{lof} / v}{SS_e / m} \sim F_{(v,m)} \,, \qquad (5.1.1.3)$$

where $F_{(v,m)}$ is F distribution with v and m degrees of freedom.

Table 5.1.1.4 shows the results for the response surface analysis of the best α levels for the five factor P, V, Δ^2, D, and N. In this case the critical value at α =

0.05 for the lack-of-fit test is $F_{(22,5)} = 4.54$. Since the lack-of-fit test is not significant ($p > 0.308$), the quadratic surface fits the data well.

It is noticed that only the factor P is statistically significant ($p < 0.01$) and that the factor D is marginally not significant ($p < 0.09$).

The result for the factor P suggests that the best α level depends on the number of predictor variables in the data and that α should be specified with a value that increases as number of predictor variables increases (see figure 5.1.1.3). This result makes sense in that when there are many predictor variables, there is a lot of competition among them to be included into the model. An individual variable has less chance to be selected, thus it should be given a more generous α level.

As explained in section 4.1, the factor D indicates the rate of increase from $\Delta^2_{(q)}$ to $\Delta^2_{(p)}$; the larger D is, the slower is the rate of increase from $\Delta^2_{(q)}$ to $\Delta^2_{(p)}$. Then the result for the factor D implies that when only a small improvement in discriminating two populations can be obtained by adding the other variables, α should be smaller (see figure 5.1.1.6). Thus we would not include any variables which have little additional information about the populations.

5.1.2 Multivariate binary case

The tables and graphs for the multivariate binary case are analogous to those in the multivariate normal case.

Table 5.1.2.1 gives the mean of $\hat{ERR}$ over 81 sampling situations for each level of significance with the seven selection criteria (LR, LS, WD, SC, PH, LK, and SW) for the $\chi^2_{(\alpha)}$ stopping criterion. The best α level among 19 α levels was found to be 0.15. The minimum values of $\hat{ERR}$ occur at $\alpha = 0.15$ for all seven selection criteria.

Figure 5.1.2.1 displays the mean of ARR and Bias over all 81 sampling situations by α level of significance for the LR selection criterion. Figure 5.1.2.2 shows that $\hat{ERR}$ function is U-shaped with the minimum at $\alpha = 0.15$. Both figures 5.1.2.1 and 5.1.2.2 are very similar to the corresponding figures 5.1.1.1 and 5.1.1.2, respectively, in the multivariate normal case.

Table 5.1.2.2 gives the best α levels for 81 sampling situations. The definition of the best α level for each sampling situation is the same as for the multivariate normal case. The best α levels are between 0.05 and 0.40.

Table 5.1.2.3 gives the means of the best α levels by level for the four factors P, B, M, and N. The results of table 5.1.2.3 are graphically presented in figures 5.1.2.3 through 5.1.2.6 for the four factors P, B, M, and N, respectively.

An analysis of variance was employed to assess the effects of the four factors P, B, M, and N on the best α levels in table 5.1.2.2. We note that there is only one observation per cell; that is, there is the best α level for each sampling situation. With only a single observation per cell, the error term, that is the mean square for error is not estimable (see e.g. Montgomery, 1976, p.211). Thus, the hypotheses concerning main effects and interactions can not be tested. The usual way to remedy this problem is to assume that certain high-order interactions are negligible, and then since their mean square have all expectation σ^2, they may be combined to estimate the error. This approach has been taken in this study; it was assumed that the effects of third-order and fourth-order interactions are zeroes in the analysis of variance in table 5.1.2.4.

The coefficient of determination, that is R^2 is used for model adequacy checking. This is the proportion of variability in the data explained by a model. Thus the results based on a small value of R^2 should be used with caution.

Table 5.1.2.4 shows the results for the analysis of variance on the best α levels for four factor P, B, M, and N. It is noted that R^2 is 0.66 which may indicate the ANOVA does not fit the data well; the magnitude as well as the significance of the F-ratio will be considered in the following discussions.

It is observed that the factor M is highly significant (F-ratio = 21.49) and that the factor P is moderately significant (F-ratio = 7.70).

For the factor M, the larger the factor M is, the smaller the best α level is (see figure 5.1.2.5). This result implies that the more distinct the two populations are, the smaller the α level should be.

For the factor P, the best α level increases with the number of predictor variables in the data (see figure 5.1.2.3). This result agrees with that in the multivariate normal case and the same interpretation be made as previously.

5.1.3 Conclusions

1. We recommend that $0.15 \leq \alpha \leq 0.20$ for the $\chi^2_{(\alpha)}$ stopping criterion be used in forward stepwise logistic regression for the purpose of prediction.

2 For the multivariate normal case, the factors P and D have large effects on the choice of the best α level.

3. For the multivariate binary case, the factors P and M have large effects on the choice of the best α level.

Table 5.1.1.1
Mean of $\hat{ERR}$ for all 48 sampling situations for each level of significance with the seven selection criteria for the $\chi^2_{(\alpha)}$ stopping criterion.

α	Selection criteria	$\overline{\hat{ERR}}$[a]	α	Selection criteria	$\overline{\hat{ERR}}$[a]
	LR	0.1664		LR	0.1646
	LS	0.1664		LS	0.1645
	WD	0.1670		WD	0.1651
0.05	SC	0.1666	0.10	SC	0.1647
	PH	0.1666		PH	0.1647
	LK	0.1666		LK	0.1647
	SW	0.1668		SW	0.1649
	LR	0.1644		LR	0.1637
	LS	0.1643		LS	0.1637
	WD	0.1647		WD	0.1643
0.15	SC	0.1644	0.20[b]	SC	0.1639
	PH	0.1644		PH	0.1639
	LK	0.1644		LK	0.1639
	SW	0.1644		SW	0.1643
	LR	0.1641		LR	0.1650
	LS	0.1642		LS	0.1651
	WD	0.1643		WD	0.1653
0.25	SC	0.1642	0.30	SC	0.1652
	PH	0.1642		PH	0.1652
	LK	0.1642		LK	0.1652
	SW	0.1645		SW	0.1652
	LR	0.1661		LR	0.1679
	LS	0.1661		LS	0.1680
	WD	0.1664		WD	0.1680
0.35	SC	0.1662	0.40	SC	0.1680
	PH	0.1662		PH	0.1681
	LK	0.1662		LK	0.1681
	SW	0.1665		SW	0.1680
	LR	0.1687		LR	0.1697
	LS	0.1687		LS	0.1697
	WD	0.1686		WD	0.1696
0.45	SC	0.1686	0.50	SC	0.1697
	PH	0.1686		PH	0.1697
	LK	0.1686		LK	0.1697
	SW	0.1686		SW	0.1696

Table 5.1.1.1 (continued)
Mean of $\hat{ERR}$ for all 48 sampling situations for each level of significance with the seven selection criteria for the $\chi^2_{(\alpha)}$ stopping criterion.

α	Selection criteria	$\overline{\hat{ERR}}$[a]	α	Selection criteria	$\overline{\hat{ERR}}$[a]
0.55	LR	0.1708	0.60	LR	0.1726
	LS	0.1709		LS	0.1726
	WD	0.1709		WD	0.1725
	SC	0.1709		SC	0.1725
	PH	0.1709		PH	0.1725
	LK	0.1709		LK	0.1725
	SW	0.1710		SW	0.1725
0.65	LR	0.1740	0.70	LR	0.1751
	LS	0.1741		LS	0.1751
	WD	0.1739		WD	0.1751
	SC	0.1740		SC	0.1751
	PH	0.1740		PH	0.1751
	LK	0.1740		LK	0.1751
	SW	0.1739		SW	0.1751
0.75	LR	0.1762	0.80	LR	0.1773
	LS	0.1762		LS	0.1772
	WD	0.1762		WD	0.1772
	SC	0.1761		SC	0.1772
	PH	0.1761		PH	0.1772
	LK	0.1761		LK	0.1772
	SW	0.1762		SW	0.1773
0.85	LR	0.1780	0.90	LR	0.1780
	LS	0.1780		LS	0.1780
	WD	0.1779		WD	0.1780
	SC	0.1780		SC	0.1780
	PH	0.1780		PH	0.1780
	LK	0.1780		LK	0.1780
	SW	0.1780		SW	0.1780
0.95	LR	0.1781			
	LS	0.1781			
	WD	0.1781			
	SC	0.1781			
	PH	0.1781			
	LK	0.1781			
	SW	0.1781			

[a]Standard errors are in the range of 0.00199 and 0.00213

[b]The best α level of significance at which the minimum value of $\overline{\hat{ERR}}$ occurs

Table 5.1.1.2

The best α levels for 48 sampling situations

Sampling situation	$^{a}\overline{\hat{ERR}}_{min}$	α	Sampling Situation	$^{a}\overline{\hat{ERR}}_{min}$	α
1	0.222	0.10	25	0.230	0.10
2	0.221	0.40	26	0.222	0.05
3	0.231	0.25	27	0.228	0.05
4	0.204	0.10	28	0.230	0.20
5	0.106	0.10	29	0.111	0.15
6	0.086	0.40	30	0.099	0.15
7	0.107	0.05	31	0.109	0.10
8	0.099	0.20	32	0.101	0.35
9	0.237	0.15	33	0.150	0.10
10	0.216	0.25	34	0.156	0.15
11	0.224	0.10	35	0.147	0.30
12	0.211	0.10	36	0.225	0.10
13	0.102	0.20	37	0.300	0.15
14	0.108	0.20	38	0.062	0.20
15	0.122	0.20	39	0.158	0.30
16	0.083	0.20	40	0.170	0.20
17	0.220	0.15	41	0.164	0.05
18	0.212	0.20	42	0.163	0.25
19	0.236	0.25	43	0.152	0.20
20	0.218	0.20	44	0.145	0.30
21	0.103	0.25	45	0.159	0.10
22	0.096	0.20	46	0.155	0.25
23	0.097	0.05	47	0.147	0.20
24	0.104	0.15	48	0.161	0.20

[a]Based on 20 replications for each sampling situation.

Table 5.1.1.3
Means of the best α levels over the level of the five factors P, V, Δ^2, D, and N

Factor	Level	Value	Number of sampling situations	$\bar{\alpha}$
P	-2	5	1	0.05
	-1	10	16	0.14
	0	15	14	0.20
	+1	20	16	0.21
	+2	25	1	0.25
V	-2	0.2	1	0.30
	-1	0.4	16	0.19
	0	0.6	14	0.18
	+1	0.8	16	0.16
	+2	1.0	1	0.20
Δ^2	-2	1.0	1	0.15
	-1	1.5	16	0.17
	0	2.0	14	0.19
	+1	2.5	16	0.18
	+2	3.0	1	0.20
D	-2	0.2	1	0.30
	-1	0.4	16	0.19
	0	0.6	14	0.19
	+1	0.8	16	0.16
	+2	1.0	1	0.10
N	-2	100	1	0.10
	-1	150	16	0.19
	0	200	14	0.20
	+1	250	16	0.16
	+2	300	1	0.15

Table 5.1.1.4

Response surface analysis of the best α levels for the five factors P, V, M, D, and N

Factor	d.f	$\hat{\beta}$	s.e	t-value	p-value
[a]P	1	3.75	1.29	2.90	0.01
V	1	-1.75	1.29	-1.35	0.19
M	1	1.00	1.29	0.77	0.45
D	1	-2.25	1.29	-1.74	0.09
N	1	-0.75	1.29	-0.58	0.57
P^2	1	-1.41	1.48	-0.95	0.35
V^2	1	1.09	1.48	0.73	0.47
M^2	1	-0.79	1.48	-0.53	0.60
D^2	1	-0.16	1.48	-0.11	0.91
N^2	1	-2.04	1.48	-1.37	0.18
PV	1	-0.63	1.45	-0.43	0.67
PM	1	1.25	1.45	0.86	0.40
PD	1	-0.63	1.45	-0.43	0.70
PN	1	-0.94	1.45	-0.65	0.52
VM	1	-0.63	1.45	-0.43	0.70
VD	1	1.88	1.45	1.30	0.21
VN	1	2.19	1.45	1.51	0.14
MD	1	2.50	1.45	1.73	0.10
MN	1	0.31	1.45	0.22	0.83
DN	1	-0.31	1.45	-0.22	0.83

[a]significant with $p < 0.05$

Lack-of-fit test: $F_{22,5} = 1.634$ ($p > 0.308$)

Figure 5.1.1.1

Mean of ARR and Bias by alpha level of significance

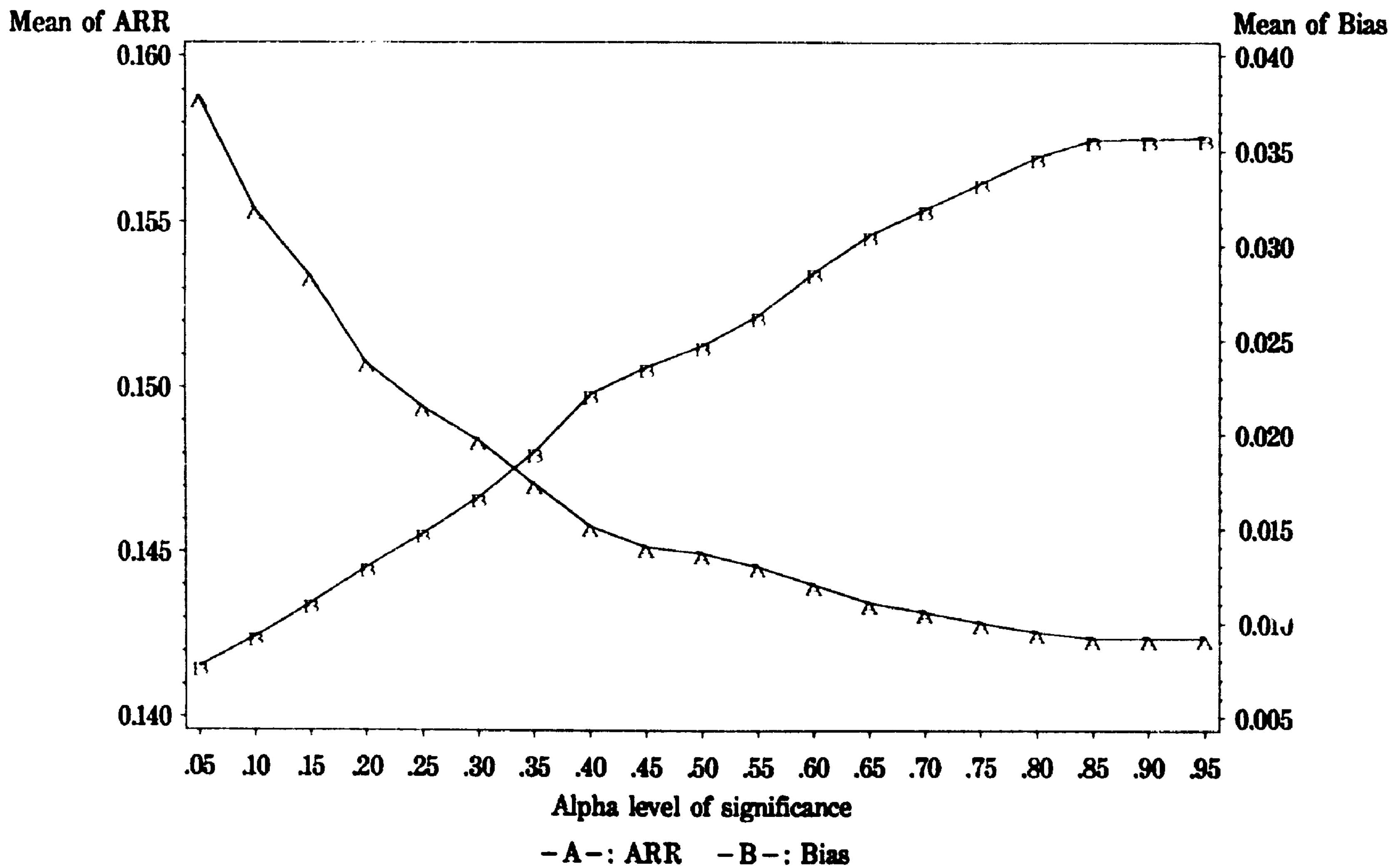


Figure 5.1.12

Mean of estimated ERR by alpha level of significance

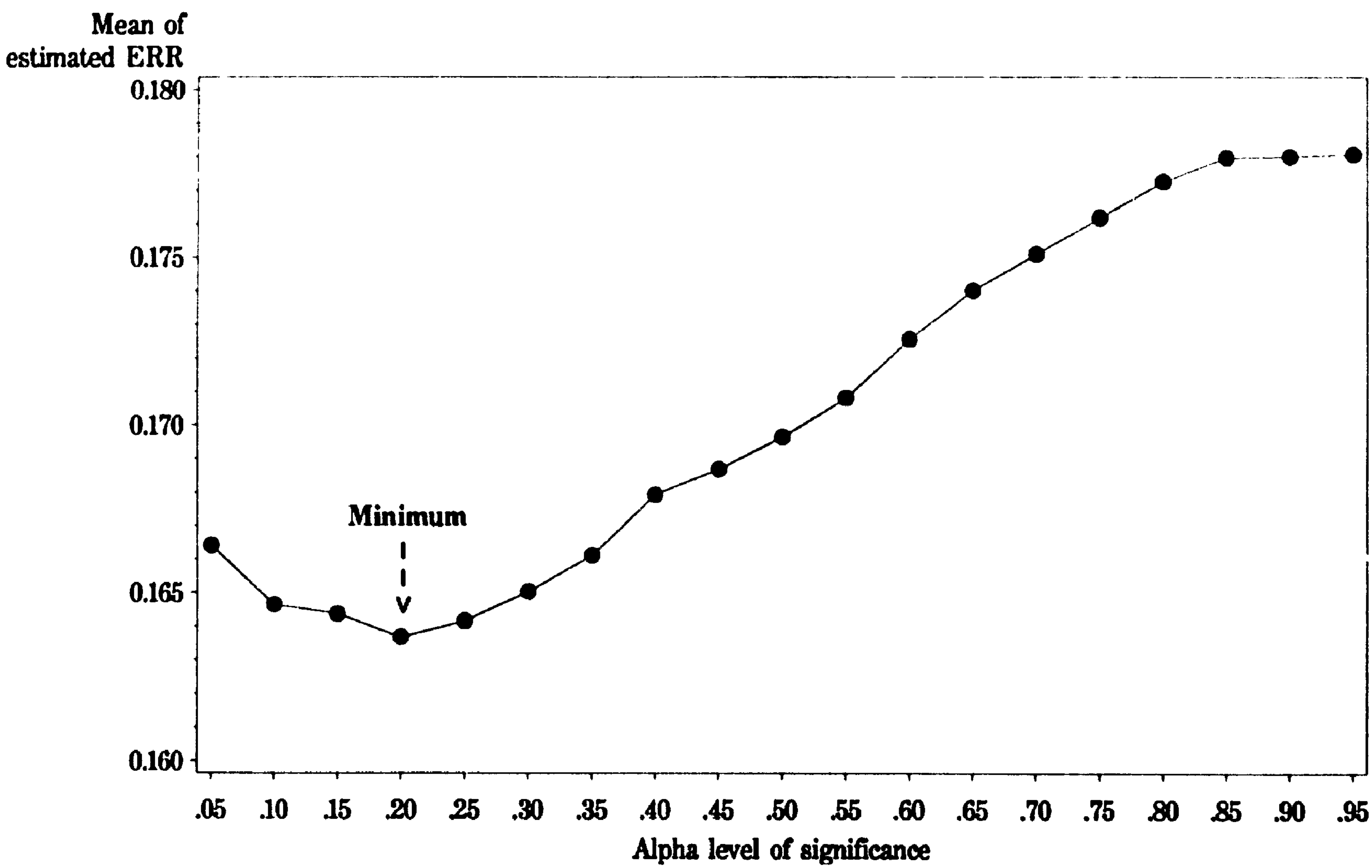


Figure 5.1.1.3

Mean of the best alpha levels for the factor P

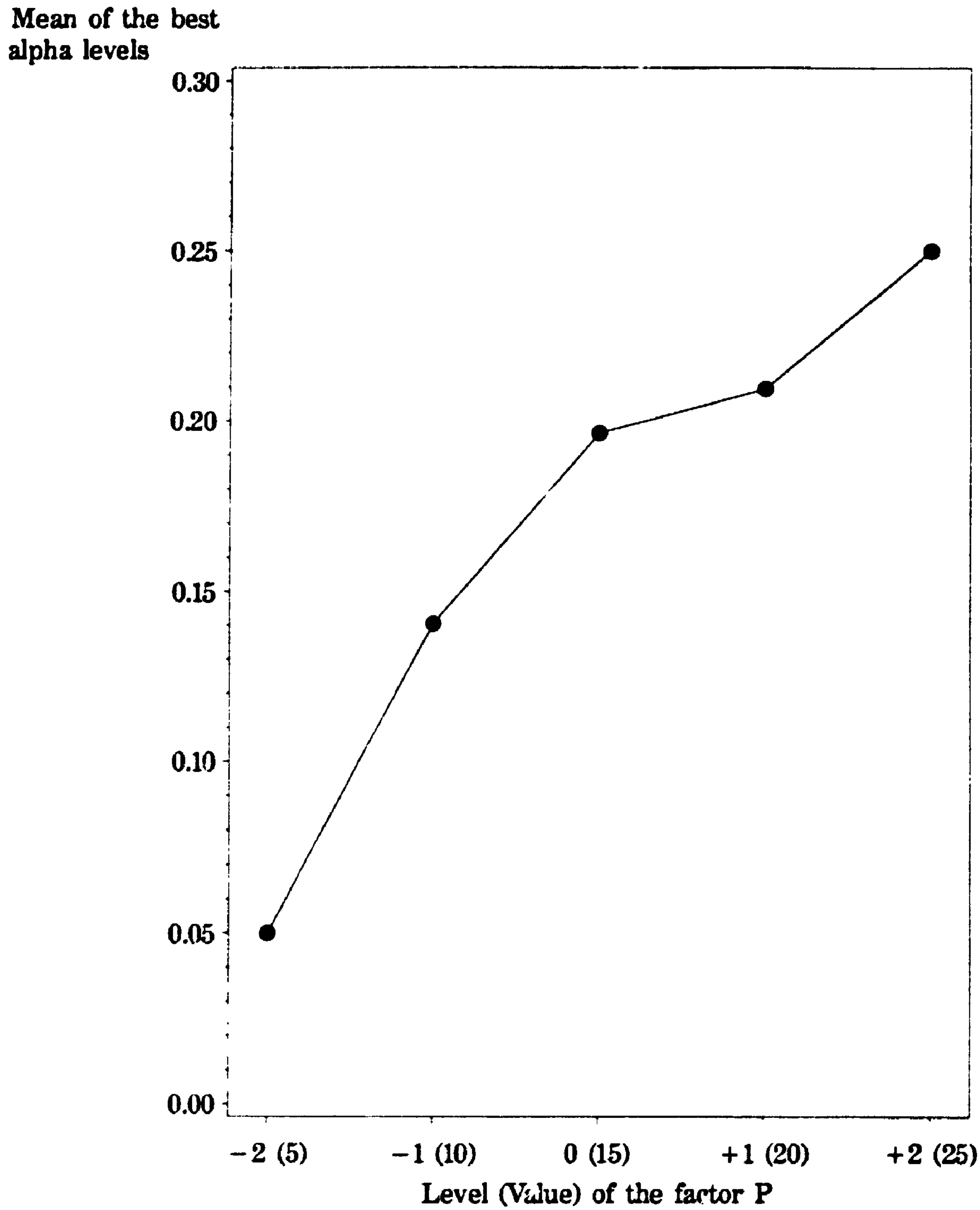


Figure 5.1.1.4

Mean of the best alpha levels for the factor V

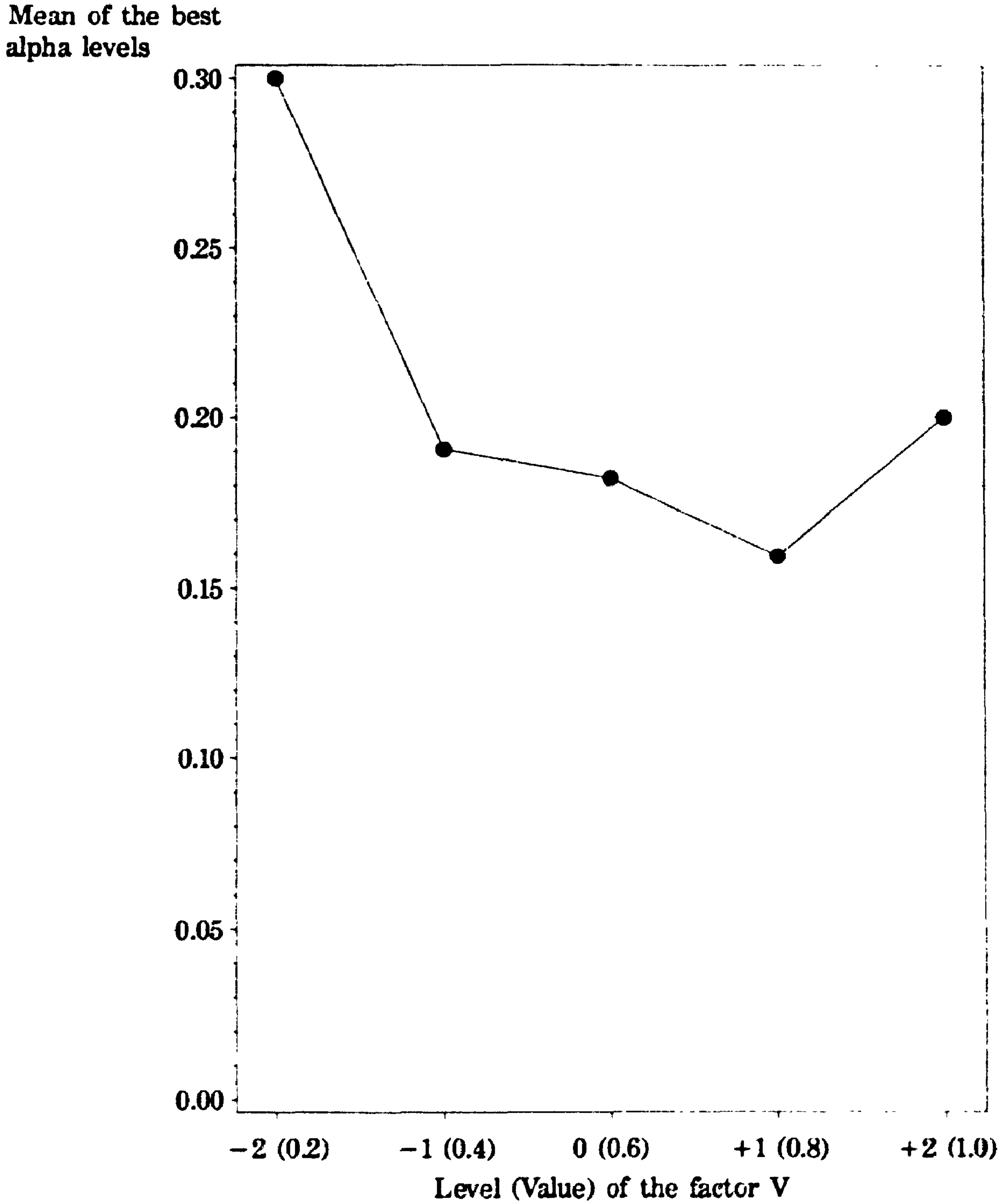


Figure 5.1.1.5
Mean of the best alpha levels for the factor M

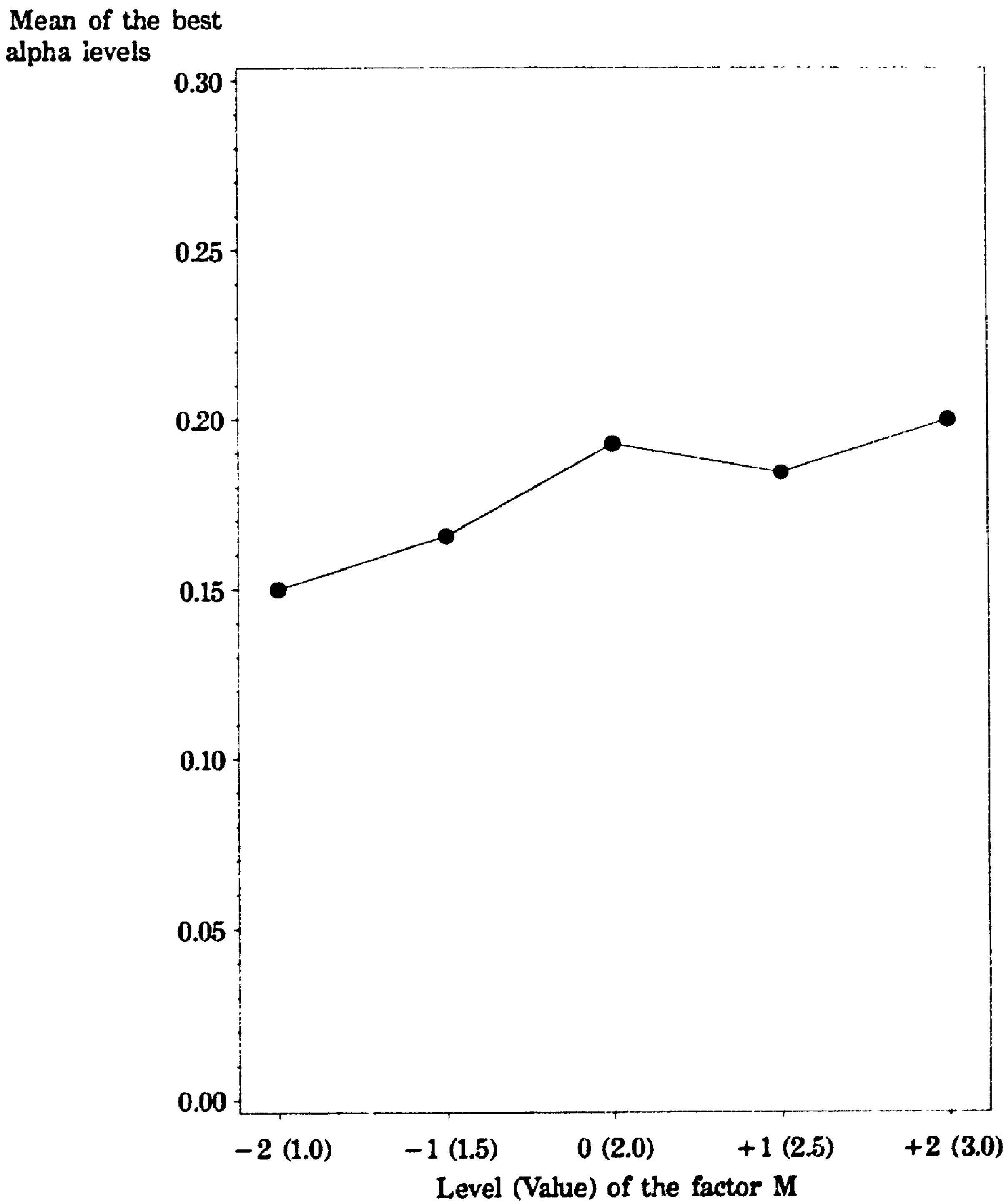


Figure 5.1.1.6

Mean of the best alpha levels for the factor D

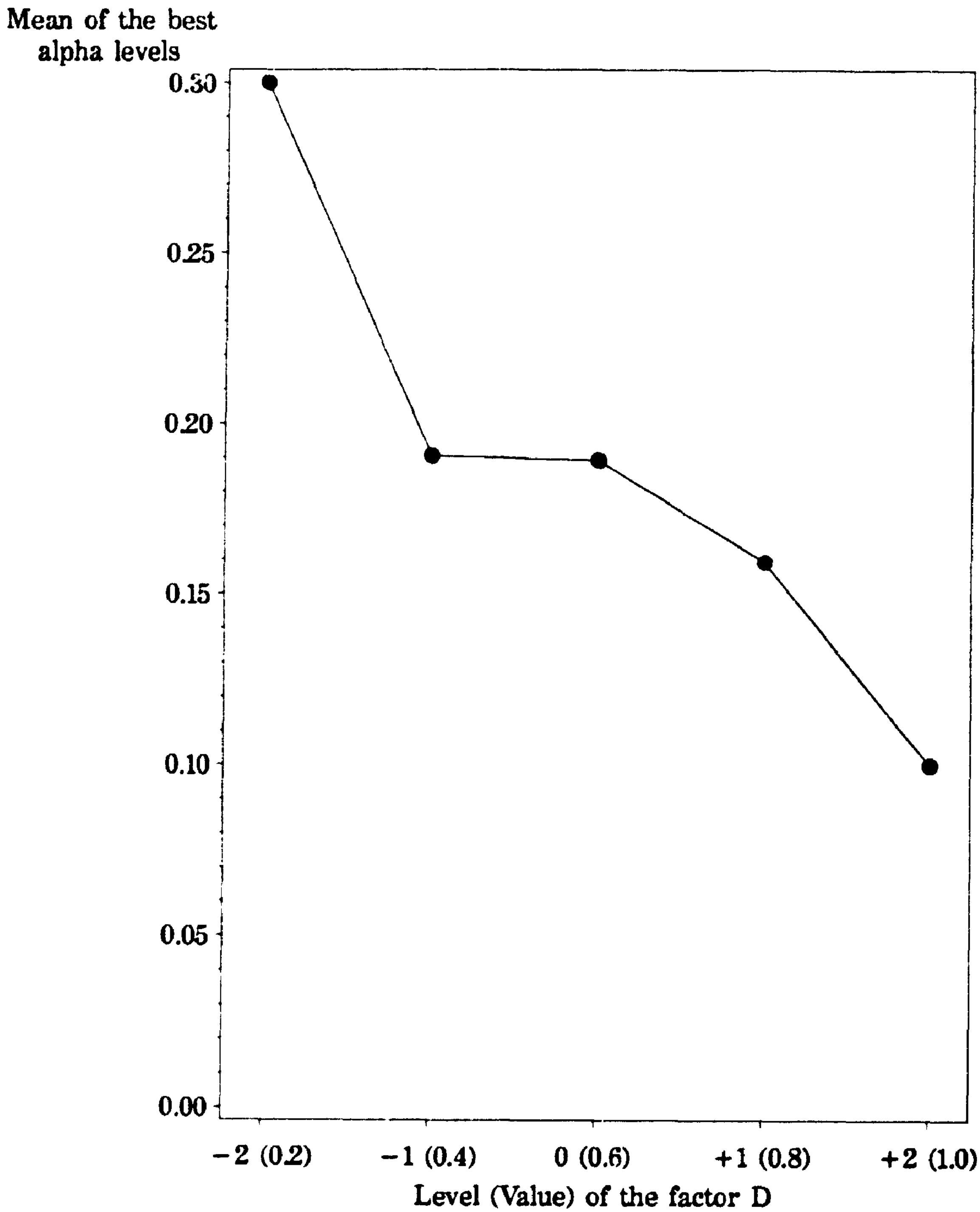


Figure 5.1.1.7

Mean of the best alpha levels for the factor N

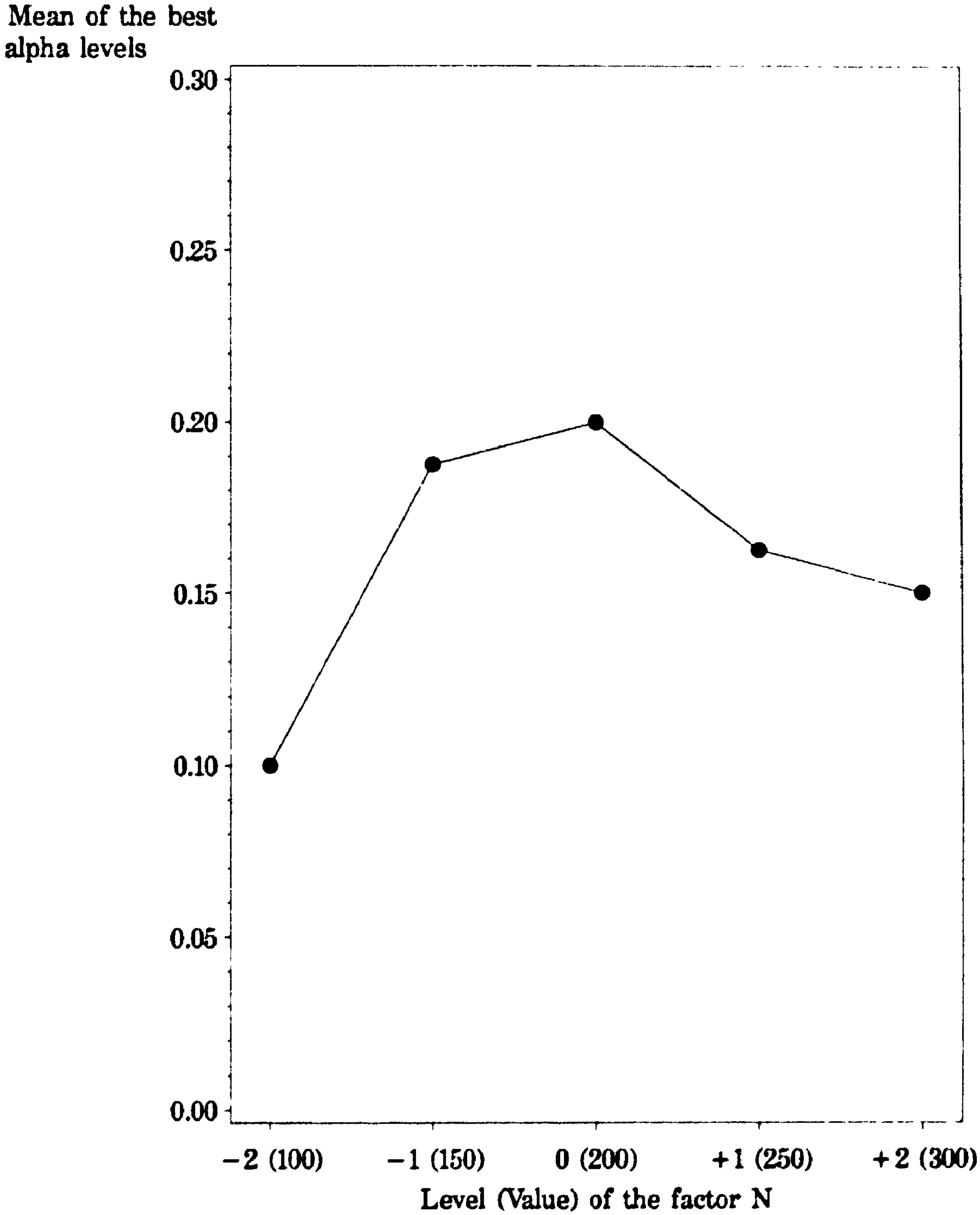


Table 5.1.2.1
Mean of $\hat{ERR}$ for all 81 sampling situations for each level of significance with the seven selection criteria for the $\chi^2_{(\alpha)}$ stopping criterion.

α	Selection criteria	$\overline{\hat{ERR}}$[a]	α	Selection criteria	$\overline{\hat{ERR}}$[a]
0.05	LR	0.2590	0.10	LR	0.2568
	LS	0.2591		LS	0.2571
	WD	0.2596		WD	0.2573
	SC	0.2593		SC	0.2572
	PH	0.2593		PH	0.2572
	LK	0.2593		LK	0.2572
	SW	0.2620		SW	0.2577
0.15[b]	LR	0.2567	0.20	LR	0.2571
	LS	0.2569		LS	0.2572
	WD	0.2569		WD	0.2572
	SC	0.2568		SC	0.2572
	PH	0.2568		PH	0.2572
	LK	0.2568		LK	0.2572
	SW	0.2570		SW	0.2571
0.25	LR	0.2583	0.30	LR	0.2593
	LS	0.2584		LS	0.2594
	WD	0.2584		WD	0.2594
	SC	0.2584		SC	0.2594
	PH	0.2584		PH	0.2594
	LK	0.2584		LK	0.2594
	SW	0.2581		SW	0.2592
0.35	LR	0.2603	0.40	LR	0.2615
	LS	0.2605		LS	0.2616
	WD	0.2604		WD	0.2615
	SC	0.2604		SC	0.2616
	PH	0.2604		PH	0.2616
	LK	0.2604		LK	0.2616
	SW	0.2603		SW	0.2616
0.45	LR	0.2627	0.50	LR	0.2639
	LS	0.2628		LS	0.2639
	WD	0.2628		WD	0.2639
	SC	0.2629		SC	0.2639
	PH	0.2629		PH	0.2639
	LK	0.2629		LK	0.2639
	SW	0.2627		SW	0.2639

Table 5.1.2.1 (continued)
Mean of $\hat{ERR}$ for all 81 sampling situations for each level of significance with the seven selection criteria for the $\chi^2_{(\alpha)}$ stopping criterion.

α	Selection criteria	$\overline{\hat{ERR}}$[a]	α	Selection criteria	$\overline{\hat{ERR}}$[a]
	LR	0.2645		LR	0.2653
	LS	0.2645		LS	0.2653
	WD	0.2644		WD	0.2653
0.55	SC	0.2645	0.60	SC	0.2653
	PH	0.2645		PH	0.2653
	LK	0.2645		LK	0.2653
	SW	0.2646		SW	0.2654
	LR	0.2657		LR	0.2661
	LS	0.2657		LS	0.2661
	WD	0.2657		WD	0.2661
0.65	SC	0.2657	0.70	SC	0.2661
	PH	0.2657		PH	0.2661
	LK	0.2657		LK	0.2661
	SW	0.2658		SW	0.2661
	LR	0.2663		LR	0.2664
	LS	0.2663		LS	0.2664
	WD	0.2663		WD	0.2664
0.75	SC	0.2663	0.80	SC	0.2664
	PH	0.2663		PH	0.2664
	LK	0.2663		LK	0.2664
	SW	0.2664		SW	0.2664
	LR	0.2664		LR	0.2664
	LS	0.2664		LS	0.2664
	WD	0.2664		WD	.2664
0.85	SC	0.2664	0.90	SC	[illegible]
	PH	0.2664		PH	0.266[illegible]
	LK	0.2664		LK	2664
	SW	0.2664		SW	0.2664
	LR	0.2664			
	LS	0.2664			
	WD	0.2664			
0.95	SC	0.2664			
	PH	0.2664			
	LK	0.2664			
	SW	0.2664			

[a]Standard errors are in the range of 0.00240 and 0.00262

[b]The best α level of significance at which the minimum value of $\overline{\hat{ERR}}$ occurs

Table 5.1.2.2

The best α levels for 81 sampling situations

Sampling situation	[a]$\overline{\hat{ERR}}_{min}$	α	Sampling Situation	[a]$\overline{\hat{ERR}}_{min}$	α
1	0.386	0.20	42	0.215	0.10
2	0.337	0.30	43	0.235	0.10
3	0.309	0.20	44	0.173	0.25
4	0.280	0.10	45	0.120	0.35
5	0.221	0.15	46	0.391	0.25
6	0.172	0.15	47	0.390	0.20
7	0.180	0.15	48	0.337	0.25
8	0.104	0.10	49	0.321	0.05
9	0.050	0.15	50	0.285	0.10
10	0.378	0.15	51	0.242	0.10
11	0.350	0.40	52	0.203	0.05
12	0.327	0.40	53	0.176	0.10
13	0.298	0.10	54	0.135	0.10
14	0.258	0.05	55	0.384	0.30
15	0.196	0.30	56	0.335	0.25
16	0.240	0.05	57	0.322	0.20
17	0.168	0.10	58	0.264	0.05
18	0.107	0.25	59	0.223	0.05
19	0.388	0.25	60	0.187	0.25
20	0.381	0.20	61	0.154	0.15

[a]Based on 20 replications for each sampling situation.

Table 5.1.2.2 (continued)

The best α levels for 81 sampling situations

Sampling situation	$^{a}\overline{\hat{ERR}}_{min}$	α	Sampling Situation	$^{a}\overline{\hat{ERR}}_{min}$	α
21	0.335	0.20	62	0.109	0.20
22	0.318	0.05	63	0.074	0.25
23	0.275	0.30	64	0.386	0.25
24	0.241	0.10	65	0.365	0.30
25	0.197	0.05	66	0.333	0.20
26	0.174	0.15	67	0.297	0.05
27	0.130	0.15	68	0.258	0.10
28	0.387	0.15	69	0.209	0.10
29	0.340	0.15	70	0.233	0.05
30	0.311	0.40	71	0.161	0.20
31	0.271	0.05	72	0.113	0.30
32	0.217	0.15	73	0.407	0.20
33	0.184	0.05	74	0.381	0.25
34	0.169	0.15	75	0.353	0.15
35	0.108	0.15	76	0.315	0.10
36	0.075	0.25	77	0.282	0.05
37	0.389	0.10	78	0.243	0.05
38	0.368	0.25	79	0.201	0.05
39	0.320	0.20	80	0.184	0.05
40	0.298	0.10	81	0.139	0.05
41	0.244	0.25			

[a]Based on 20 replications for each sampling situation.

Table 5.1.2.3
Means of the best α levels over the level of the four factors P, B, M, and N

Factor	Level	Value	Number of sampling situations	$\bar{\alpha}$
P	-1	10	27	0.12
	0	15	27	0.18
	+1	20	27	0.19
B	-1	0.2	27	0.17
	0	0.4	27	0.19
	+1	0.6	27	0.13
M	-1	0.1	27	0.24
	0	0.2	27	0.11
	+1	0.3	27	0.15
N	-1	150	27	0.18
	0	200	27	0.16
	+1	250	27	0.16

Table 5.1.2.4
Analysis of variance of the best α levels
for the four factors P, B, M, and N

Effect	d.f.	Mean Square	F-ratio	p-value
[a]P	2	386.11	7.70	0.001
[b]B	2	212.04	4.23	0.020
[a]M	2	1077.78	21.49	0.001
N	2	28.70	0.57	0.568
PB	4	113.43	2.26	0.076
PM	4	47.22	0.94	0.448
PN	4	21.76	0.43	0.784
BM	4	50.93	1.02	0.409
MN	4	56.48	1.13	0.355
Error	48	50.15		

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

$R^2 = 0.66$

Figure 5.1.2.1

Mean of ARR and Bias by alpha level of significance

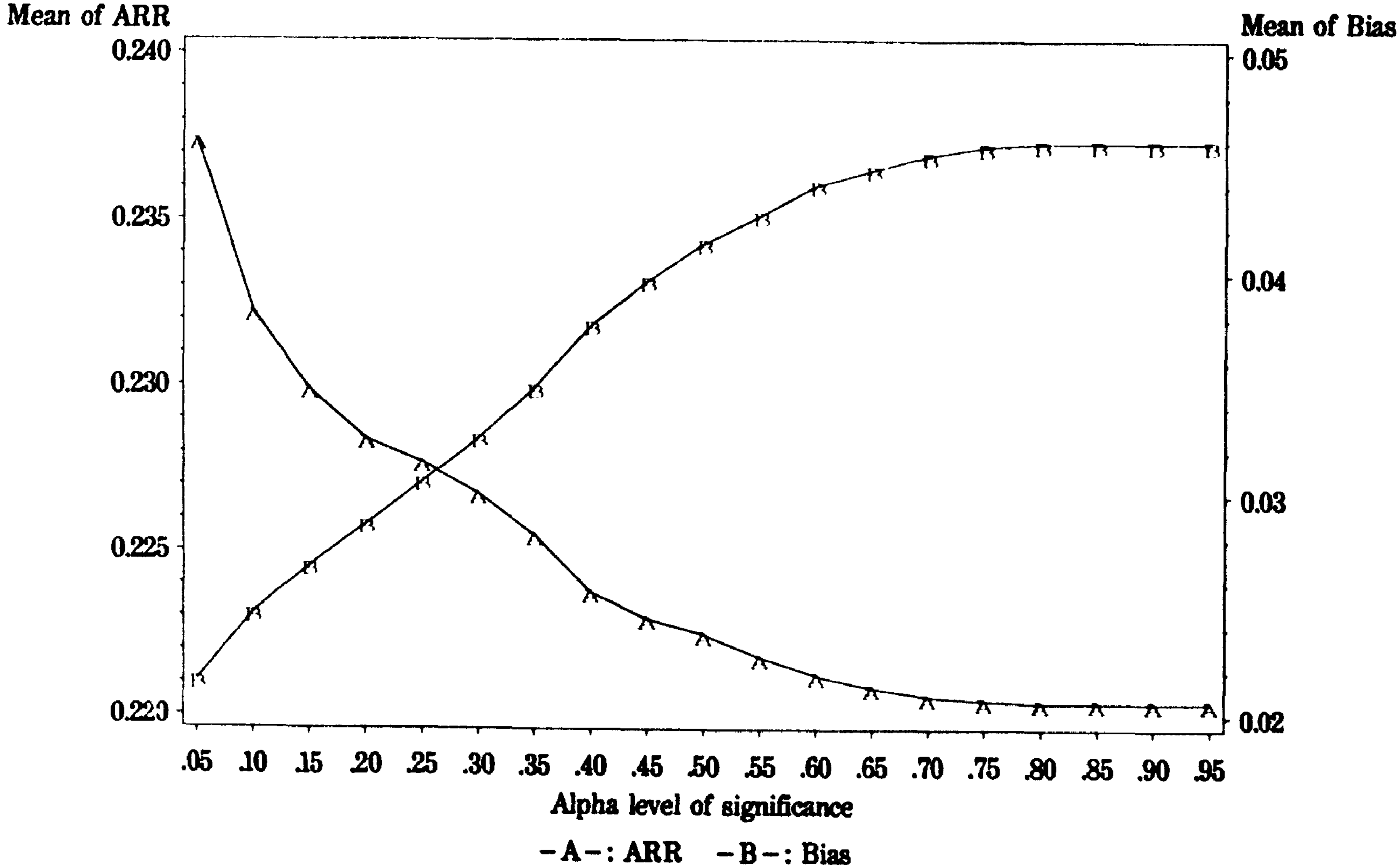


Figure 5.1.2.2

Mean of estimated ERR by alpha level of significance

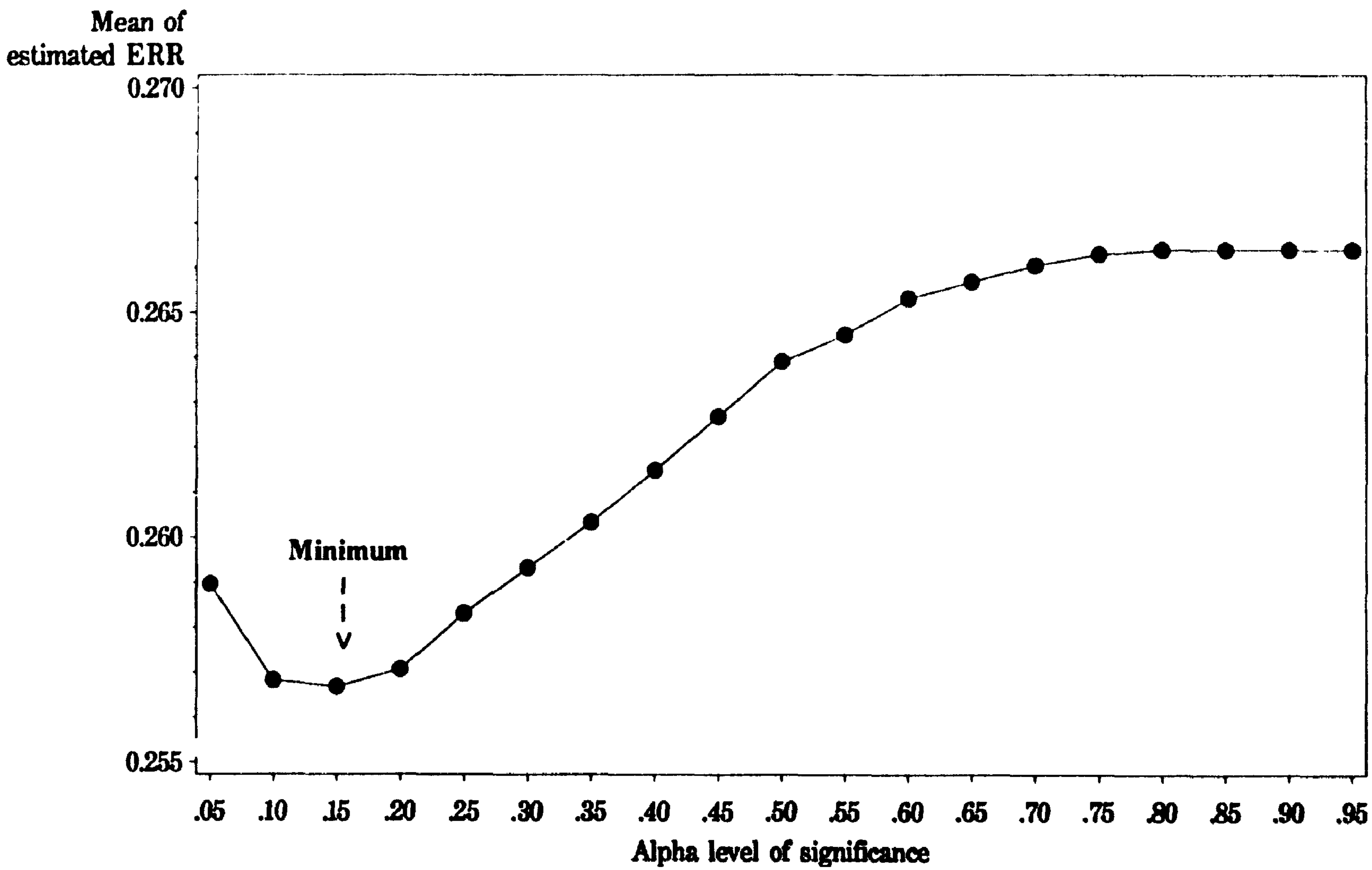


Figure 5.1.2.3

Mean of the best alpha levels for the factor P

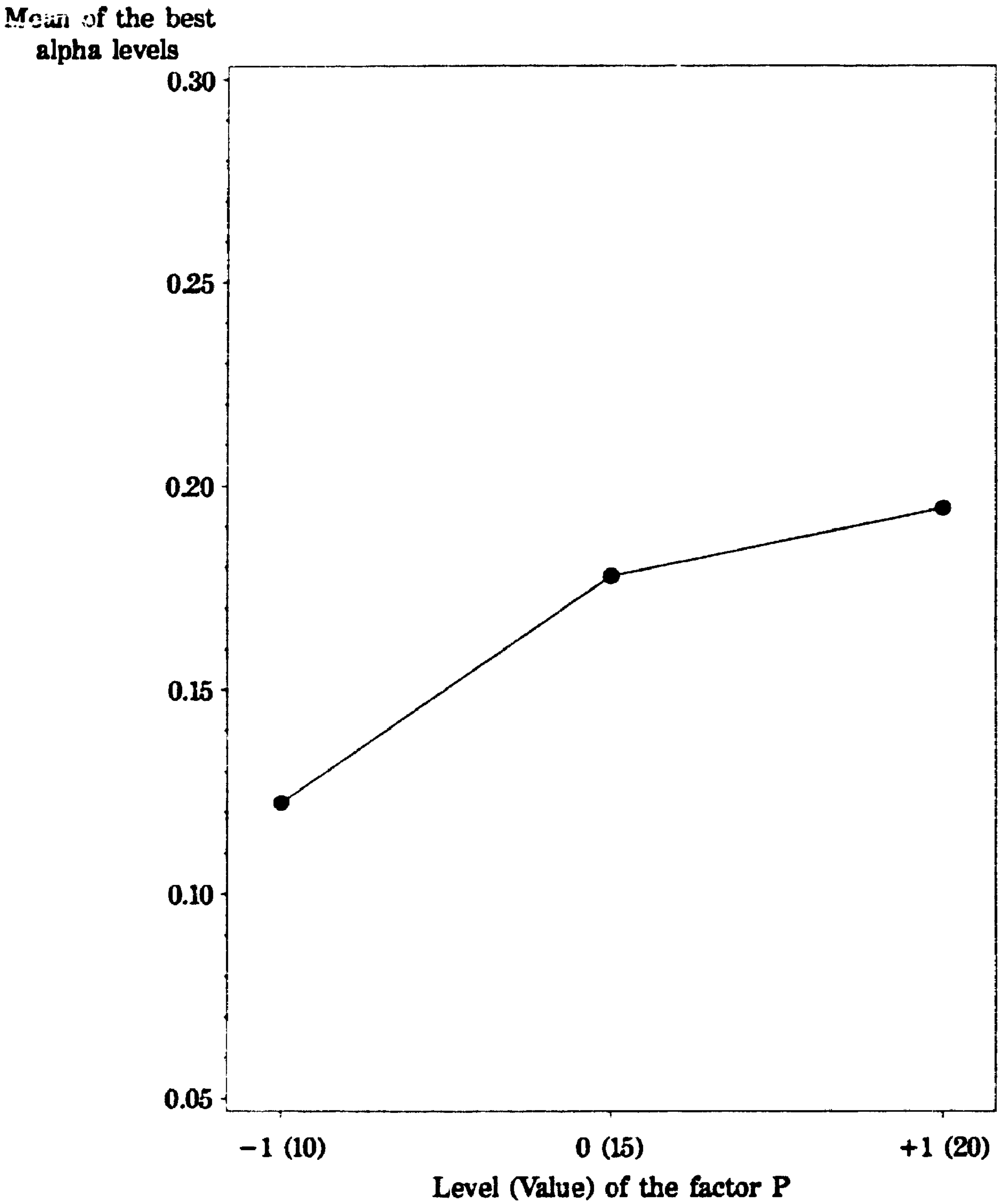


Figure 5.1.2.4

Mean of the best alpha levels for the factor B

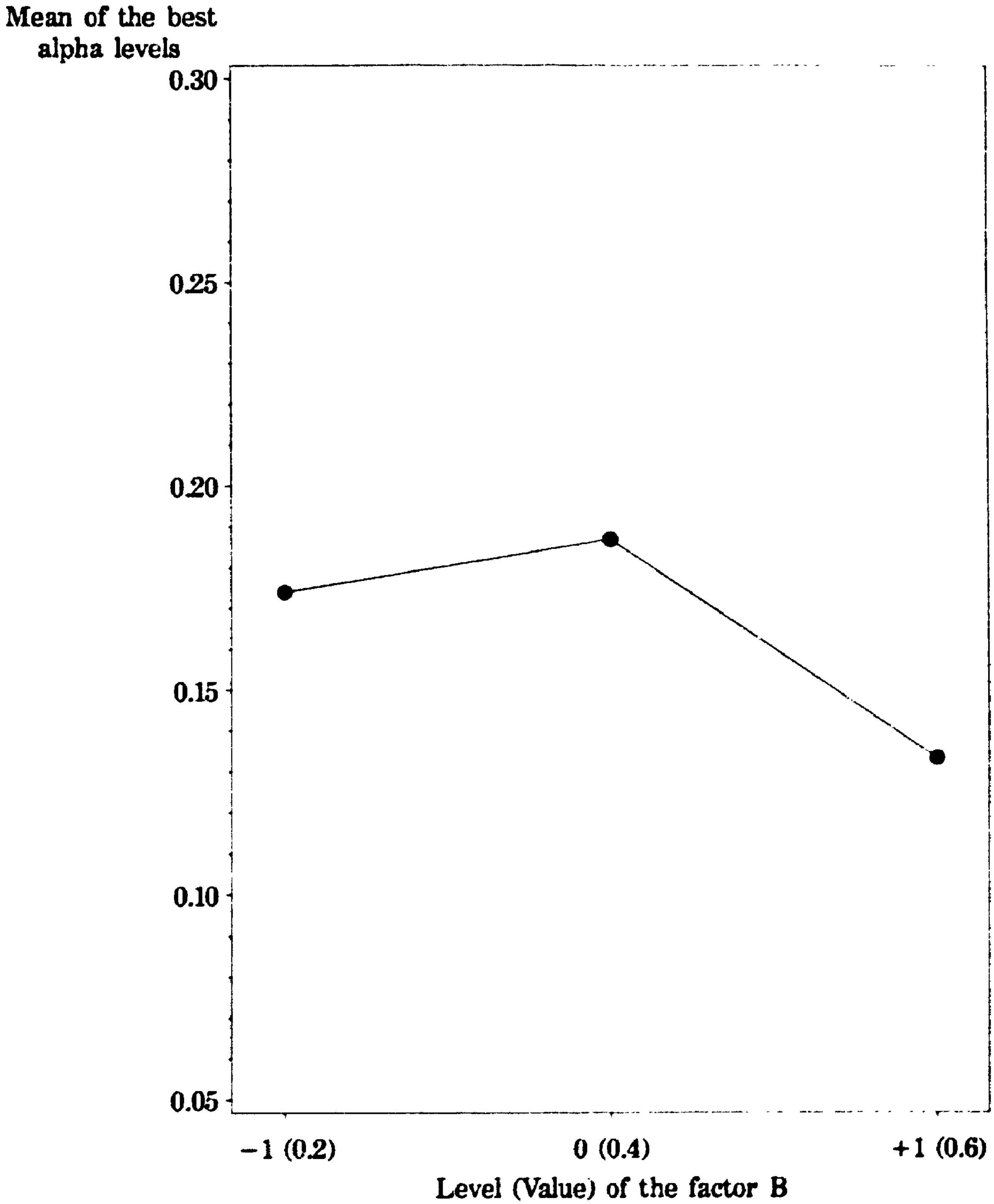


Figure 5.1.2.5

Mean of the best alpha levels for the factor M

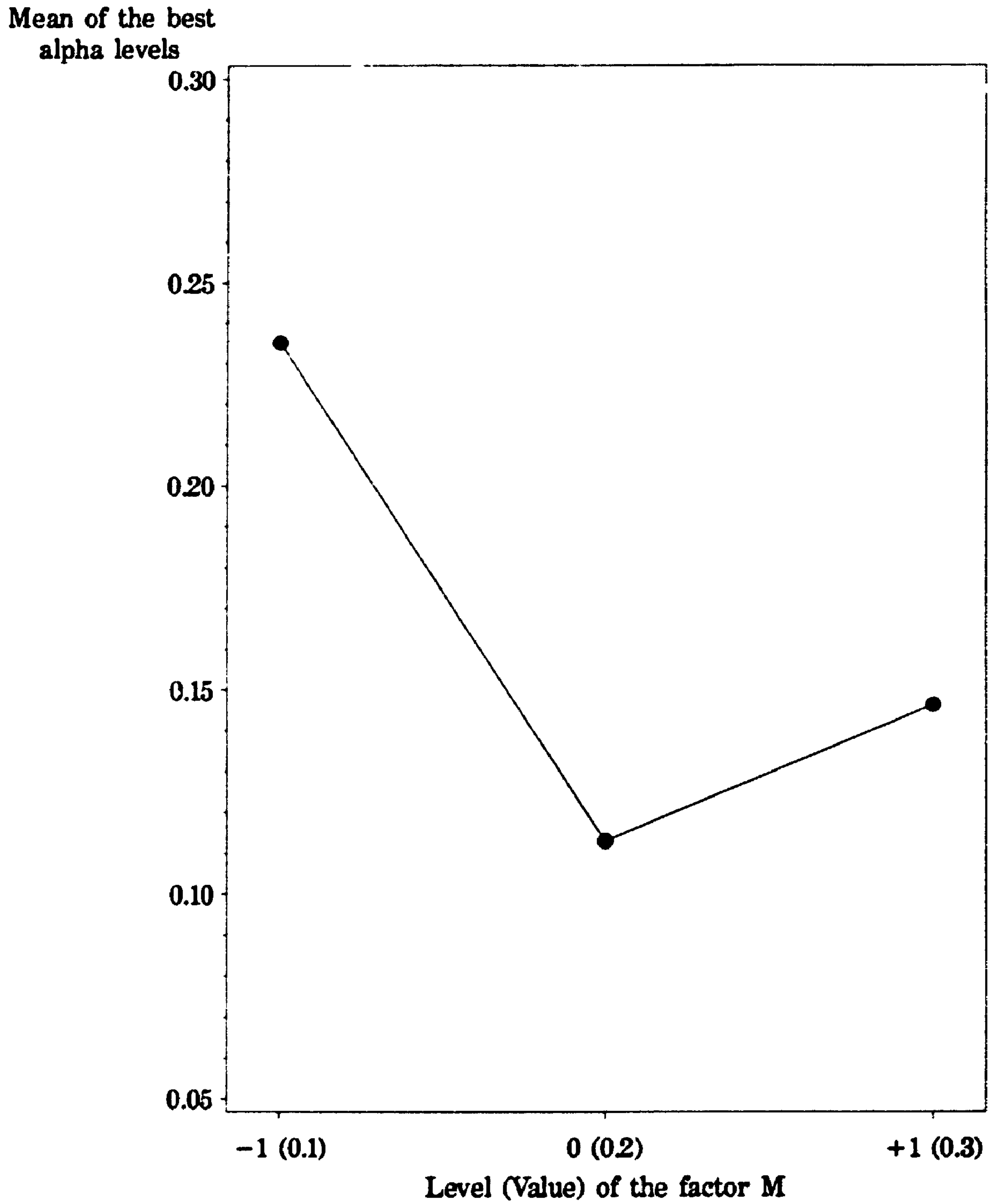
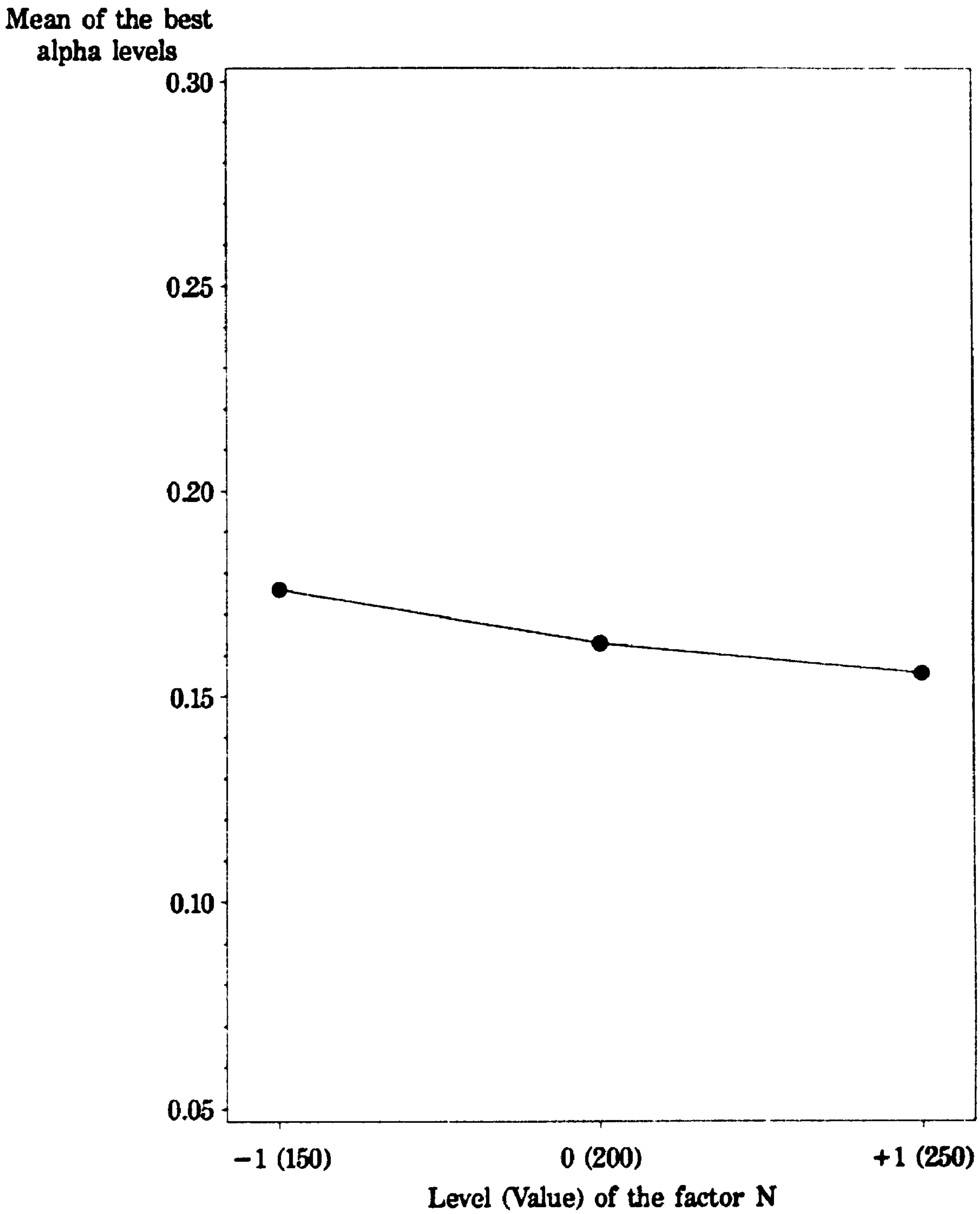


Figure 5.1.2.6

Mean of the best alpha levels for the factor N



5.2 Order of variables selected by the selection criteria

The seven selection criteria LR, LS, WD, SC, PH, LK, and SW, which can be used in forward stepwise logistic regression, have been defined in section 3.2. Among them the LR, WD, and SC selection criteria are known to be asymptotically equivalent statistics under certain regularity conditions. This means that they may select variables in different orders in a forward stepwise procedure in some situations.

This section has two main purposes: 1) to investigate how the six selection criteria LS, WD, SC, PH, LK, and SW perform in the selection of variables by comparing the orders of variables selected by these six selection criteria to those of variables selected by the LR selection criterion; and 2) to examine the effects of the five factors P, V, Δ^2, D, and N on the proportion of disagreement in the multivariate normal case, and of the four factors P, B, M, and N on the proportion of disagreement in the multivariate binary case. The definition of the proportion of disagreement will be defined below.

5.2.1 Multivariate normal case

The orders of variables selected by the six selection criteria LS, WD, SC, PH, LK, and WS were compared with those of variables selected by the LR selection criterion. Table 5.2.1.1 gives the proportion of 20 replications with different orders of variables for these six selection criteria for each of 48 sampling situations. We refer to this proportion as 'the proportion of disagreement'.

The LS selection criterion which is a first order approximation to the LR selection criterion selects exactly the same variables as the LR selection criterion for all 48 sampling situations.

For the WD selection criterion, the proportion of disagreement ranges from 0.00 (sampling situation = 3, 10, 17, and 18) to 1.00 (sampling situation = 13 and 38). The WD selection criterion performs poorly the selection of variables with an average of 34 percent chance of disagreement over all 48 sampling situations (see table 5.2.1.2). This result implies that the WD selection criterion is not a good approximation to the LR selection criterion.

The proportions of disagreement for the SC, PH, and LK selection criteria are the same except for two sampling situations (16 and 38). The range of the proportions of disagreement for these three selection criteria is from 0.00 (sampling situation = 3, 5, 9, 17, 20, 25, 29, 34, and 41) to 0.55 (sampling situation = 6). The number of sampling situations with a 0.00 proportion of disagreement for these three selection criteria, compared for the WD selection criterion, increases from 4 to 9, and none of 48 sampling situations for these three selection criteria had 1.00 proportion of disagreement. However, these three selection criteria are not universally superior to the WD selection criterion (see, for example, the results of sampling situation = 10, 18, and 37). These three selection criteria have an average of 13 percent chance of disagreement over all 48 sampling situations (see table 5.2.1.2). This result implies that they give much better approximations to the LR selection criterion than the WD selection criterion.

As explained in section 3.2, the SW selection criterion is different from the other selection criteria in the sense that it is not an approximate of the LR selection criterion. We note again that the SW selection criterion does not need MLE's for β to perform the selection of variables. In other words, the SW selection criterion is solely used as a tool for selecting variables, and MLE's for β are determined only when a variable has been selected in the model. There are two sampling situations (17 and 41) which have 0.00 proportion of disagreement and none of 48 sampling situations have 1.00 proportion of disagreement. There are 16 sampling situations (5, 7, 13, 17, 21, 23, 29, 30, 34, 38, 39, 41, 43, 44, 45, and 48) in which the SW selection criterion is equal or superior to the WD selection criterion. There are even two sampling situations (21 and 41) in which the SW selection criterion is equal or superior to the SC, PH, and LK selection criteria. The SW selection criterion has an average of 43 percent chance of disagreement over all 48 sampling situations (see table 5.2.1.2).

In any case discussed above, the proportion of disagreement for each selection criterion depends on the sampling situations. Thus we examine the effects of the five factors P, V, Δ^2, D, and N on the proportion of disagreement.

Table 5.2.1.2 gives the proportion of disagreement for the six selection criteria LS, WD, SC, PH, LK, and SW over the levels of the five factors P, V, Δ^2, D, and N. The results of table 5.2.1.2 are graphically presented in figures 5.2.1.1 through 5.2.1.5 for the five factors P, V, Δ^2, D, and N, respectively. These figures include only the WD and SC selection criteria because the LS selection criterion has always 0.00 percent disagreement and the results for the PH and LK selection criteria are almost the same as the SC selection criterion.

Response surface analyses were employed to assess the effects of the five factors P, V, Δ^2, D, and N on the proportion of disagreement in table 5.2.1.1 for the WD and SC selection criteria. Tables 5.2.1.3 and 5.2.1.4 give the results for the response surface analyses for the WD and SC selection criteria, respectively. Before discussing the principal findings, one remark with respect to the lack-of-fit tests may be made; the lack-of-fit tests for both tables are not statistically significant, hence the quadratic surfaces fit the data well.

Table 5.2.1.3 for the WD selection criterion shows that the factor M is highly significant ($t = 11.10$) and that the factor D is moderately significant ($t = 2.59$) (see figures 5.2.1.3 and 5.2.1.4 for the factors M and D, respectively). The term M^2 is moderately significant ($t = 2.40$), which means a non-linear effect for the factor M on the proportion of disagreement.

Before discussing the interaction terms, we note that the graph for an interaction can be drawn only at the factorial levels (± 1) because the inclusions of the star levels (± 2) and centre level (0) for the graph do not have all five levels for each of the five levels.

The two interaction terms, VM and VD, are moderately significant ($t = -2.17$ and -2.77, respectively). The graphs for the VM and VD interactions are presented in figures 5.2.1.6 and 5.2.1.7, respectively, to exhibit the phenomena of these interactions on the proportion of disagreement at the factorial levels (± 1). For the VM interaction, the proportion of disagreement increases as the factor V increases for the small value (1.5) of the factor M, whereas the proportion of disagreement decreases as the factor V increases for the large (2.5) value of the factor M. For the VD interaction, the proportion of disagreement increases as the factor V increases for the small value (0.4) of the factor D, whereas the proportion of disagreement decreases as the factor V increases for the large (0.8) value of the factor D.

It is interesting to see that Δ^2 is not only the most influential factor on the proportion of disagreement, but also the sign of $\hat{\beta}$ for the factor Δ^2 is positive. This result means that as Δ^2 increases ,the proportion of disagreement increases. One of possible explanations for this result may be the fact that as Δ^2 increases the difference in values between the LR selection criterion and the WD selection criterion increases. Table 5.2.1.5 shows that as Δ^2 increases from 1.5 to 2.5, the overall mean of the difference in values between the LR selection criterion and the WD selection criterion increases from 2.080 to 6.632. The difference is particularly large in selecting the first variable in the model; the difference is 20.784 and

66.270 for Δ^2 = 1.5 and 2.5, respectively. This implies that the WD selection criterion may have a high chance of selecting a different variable from the LR selection criterion at the first step in forward stepwise logistic regression. Table 5.2.1.6 supports that proposition. The WD selection criterion has a 27 percent of chance of selecting a different variable at the first step. Table 5.2.1.6 also shows that, given that there was an 'overall' 34 percent of chance of disagreement, once the first few variables have been successfully selected there is a good chance of getting a model with the same order of selected variables; the first few variables tend to determine the proportion of disagreement.

Table 5.2.1.4 for the SC selection criterion shows that the factors P and Δ^2 are highly significant (t = 4.23 for both) (see figures 5.2.1.1 and 5.2.1.3 for the factors P and Δ^2, respectively). The result for the factor P is expected; as the number of variables increase more variables compete for selection leading to more disagreements. The interpretation of the result for the factor Δ^2 is the same as in the case of the WD selection criterion. However, the scale of the effect for the factor Δ^2 in the SC selection criterion is much smaller than that in the WD selection criterion (see table 5.2.1.5). For instance, when Δ^2 = 1.5 the overall mean of the difference in values between the LR selection criterion and the SC selection criterion is less than half of that between the LR selection criterion and the WD selection criterion (i.e., 0.818 versus 2.080). Table 5.2.1.6 supports this finding; there are 3 and 27 percents of chance of selecting different variables at the first step for the SC and WD selection criteria, respectively.

5.2.2 Multivariate binary case

The tables and graphs for the multivariate binary case are analogous to those in the multivariate normal case.

Table 5.2.2.1 gives the proportion of disagreement for the six selection criteria LS, WD, SC, PH, LK, and SW for each of 81 sampling situations. The LS selection criterion selects exactly the same variables as the LR selection criterion for all 81 sampling situations. This result is consistent with that in the multivariate normal case.

For the WD selection criterion, the proportion of disagreement range from 0.00 (sampling situation = 1, 11, 19, 29, 37, 38, 39, 48, 58, 64, 73, 74, and 75) to 0.70 (sampling situation = 9, 26, 27, 54, and 81). The WD selection criterion performs poorly the selection of variables with an average of 23 percent chance of disagreement over all 81 sampling situations (see table 5.2.2.2).

The proportions of disagreement for the SC, PH, and LK selection criteria are the same except for one sampling situation (9). The range of the proportion of disagreement for these three selection criteria is from 0.00 (sampling situation = 16, 19, 28, 29, 37, 38, 39, 40, 48, 55, 58, 59, 61, 64, 73, 74, and 75) to 0.70 (sampling situation = 9). The number of sampling situations with 0.00 proportion of disagreement for these three selection criteria, compared for the WD selection criterion, increases from 13 to 17, and none of 81 sampling situations for these three selection criteria had 1.00 proportion of disagreement. However, these three selection criteria are not universally superior to the WD selection criterion (see, for example, sampling situation = 1, 2, 11, 15, 21, 60, and 70). These three selection criteria have an average 15 percent chance of disagreement over all 81 sampling situations (see table 5.2.2.2).

The SW selection criterion has only one sampling situation (46) with 0.00 proportion of disagreement and has twenty two sampling situations with 1.00 proportion of disagreement (6, 8, 9, 15, 18, 24, 27, 33,35, 36, 42, 44, 45,51, 54, 60, 62, 63, 69, 72, 80, and 81). The SW selection criterion has an average 66 percent chance of disagreement with the LR selection criterion (see table 5.2.2.2).

Since the proportion of disagreement for each selection selection criterion depends on the sampling situations, we examine the effects of the four factors P, B, M, and N on the proportion of disagreement.

Table 5.2.2.2 gives the proportion of disagreement for the six selection criteria LS, WD, SC, PH, LK, and SW over the levels of the four factors P, B, M, and N. The results of table 5.2.2.2 are graphically presented in figures 5.2.2.1 through 5.2.1.4 for the four factors P, B, M, and N, respectively. These figures include only the WD and SC selection criteria with the same reasons as in the multivariate normal case.

Analyses of variance were employed to assess the effects of the four factors P, B, M, and N on the proportion of disagreement in table 5.2.2.1 for the WD and SC selection criteria. Tables 5.2.2.3 and 5.2.2.4 give the results for analyses of variance for the WD and SC selection criteria, respectively. We note that the values of R^2 for tables 5.2.2.3 and 5.2.2.4 are 0.92 and 0.83, respectively, which indicate that these ANOVAs fit the data well.

The factors P and M are highly significant ($p < 0.001$) in both tables. The interpretation of the result for the factor P can be the same as in the multivariate normal case.

M is the most influential factor on the proportion of disagreement in both tables. Since the concept of the factor M is analogous to that of the factor Δ^2 in the multivariate normal case, the same results for the factor M are expected as in the multivariate normal case. Tables 5.2.2.5 and 5.2.2.6 confirm that proposition.

Three interaction terms, PB, PM, and BM, for the WD selection criterion are statistically significant. The graphs for the PB, PM, and BM interactions are presented in figures 5.2.2.5 through 5.2.2.7, respectively, to depict the behaviors of these interactions on the proportion of disagreement. Only the PM interaction for the SC selection criterion is statistically significant. The graph for the PM interaction is presented in figure 5.2.2.8. All of these graphs show that these interactions are merely nonparallel; this means that the direction of the effect of a factor on the the proportion of disagreement is the same for all levels of the other factor.

Finally, as in the multivariate normal case, it is noticed in table 5.2.2.5 that the values of the LS selection criterion are always larger than those of the LR selection criterion, whereas the values of the WD and SC selection criteria tend to be smaller than those of the LR selection criterion.

5.2.3 Conclusions

1. The LS selection criterion always select variables in the same orders as the LR selection criterion in forward stepwise logistic regression.
2. The SC, PH, and LK selection criteria have the equal proportions of disagreement in most sampling situations.
3. Δ^2 and M are the most important factors influencing the proportions of disagreement in the multivariate normal and multivariate binary cases, respectively.

Table 5.2.1.1

The proportion of disagreement for the six selection criteria LS, WD, SC, PH, LK, and SW for each sampling situation

Sampling situation	Number of replications	Selection criteria					
		LS	WD	SC	PH	LK	SW
1	20	0.00	0.05	0.05	0.05	0.05	0.10
2	20	0.00	0.05	0.10	0.10	0.10	0.55
3	20	0.00	0.00	0.00	0.00	0.00	0.05
4	20	0.00	0.10	0.10	0.10	0.10	0.45
5	20	0.00	0.55	0.00	0.00	0.00	0.20
6	20	0.00	0.70	0.55	0.55	0.55	0.90
7	20	0.00	0.55	0.10	0.10	0.10	0.20
8	20	0.00	0.75	0.40	0.40	0.40	0.85
9	20	0.00	0.05	0.00	0.00	0.00	0.10
10	20	0.00	0.00	0.05	0.05	0.05	0.65
11	20	0.00	0.25	0.15	0.15	0.15	0.40
12	20	0.00	0.20	0.15	0.15	0.15	0.70
13	20	0.00	1.00	0.25	0.25	0.25	0.40
14	20	0.00	0.90	0.15	0.15	0.15	0.80
15	20	0.00	0.25	0.15	0.15	0.15	0.50
16	20	0.00	0.65	0.40	0.35	0.40	0.95
17	20	0.00	0.00	0.00	0.00	0.00	0.00
18	20	0.00	0.00	0.05	0.05	0.05	0.50
19	20	0.00	0.05	0.05	0.05	0.05	0.10
20	20	0.00	0.05	0.00	0.00	0.00	0.65
21	20	0.00	0.45	0.10	0.10	0.10	0.05
22	20	0.00	0.20	0.15	0.15	0.15	0.70
23	20	0.00	0.55	0.10	0.10	0.10	0.25
24	20	0.00	0.45	0.25	0.25	0.25	0.60

Table 5.2.1.1 (continued)

The proportion of disagreement for the six selection criteria LS, WD, SC, PH, LK, and SW for each sampling situation

Sampling situation	Number of replications	Selection criteria					
		LS	WD	SC	PH	LK	SW
25	20	0.00	0.05	0.00	0.00	0.00	0.20
26	20	0.00	0.40	0.30	0.30	0.30	0.55
27	20	0.00	0.10	0.05	0.05	0.05	0.20
28	20	0.00	0.10	0.05	0.05	0.05	0.50
29	20	0.00	0.95	0.00	0.00	0.00	0.10
30	20	0.00	0.95	0.15	0.15	0.15	0.70
31	20	0.00	0.40	0.25	0.25	0.25	0.55
32	20	0.00	0.55	0.25	0.25	0.25	0.95
33	20	0.00	0.40	0.15	0.15	0.15	0.60
34	20	0.00	0.25	0.00	0.00	0.00	0.20
35	20	0.00	0.05	0.05	0.05	0.05	0.25
36	20	0.00	0.10	0.05	0.05	0.05	0.30
37	20	0.00	0.05	0.10	0.10	0.10	0.50
38	20	0.00	1.00	0.25	0.30	0.30	0.75
39	20	0.00	0.20	0.15	0.15	0.15	0.20
40	20	0.00	0.15	0.10	0.10	0.10	0.50
41	20	0.00	0.15	0.00	0.00	0.00	0.00
42	20	0.00	0.35	0.30	0.30	0.30	0.90
43	20	0.00	0.30	0.05	0.05	0.05	0.25
44	20	0.00	0.45	0.10	0.10	0.10	0.30
45	20	0.00	0.35	0.20	0.20	0.20	0.35
46	20	0.00	0.20	0.10	0.10	0.10	0.35
47	20	0.00	0.40	0.10	0.10	0.10	0.45
48	20	0.00	0.45	0.15	0.15	0.15	0.40

Table 5.2.1.2
The proportion of disagreement for the six selection criteria over the levels of the five factors P, V, Δ^2, D, and N

Factor	[a]Level	Value	Selection criteria					
			LS	WD	SC	PH	LK	SW
P	-2	5	0.00	0.15	0.00	0.00	0.00	0.00
	-1	10	0.00	0.33	0.08	0.08	0.08	0.21
	0	15	0.00	0.31	0.11	0.11	0.11	0.39
	+1	20	0.00	0.38	0.19	0.19	0.19	0.69
	+2	25	0.00	0.35	0.30	0.30	0.30	0.90
V	-2	0.2	0.00	0.20	0.15	0.15	0.15	0.20
	-1	0.4	0.00	0.39	0.12	0.12	0.12	0.41
	0	0.6	0.00	0.32	0.11	0.12	0.12	0.40
	+1	0.8	0.00	0.31	0.15	0.15	0.15	0.49
	+2	1.0	0.00	0.15	0.10	0.10	0.10	0.50
Δ^2	-2	1.0	0.00	0.05	0.10	0.10	0.10	0.50
	-1	1.5	0.00	0.09	0.07	0.07	0.07	0.36
	0	2.0	0.00	0.27	0.11	0.11	0.11	0.36
	+1	2.5	0.00	0.62	0.20	0.20	0.20	0.54
	+2	3.0	0.00	1.00	0.25	0.30	0.30	0.75
D	-2	0.2	0.00	0.05	0.05	0.05	0.05	0.25
	-1	0.4	0.00	0.28	0.13	0.13	0.13	0.38
	0	0.6	0.00	0.34	0.13	0.13	0.13	0.41
	+1	0.8	0.00	0.43	0.15	0.14	0.15	0.52
	+2	1.0	0.00	0.10	0.05	0.05	0.05	0.30
N	-2	100	0.00	0.40	0.15	0.15	0.15	0.60
	-1	150	0.00	0.38	0.16	0.16	0.16	0.49
	0	200	0.00	0.30	0.12	0.13	0.13	0.39
	+1	250	0.00	0.33	0.11	0.11	0.11	0.41
	+2	300	0.00	0.25	0.00	0.00	0.00	0.20
Overall			0.00	0.34	0.13	0.13	0.13	0.43

[a]The number of sampling situations for each level is 1 for levels ±2, 16 for levels ±1, and 14 for level 0.

Table 5.2.1.3

Response surface analysis of the proportion of disagreement for the WD selection criterion

Factor	d.f	$\hat{\beta}$	s.e	t-value	p-value
P	1	0.030	0.023	1.29	0.21
V	1	-0.035	0.023	-1.51	0.14
[a]M	1	0.258	0.023	11.10	0.00
[b]D	1	0.060	0.023	2.59	0.02
N	1	-0.028	0.023	-1.19	0.25
P^2	1	-0.005	0.027	-0.19	0.85
V^2	1	-0.024	0.027	-0.89	0.38
[b]M^2	1	0.064	0.027	2.40	0.02
D^2	1	-0.049	0.027	-1.83	0.08
N^2	1	0.014	0.027	0.52	0.61
PV	1	0.019	0.026	0.72	0.48
PM	1	0.003	0.026	0.12	0.91
PD	1	0.019	0.026	0.72	0.48
PN	1	-0.016	0.026	-0.60	0.55
[b]VM	1	-0.056	0.026	-2.17	0.04
[b]VD	1	-0.072	0.026	-2.77	0.01
VN	1	-0.006	0.026	-0.24	0.81
MD	1	0.019	0.026	0.72	0.48
MN	1	-0.028	0.026	-1.08	0.29
DN	1	0.038	0.026	1.45	0.16

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

Lack-of-fit test: $F_{22,5} = 2.578$ ($p > 0.148$)

Table 5.2.1.4

Response surface analysis of the proportion of disagreement for the SC selection criterion

Factor	d.f	$\hat{\beta}$	s.e	t-value	p-value
[a]P	1	0.061	0.014	4.23	0.00
V	1	0.011	0.014	0.78	0.44
[a]M	1	0.061	0.014	4.23	0.00
D	1	0.009	0.014	0.60	0.55
N	1	-0.029	0.014	-1.99	0.06
P^2	1	0.013	0.017	0.77	0.45
V^2	1	0.007	0.017	0.40	0.70
M^2	1	0.019	0.017	1.15	0.26
D^2	1	-0.012	0.017	-0.73	0.47
N^2	1	-0.006	0.017	-0.36	0.72
PV	1	-0.011	0.016	-0.68	0.51
PM	1	0.027	0.016	1.64	0.11
PD	1	-0.017	0.016	-1.06	0.30
PN	1	-0.017	0.016	-1.06	0.30
VM	1	0.017	0.016	1.06	0.30
VD	1	0.017	0.016	1.06	0.30
VN	1	-0.002	0.016	-0.10	0.92
MD	1	-0.014	0.016	-0.87	0.39
MN	1	-0.020	0.016	-1.25	0.22
DN	1	0.011	0.016	0.68	0.51

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

Lack-of-fit test: $F_{22,5} = 3.635$ ($p > 0.078$)

Table 5.2.1.5
Comparison of values of the LS, WD, and SC selection criteria with those of the LR selection criterion for Δ^2 = 1.5 and 2.5

Step of variable	$^{a}\Delta^2$	Selection criteria			
		LR	LS[b]	WD[b]	SC[b]
$X_{(1)}$	1.5	54.795	-0.023	20.784	8.173
	2.5	111.045	-0.017	66.270	27.231
$X_{(2)}$	1.5	0.834	-0.002	0.003	-0.001
	2.5	1.632	-0.020	0.017	-0.007
$X_{(3)}$	1.5	0.814	-0.002	0.006	0.003
	2.5	1.584	-0.013	0.029	0.001
$X_{(4)}$	1.5	0.441	0.000	0.005	0.004
	2.5	0.230	-0.001	0.000	0.000
$X_{(5)}$	1.5	0.157	0.000	0.000	0.000
	2.5	0.385	-0.001	0.001	0.001
$X_{(6)}$	1.5	0.138	0.000	0.000	0.000
	2.5	0.061	0.000	0.000	0.000
$X_{(7)}$	1.5	0.246	0.000	-0.001	-0.001
	2.5	0.050	0.000	0.000	0.000
$X_{(8)}$	1.5	0.043	0.000	0.000	0.000
	2.5	0.028	0.000	-0.001	0.000
$X_{(9)}$	1.5	0.008	0.000	0.000	0.000
	2.5	0.015	0.000	0.000	0.000
$X_{(10)}$	1.5	0.001	0.000	0.000	0.000
	2.5	0.008	0.000	0.000	0.000
Overall mean	1.5	5.748	0.003	2.080	0.818
	2.5	11.504	0.005	6.632	2.724

[a]Δ^2 = 1.5 and 2.5 come from the sampling situations=1 and 5, respectively.
[b]Numbers are the differences from the values of the LR selection criterion.

Table 5.2.1.6

Cumulative proportion of disagreement up to the fifth variable over 48 sampling situations for the six selection criteria

Order of variables	Selection criteria					
	LS	WD	SC	PH	LK	SW
1	0.00	0.27	0.03	0.03	0.03	0.03
2	0.00	0.29	0.05	0.05	0.05	0.14
3	0.00	0.29	0.05	0.05	0.05	0.21
4	0.00	0.30	0.06	0.06	0.06	0.26
5	0.00	0.31	0.07	0.07	0.07	0.31

Figure 5.2.1.1

The proportion of disagreement for the factor P

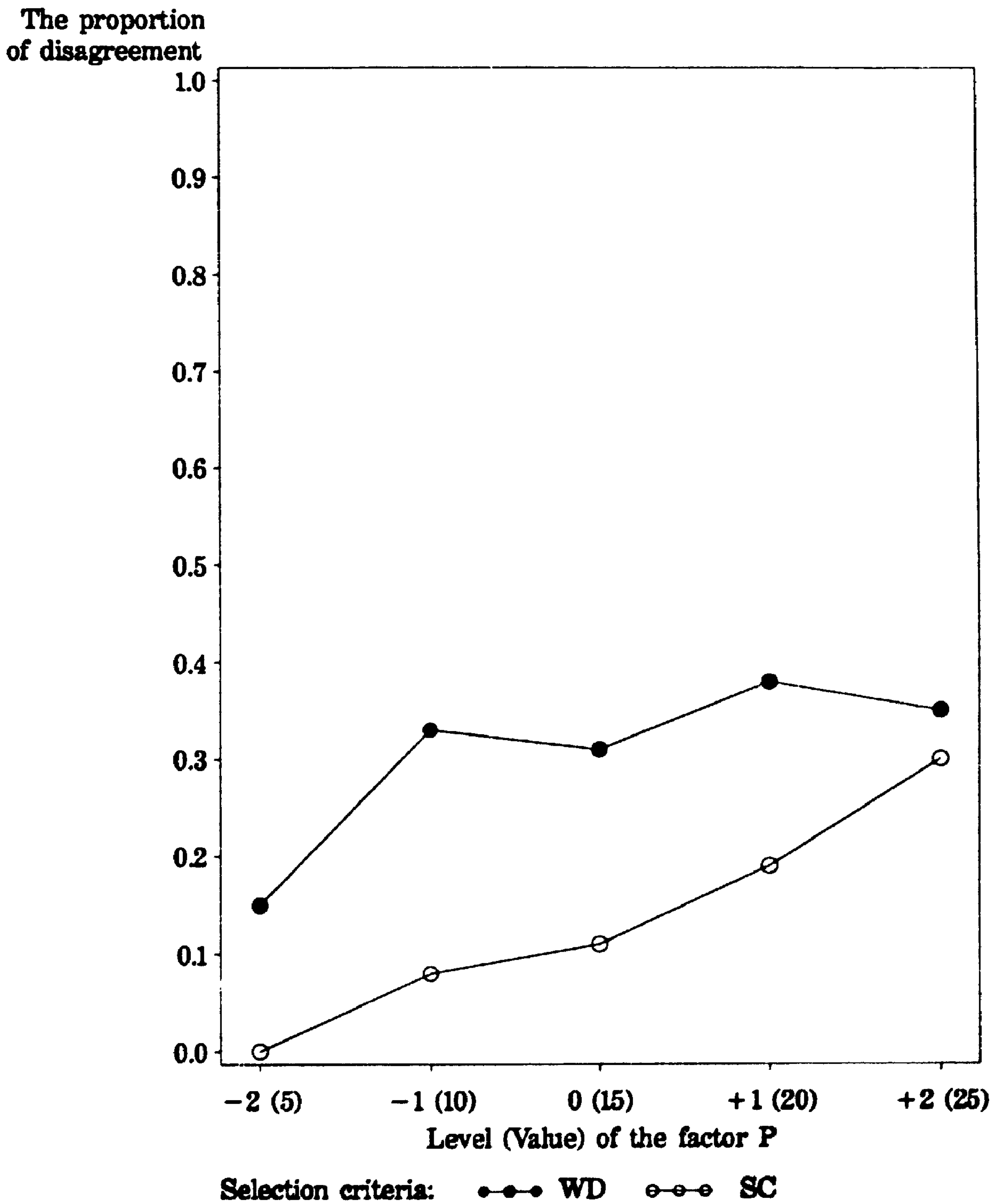


Figure 5.2.1.2

The proportion of disagreement for the factor V

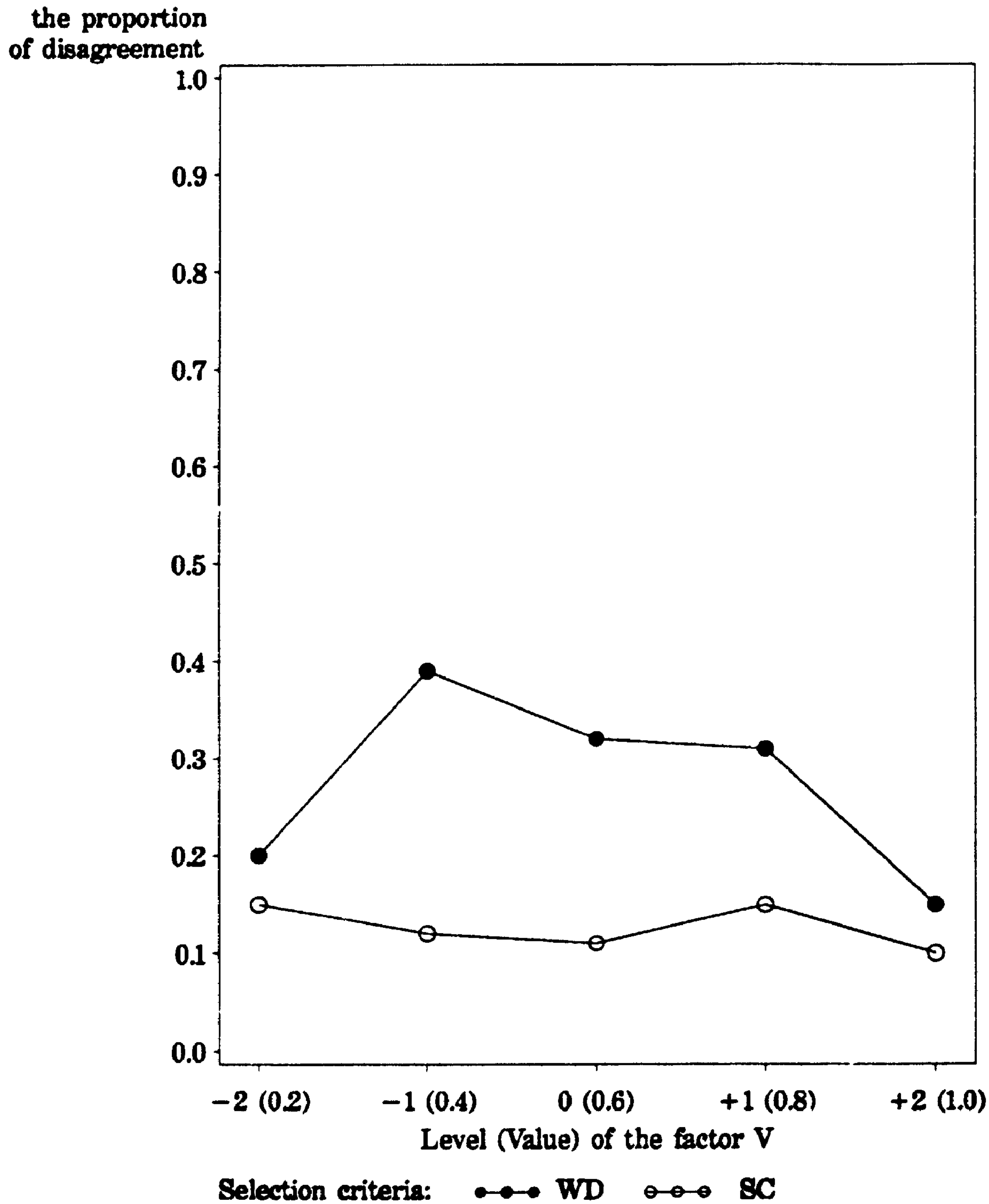


Figure 5.2.1.3

The proportion of disagreement for the factor M

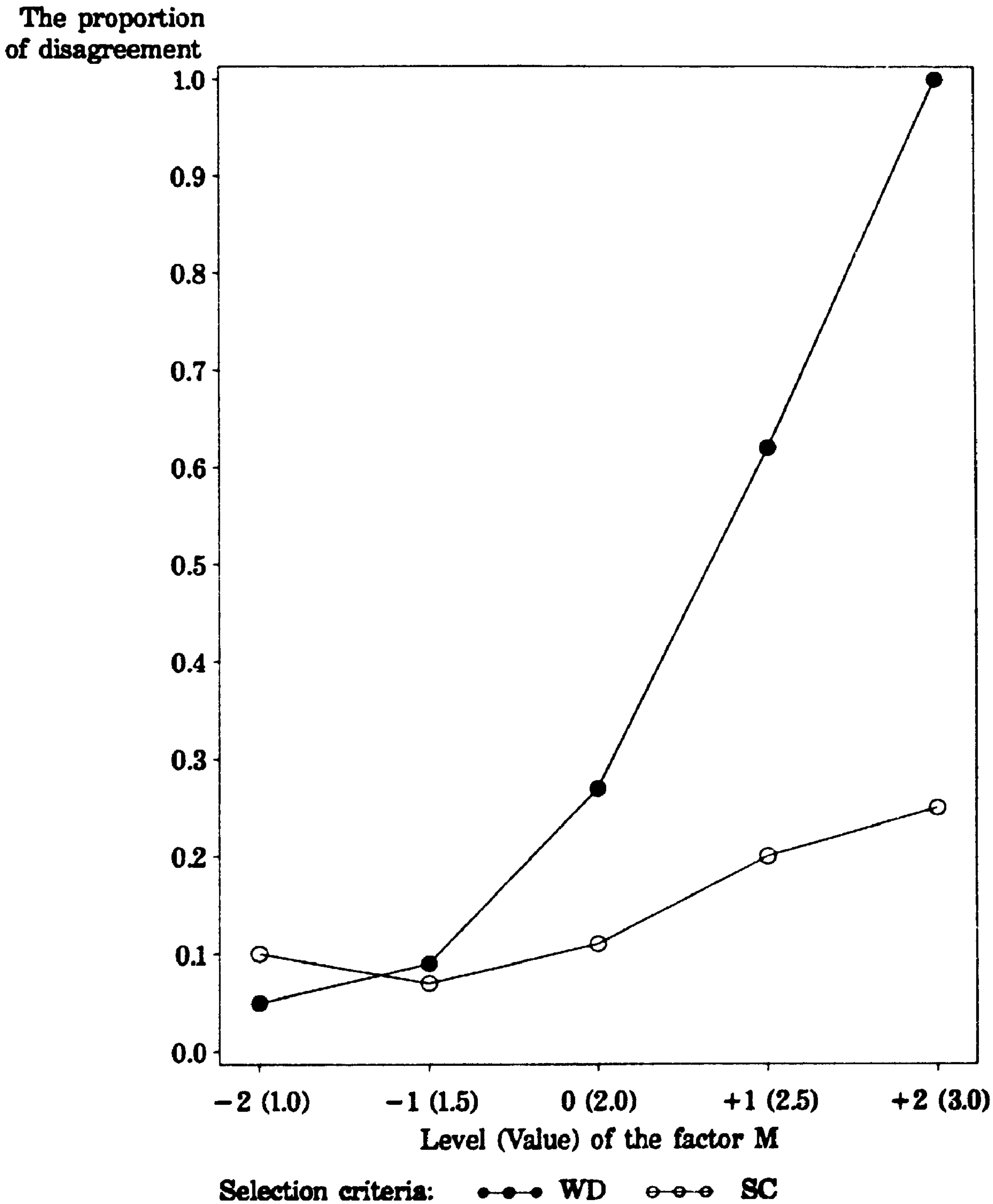


Figure 5.2.1.4

The proportion of disagreement for the factor D

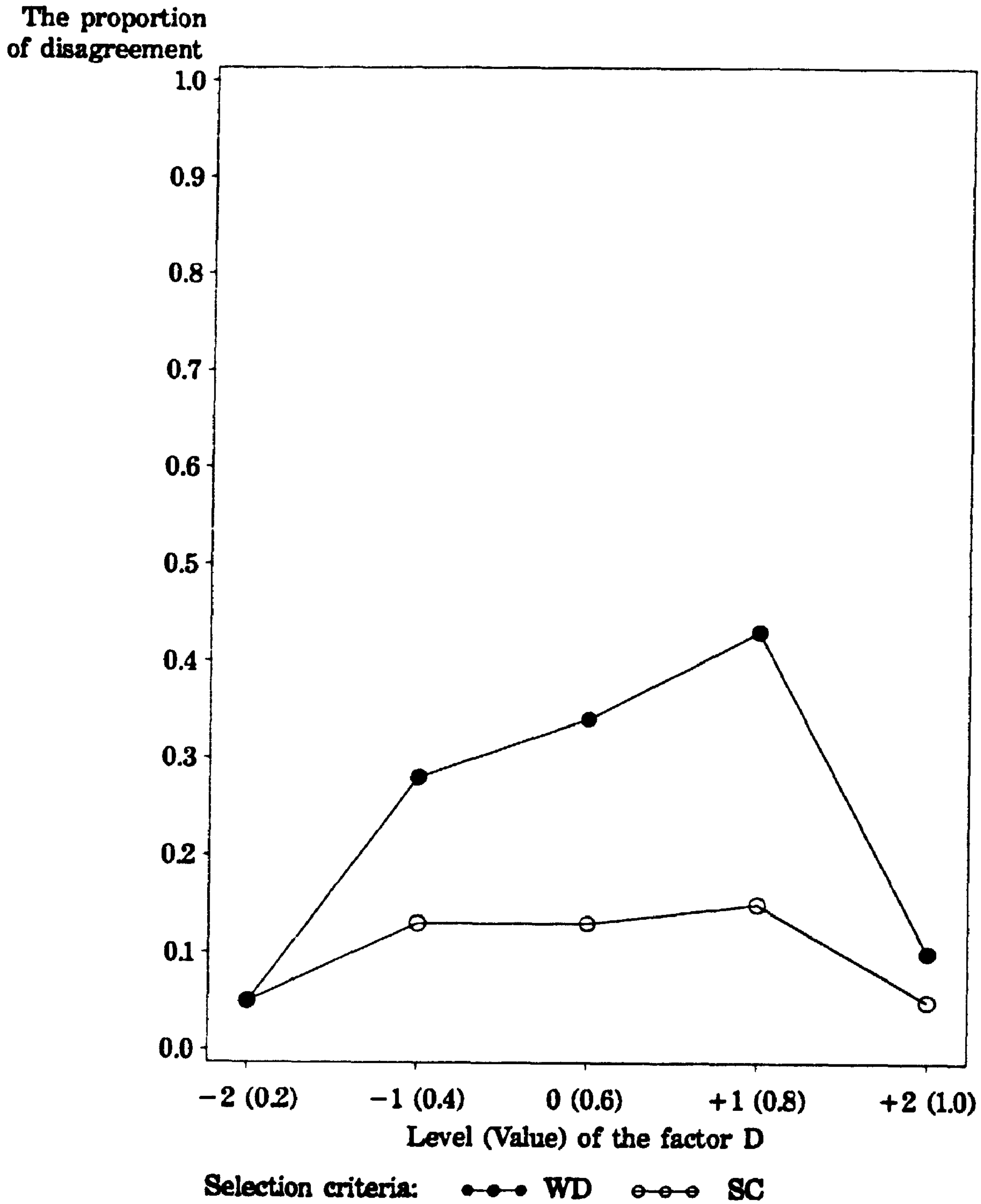


Figure 5.2.1.5

The proportion of disagreement for the factor N

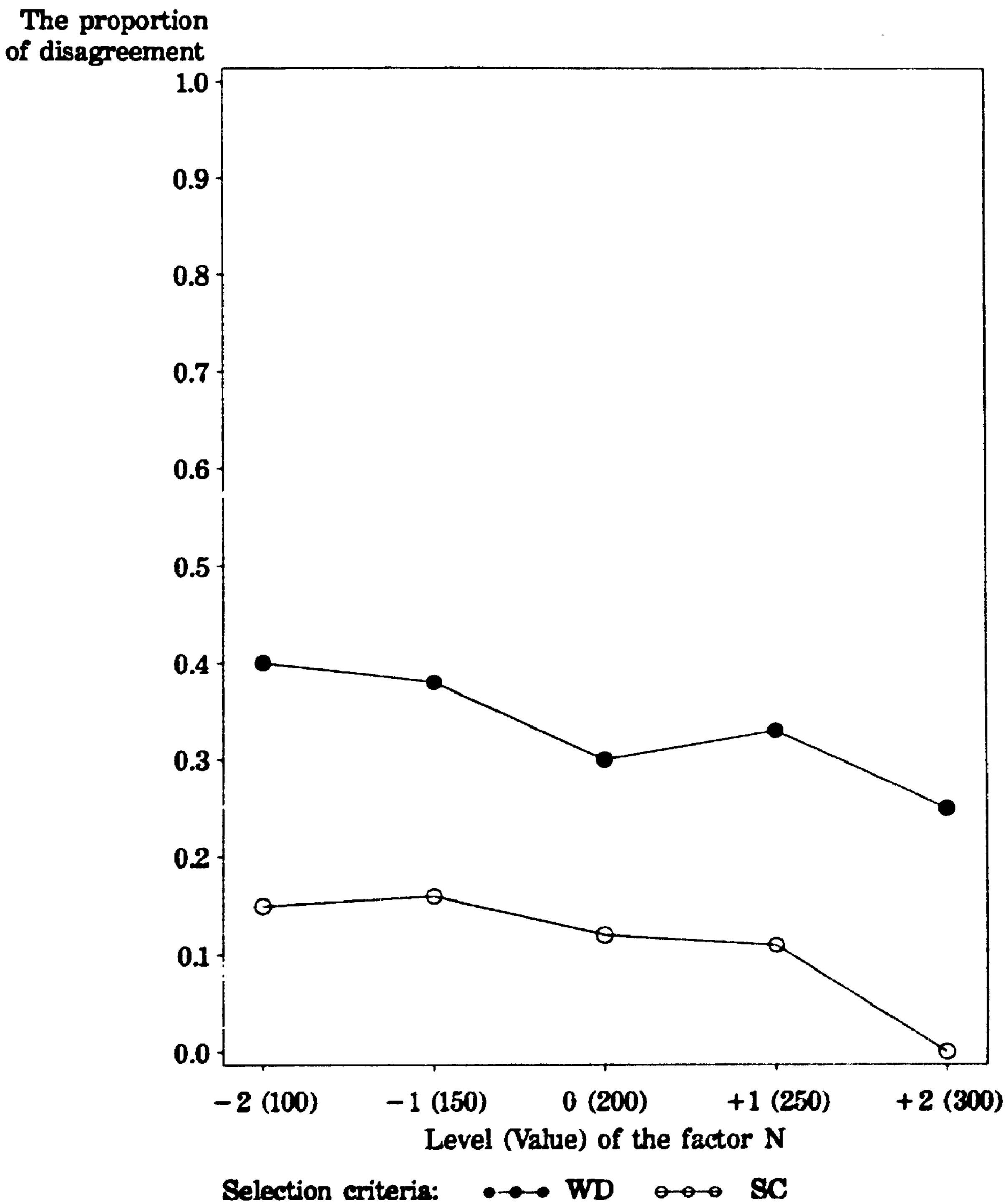


Figure 5.2.1.6

Effect of VM interaction on the proportion of disagreement at the factorial levels for the WD selection criterion

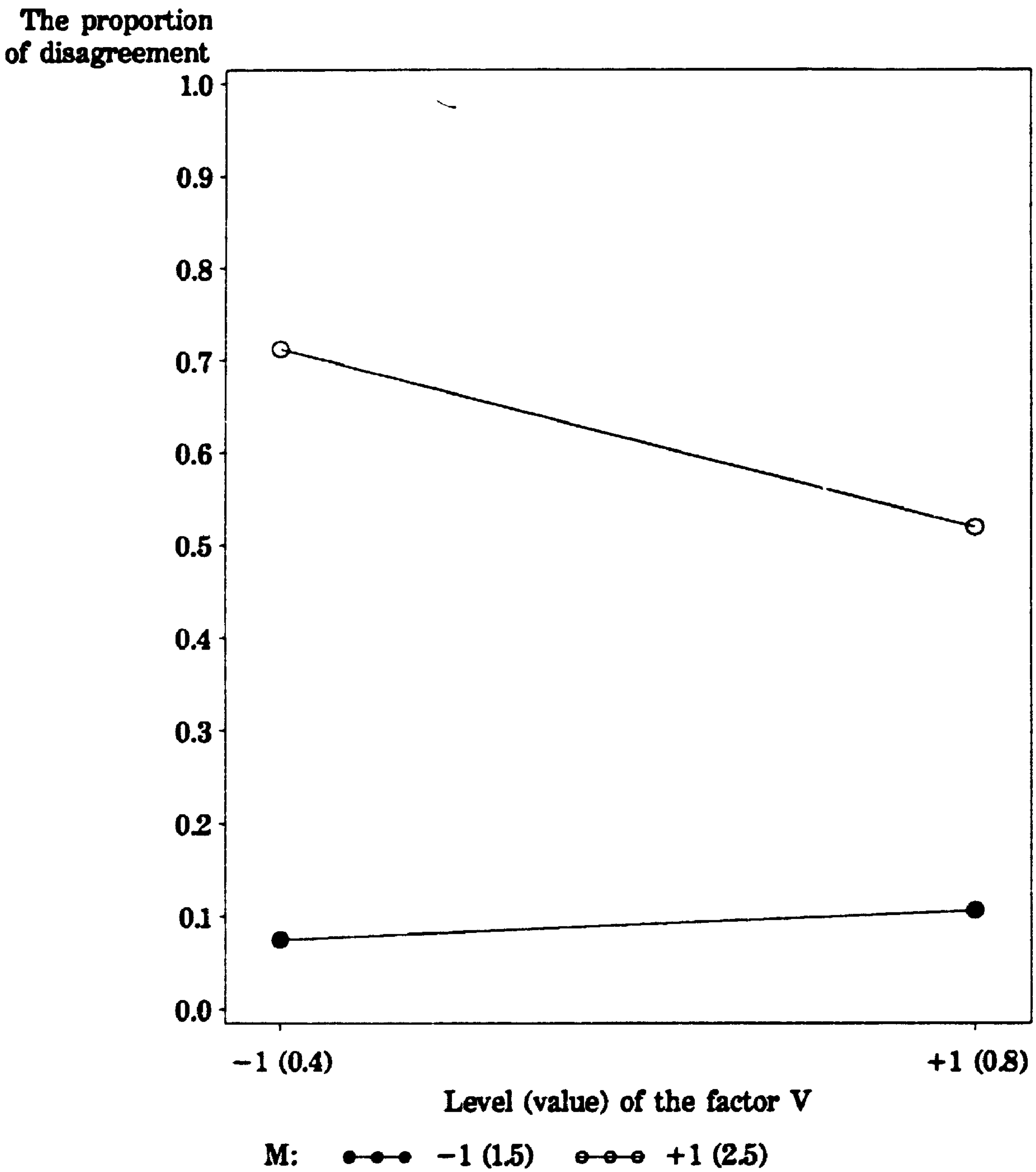


Figure 5.2.1.7

Effect of VD interaction on the proportion of disagreement at the factorial levels for the WD selection criterion

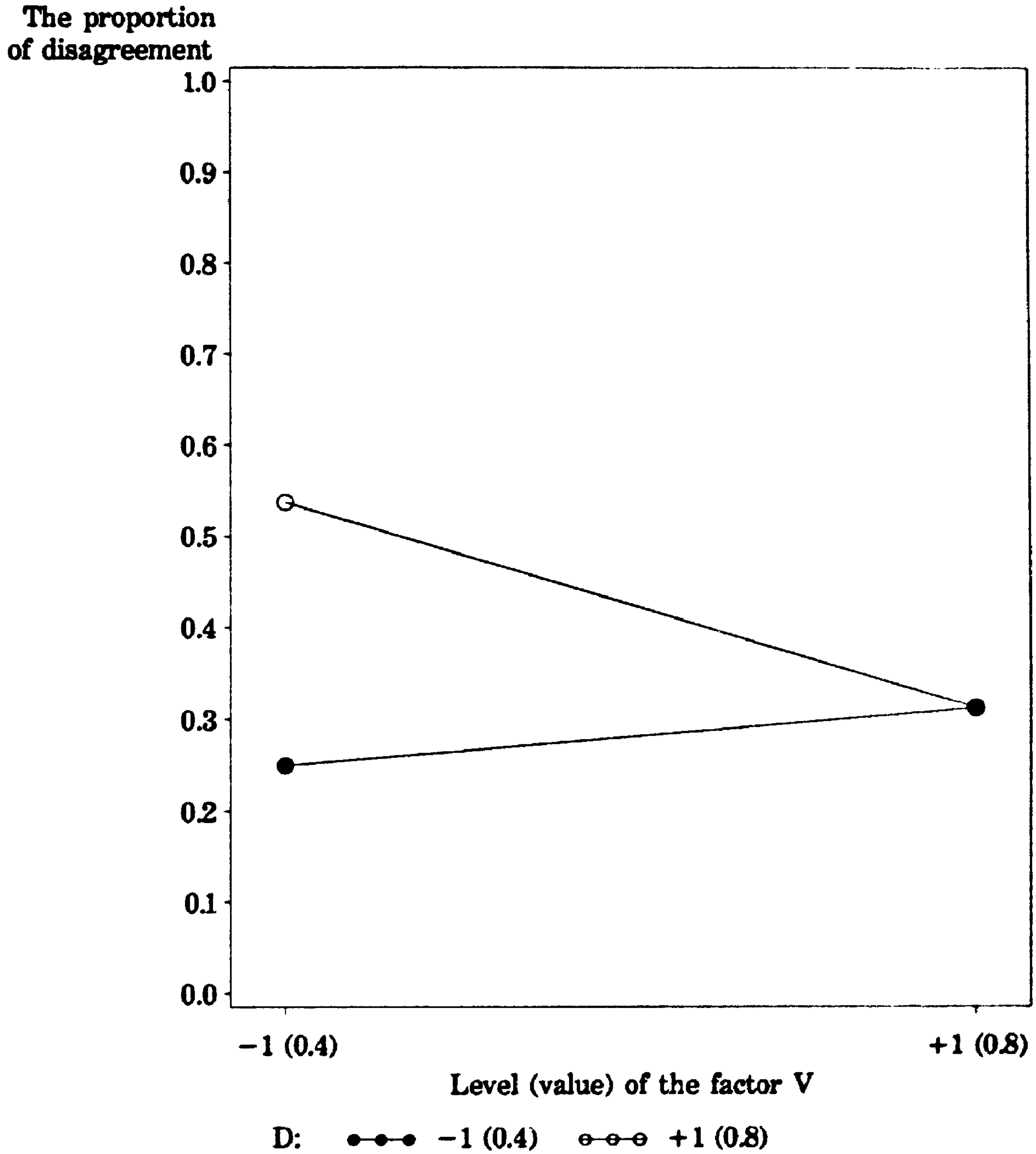


Table 5.2.2.1
The proportion of disagreement for the six selection criteria LS, WD, SC, PH, LK, and SW for each sampling situation

Sampling situation	Number of replications	Selection criteria					
		LS	WD	SC	PH	LK	SW
1	20	0.00	0.00	0.05	0.05	0.05	0.10
2	20	0.00	0.05	0.15	0.15	0.15	0.50
3	20	0.00	0.15	0.15	0.15	0.15	0.85
4	20	0.00	0.20	0.10	0.10	0.10	0.45
5	20	0.00	0.35	0.15	0.15	0.15	0.95
6	20	0.00	0.40	0.20	0.20	0.20	1.00
7	20	0.00	0.25	0.20	0.20	0.20	0.80
8	20	0.00	0.35	0.20	0.20	0.20	1.00
9	20	0.00	0.70	0.70	0.70	0.65	1.00
10	20	0.00	0.05	0.05	0.05	0.05	0.20
11	20	0.00	0.00	0.05	0.05	0.05	0.35
12	20	0.00	0.15	0.10	0.10	0.10	0.70
13	20	0.00	0.10	0.05	0.05	0.05	0.30
14	20	0.00	0.10	0.10	0.10	0.10	0.80
15	20	0.00	0.20	0.25	0.25	0.25	1.00
16	20	0.00	0.10	0.00	0.00	0.00	0.40
17	20	0.00	0.30	0.25	0.25	0.25	0.95
18	20	0.00	0.55	0.45	0.45	0.45	1.00
19	20	0.00	0.00	0.00	0.00	0.00	0.15
20	20	0.00	0.25	0.05	0.05	0.05	0.50
21	20	0.00	0.10	0.15	0.15	0.15	0.80
22	20	0.00	0.25	0.15	0.15	0.15	0.45
23	20	0.00	0.20	0.15	0.15	0.15	0.80
24	20	0.00	0.30	0.25	0.25	0.25	1.00
25	20	0.00	0.55	0.15	0.15	0.15	0.65
26	20	0.00	0.70	0.20	0.20	0.20	0.85
27	20	0.00	0.70	0.45	0.45	0.45	1.00

Table 5.2.2.1 (continued)
The proportion of disagreement for the six selection criteria
LS, WD, SC, PH, LK, and SW for each sampling situation

Sampling situation	Number of replications	Selection criteria					
		LS	WD	SC	PH	LK	SW
28	20	0.00	0.05	0.00	0.00	0.00	0.15
29	20	0.00	0.00	0.00	0.00	0.00	0.35
30	20	0.00	0.10	0.05	0.05	0.05	0.95
31	20	0.00	0.15	0.10	0.10	0.10	0.30
32	20	0.00	0.10	0.05	0.05	0.05	0.80
33	20	0.00	0.45	0.35	0.35	0.35	1.00
34	20	0.00	0.35	0.20	0.20	0.20	0.70
35	20	0.00	0.45	0.30	0.30	0.30	1.00
36	20	0.00	0.55	0.35	0.35	0.35	1.00
37	20	0.00	0.00	0.00	0.00	0.00	0.05
38	20	0.00	0.00	0.00	0.00	0.00	0.40
39	20	0.00	0.00	0.00	0.00	0.00	0.75
40	20	0.00	0.05	0.00	0.00	0.00	0.35
41	20	0.00	0.20	0.05	0.05	0.05	0.70
42	20	0.00	0.25	0.25	0.25	0.25	1.00
43	20	0.00	0.15	0.05	0.05	0.05	0.65
44	20	0.00	0.45	0.30	0.30	0.30	1.00
45	20	0.00	0.40	0.30	0.30	0.30	1.00
46	20	0.00	0.05	0.05	0.05	0.05	0.00
47	20	0.00	0.10	0.10	0.10	0.10	0.30
48	20	0.00	0.00	0.00	0.00	0.00	0.60
49	20	0.00	0.15	0.05	0.05	0.05	0.30
50	20	0.00	0.10	0.05	0.05	0.05	0.70
51	20	0.00	0.05	0.05	0.05	0.05	1.00
52	20	0.00	0.60	0.25	0.25	0.25	0.70
53	20	0.00	0.65	0.30	0.30	0.30	0.85
54	20	0.00	0.70	0.25	0.25	0.25	1.00

Table 5.2.2.1 (continued) The proportion of disagreement for the six selection criteria LS, WD, SC, PH, LK, and SW for each sampling situation							
Sampling situation	Number of replications	Selection criteria					
		LS	WD	SC	PH	LK	SW
55	20	0.00	0.05	0.00	0.00	0.00	0.15
56	20	0.00	0.10	0.05	0.05	0.05	0.35
57	20	0.00	0.05	0.05	0.05	0.05	0.75
58	20	0.00	0.00	0.00	0.00	0.00	0.50
59	20	0.00	0.15	0.00	0.00	0.00	0.85
60	20	0.00	0.25	0.30	0.30	0.30	1.00
61	20	0.00	0.15	0.00	0.00	0.00	0.65
62	20	0.00	0.35	0.30	0.30	0.30	1.00
63	20	0.00	0.50	0.35	0.35	0.35	1.00
64	20	0.00	0.00	0.00	0.00	0.00	0.10
65	20	0.00	0.15	0.15	0.15	0.15	0.45
66	20	0.00	0.10	0.10	0.10	0.10	0.55
67	20	0.00	0.15	0.15	0.15	0.15	0.25
68	20	0.00	0.10	0.05	0.05	0.05	0.80
69	20	0.00	0.25	0.20	0.20	0.20	1.00
70	20	0.00	0.15	0.20	0.20	0.20	0.65
71	20	0.00	0.35	0.35	0.35	0.35	0.90
72	20	0.00	0.60	0.55	0.55	0.55	1.00
73	20	0.00	0.00	0.00	0.00	0.00	0.20
74	20	0.00	0.00	0.00	0.00	0.00	0.40
75	20	0.00	0.00	0.00	0.00	0.00	0.50
76	20	0.00	0.10	0.05	0.05	0.05	0.35
77	20	0.00	0.25	0.25	0.25	0.25	0.55
78	20	0.00	0.20	0.15	0.15	0.15	0.85
79	20	0.00	0.45	0.15	0.15	0.15	0.45
80	20	0.00	0.35	0.10	0.10	0.10	1.00
81	20	0.00	0.70	0.40	0.40	0.40	1.00

Table 5.2.2.2
The proportion of disagreement for the six selection criteria over the levels of the four factors P, B, M, and N

Factor	[a]Level	Value	Selection criteria					
			LS	WD	SC	PH	LK	SW
P	-1	10	0.00	0.15	0.07	0.07	0.07	0.37
	0	15	0.00	0.23	0.14	0.14	0.14	0.71
	+1	20	0.00	0.31	0.24	0.24	0.24	0.90
B	-1	0.2	0.00	0.23	0.16	0.16	0.16	0.71
	0	0.4	0.00	0.18	0.15	0.15	0.15	0.64
	+1	0.6	0.00	0.28	0.14	0.14	0.14	0.63
M	-1	0.1	0.00	0.06	0.05	0.05	0.05	0.41
	0	0.2	0.00	0.19	0.13	0.13	0.13	0.71
	+1	0.3	0.00	0.45	0.27	0.27	0.27	0.86
N	-1	150	0.00	0.26	0.18	0.18	0.17	0.69
	0	200	0.00	0.23	0.13	0.13	0.13	0.65
	+1	250	0.00	0.20	0.14	0.14	0.14	0.64
Overall			0.00	0.23	0.15	0.15	0.15	0.66

[a]The number of sampling situations for each level is 27.

Table 5.2.2.3
Analysis of variance of the proportion of disagreement for the WD selection criterion

Effect	d.f.	Mean Square	F-ratio	p-value
[a]P	2	0.1713	29.39	0.0001
[a]B	2	0.0626	10.74	0.0001
[a]M	2	1.0782	184.95	0.0001
[b]N	2	0.0226	3.88	0.0274
[b]PB	4	0.0165	2.83	0.0345
[a]PM	4	0.0367	6.29	0.0004
PN	4	0.0051	0.88	0.4836
[a]BM	4	0.0578	9.92	0.0001
MN	4	0.0068	1.17	0.3347
Error	48	0.0058		

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

$R^2 = 0.92$

Table 5.2.2.4
Analysis of variance of the proportion of disagreement for the SC selection criterion

Effect	d.f.	Mean Square	F-ratio	p-value
[a]P	2	0.1823	30.59	0.0001
B	2	0.0034	0.56	0.5724
[a]M	2	0.3412	57.25	0.0001
N	2	0.0161	2.71	0.0768
PB	4	0.0093	1.56	0.2006
[a]PM	4	0.0363	6.09	0.0005
PN	4	0.0097	1.63	0.1824
BM	4	0.0007	0.11	0.9770
BN	4	0.0148	2.48	0.0567
MN	4	0.0005	0.09	0.9850
Error	48	0.0060		

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

$R^2 = 0.83$

Table 5.2.2.5
Comparison of values of the LS, WD, and SC selection criteria with those of the LR selection criterion for M = 0.1 and 0.3

Step of variable	M	Selection criteria			
		LR	LS[b]	WD[b]	SC[b]
$X_{(1)}$	0.1	10.984	-0.029	0.841	0.270
	0.3	23.538	-0.125	2.676	0.853
$X_{(2)}$	0.1	3.419	-0.005	0.081	0.023
	0.3	19.492	-0.402	1.994	0.432
$X_{(3)}$	0.1	2.634	-0.004	0.040	0.004
	0.3	18.537	-0.728	2.406	0.470
$X_{(4)}$	0.1	3.286	-0.009	0.070	0.008
	0.3	17.890	-0.722	2.546	0.682
$X_{(5)}$	0.1	2.390	-0.009	0.066	0.012
	0.3	8.128	-0.262	0.653	0.140
$X_{(6)}$	0.1	1.738	-0.004	0.021	-0.003
	0.3	5.738	-0.117	0.362	0.081
$X_{(7)}$	0.1	0.123	0.000	0.000	-0.001
	0.3	4.267	-0.072	0.141	-0.030
$X_{(8)}$	0.1	0.099	0.000	0.000	-0.001
	0.3	3.387	-0.046	0.114	0.008
$X_{(9)}$	0.1	0.039	0.000	0.000	0.000
	0.3	0.025	0.000	0.000	0.000
$X_{(10)}$	0.1	0.034	0.000	0.000	0.000
	0.3	0.016	0.000	0.000	0.000
Overall mean	0.1	2.475	0.006	0.112	0.032
	0.3	10.102	0.247	1.089	0.270

[a]M =0.1 and 0.3 come from the sampling situations = 1 and 7, respectively.
[b]Numbers are the differences from the values of the LR selection criterion.

Table 5.2.2.6
Cumulative proportion of disagreement up to the fifth variable over 81 sampling situations for the six selection criteria

Order of variables	Selection criteria					
	LS	WD	SC	PH	LK	SW
1	0.00	0.06	0.01	0.01	0.01	0.32
2	0.00	0.10	0.03	0.03	0.03	0.45
3	0.00	0.12	0.04	0.04	0.04	0.52
4	0.00	0.14	0.06	0.06	0.06	0.57
5	0.00	0.16	0.08	0.08	0.08	0.60

Figure 5.2.2.1

The proportion of disagreement for the factor P

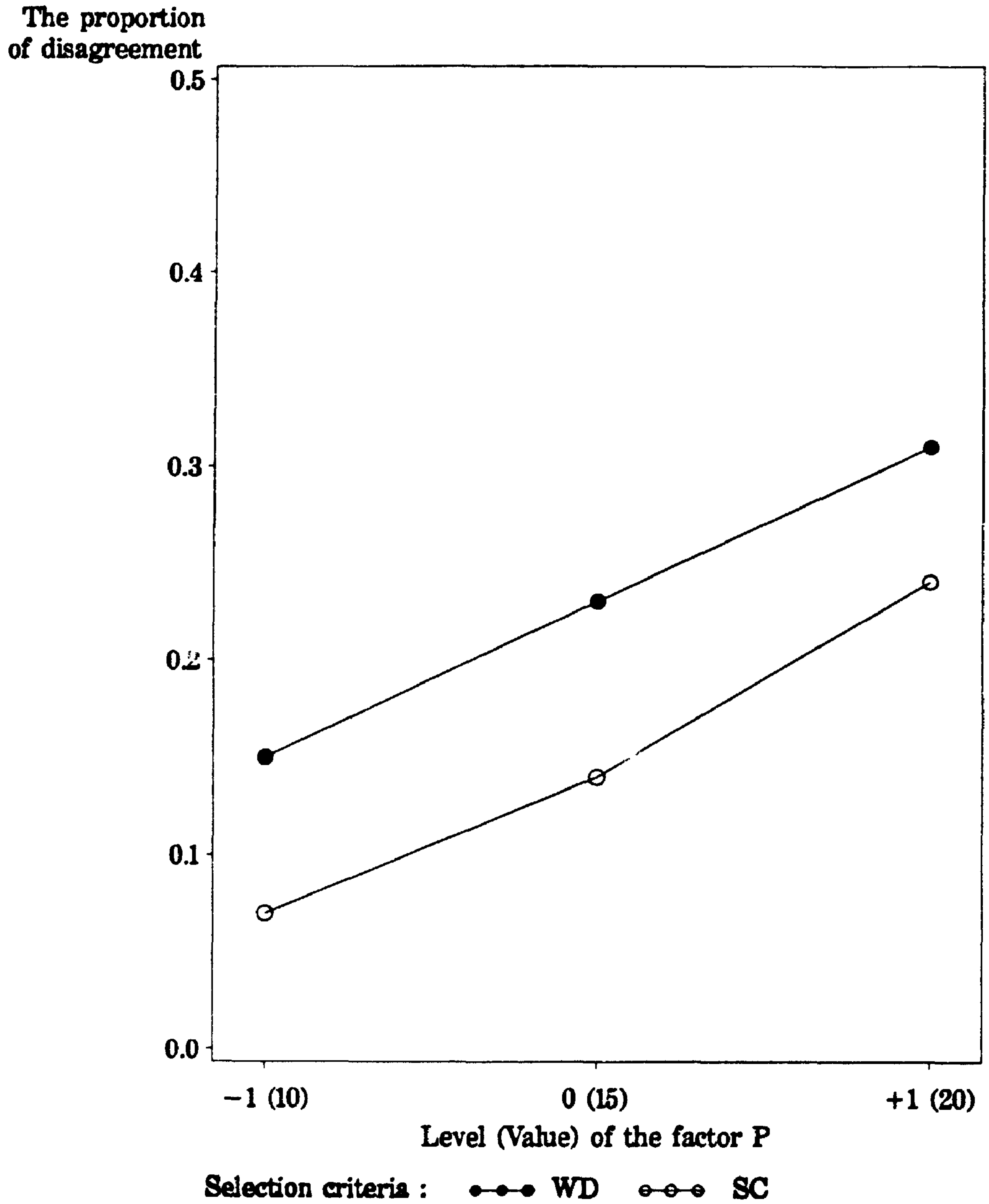


Figure 5.2.2.2

The proportion of disagreement for the factor B

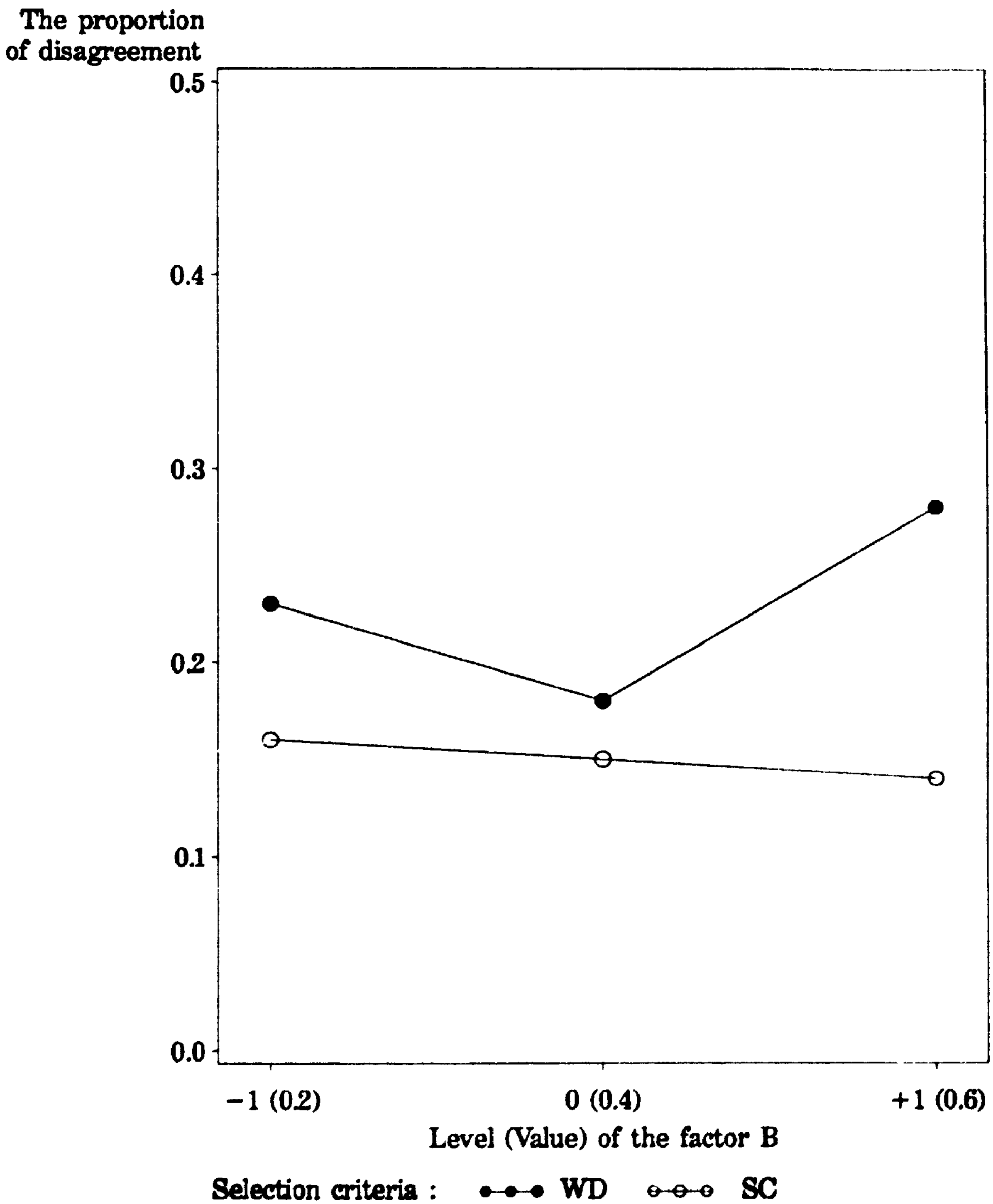


Figure 5.2.2.3

The proportion of disagreement for the M

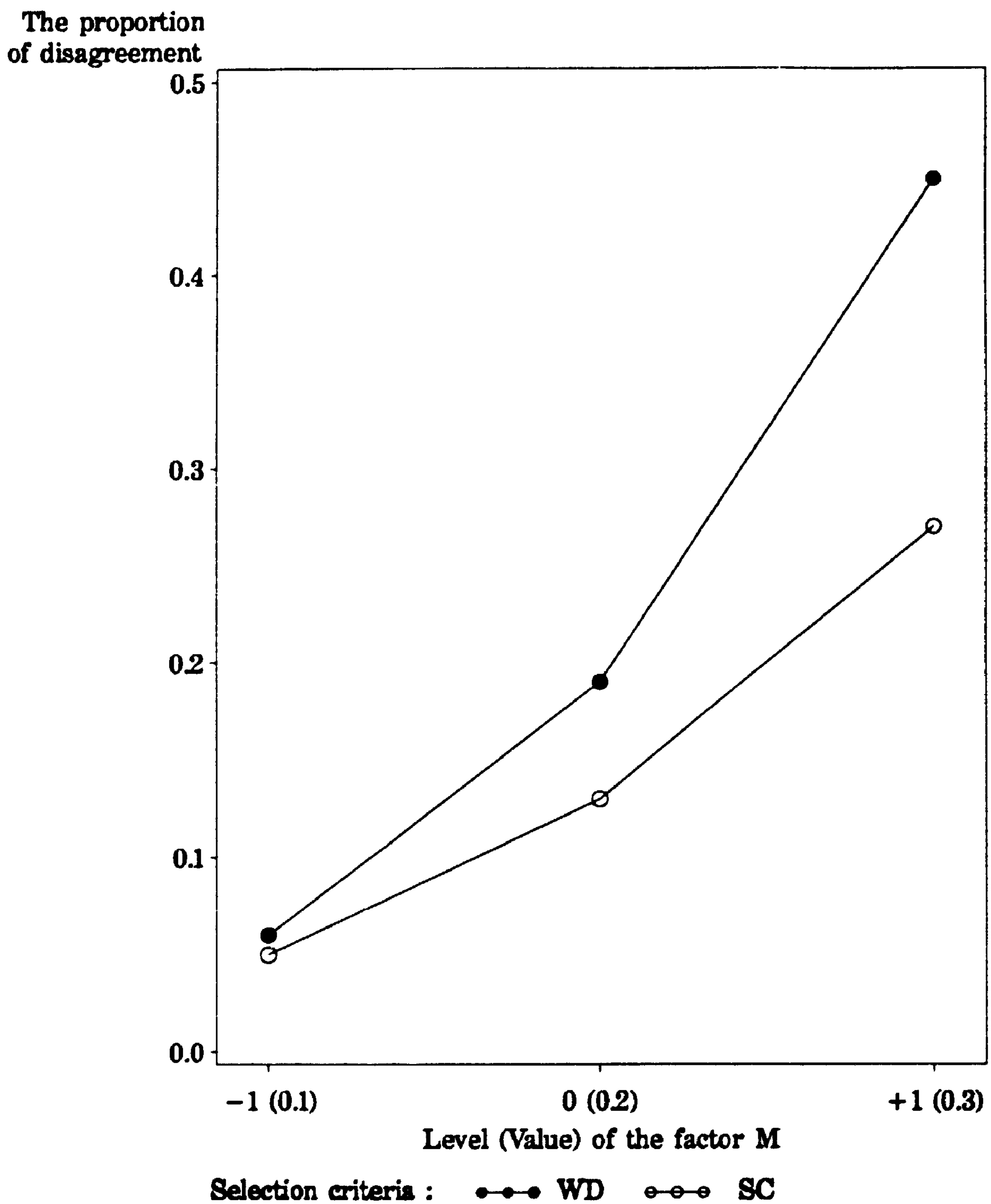


Figure 5.2.2.4

The proportion of disagreement for the factor N

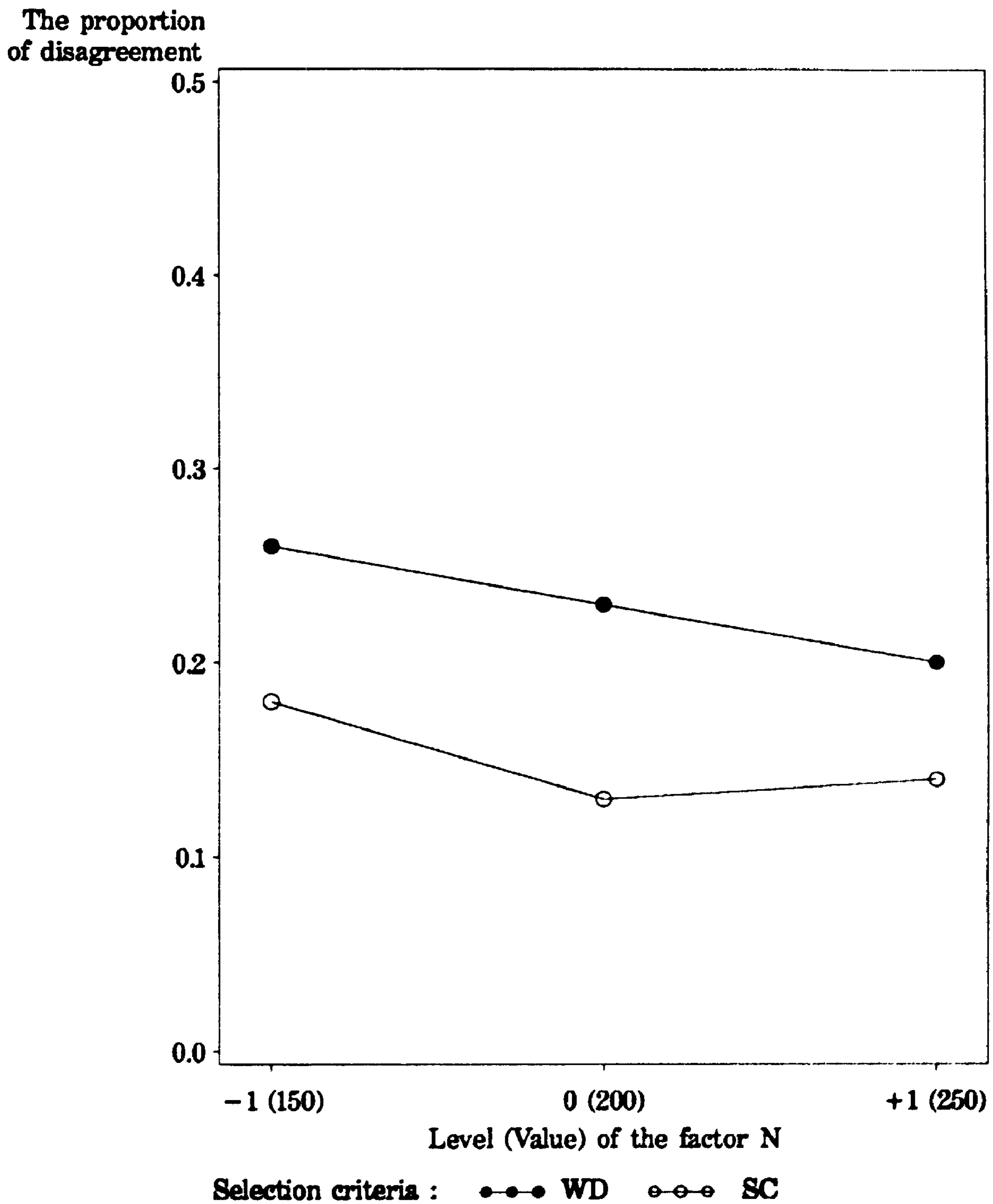


Figure 5.2.2.5

Effect of PB interaction on the proportion of disagreement for the WD selection criterion

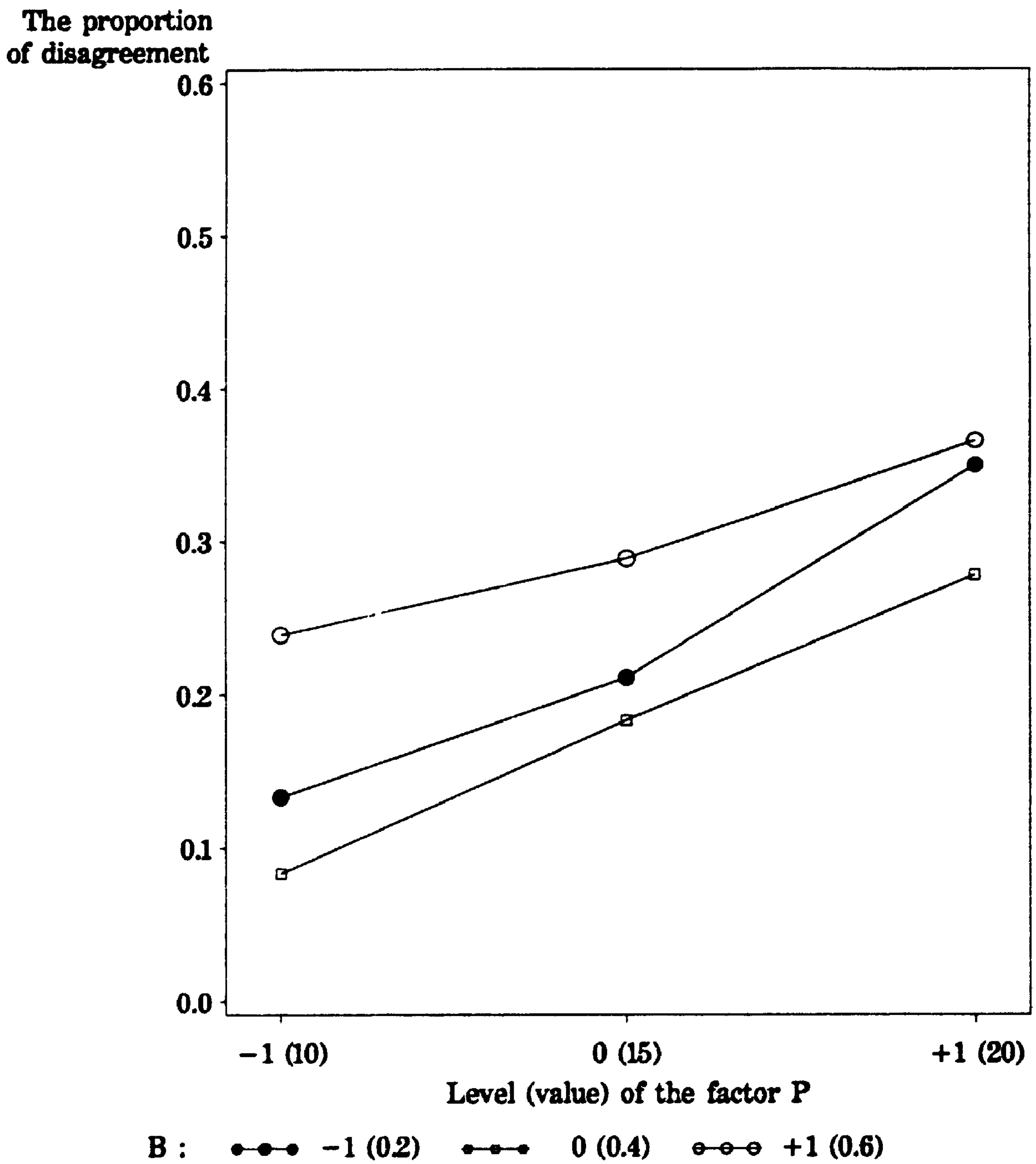


Figure 5.2.2.6

Effect of PM interaction on the proportion of disagreement for the WD selection criterion

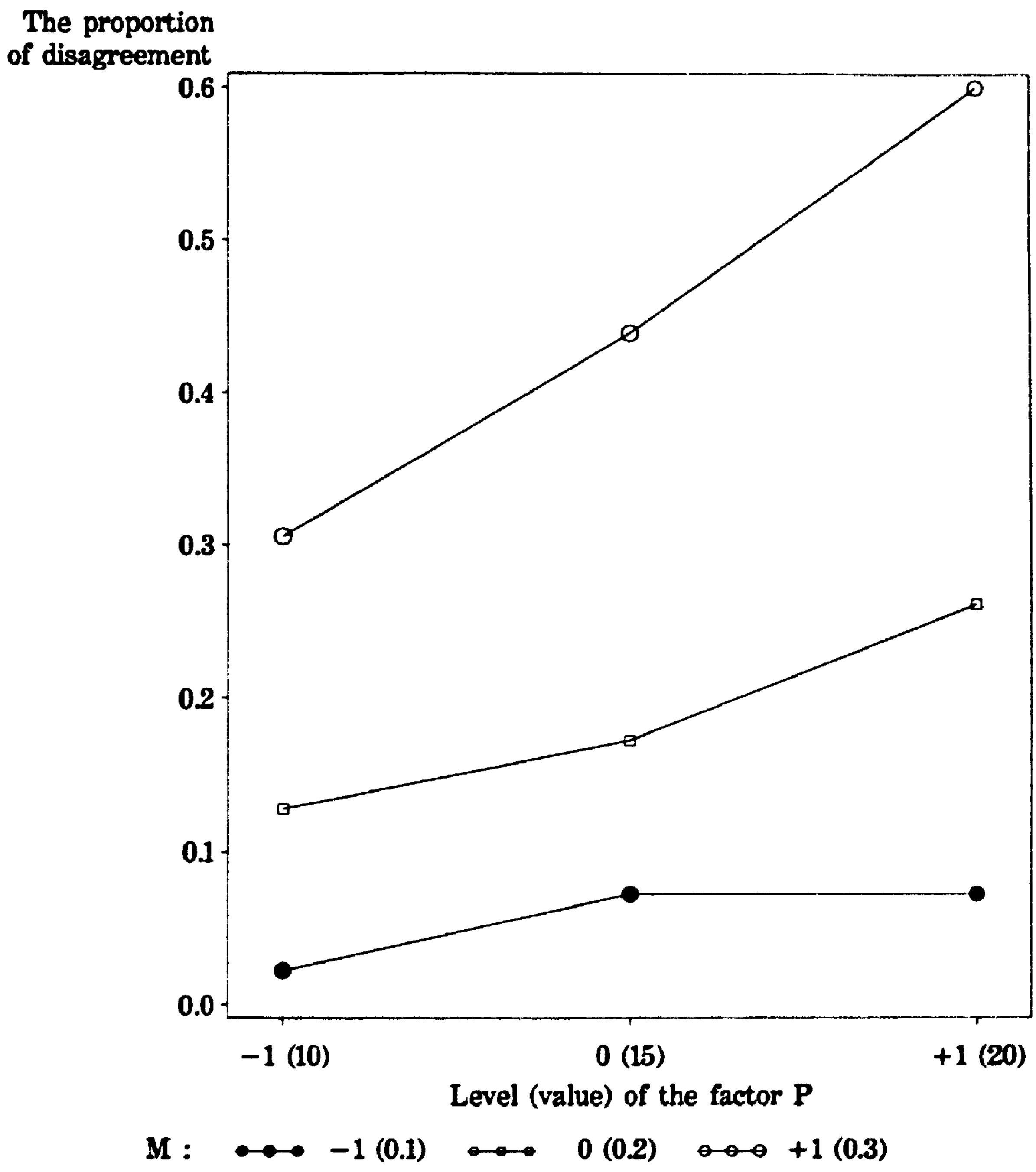


Figure 5.2.2.7

Effect of BM interaction on the proportion of disagreement for the WD selection criterion

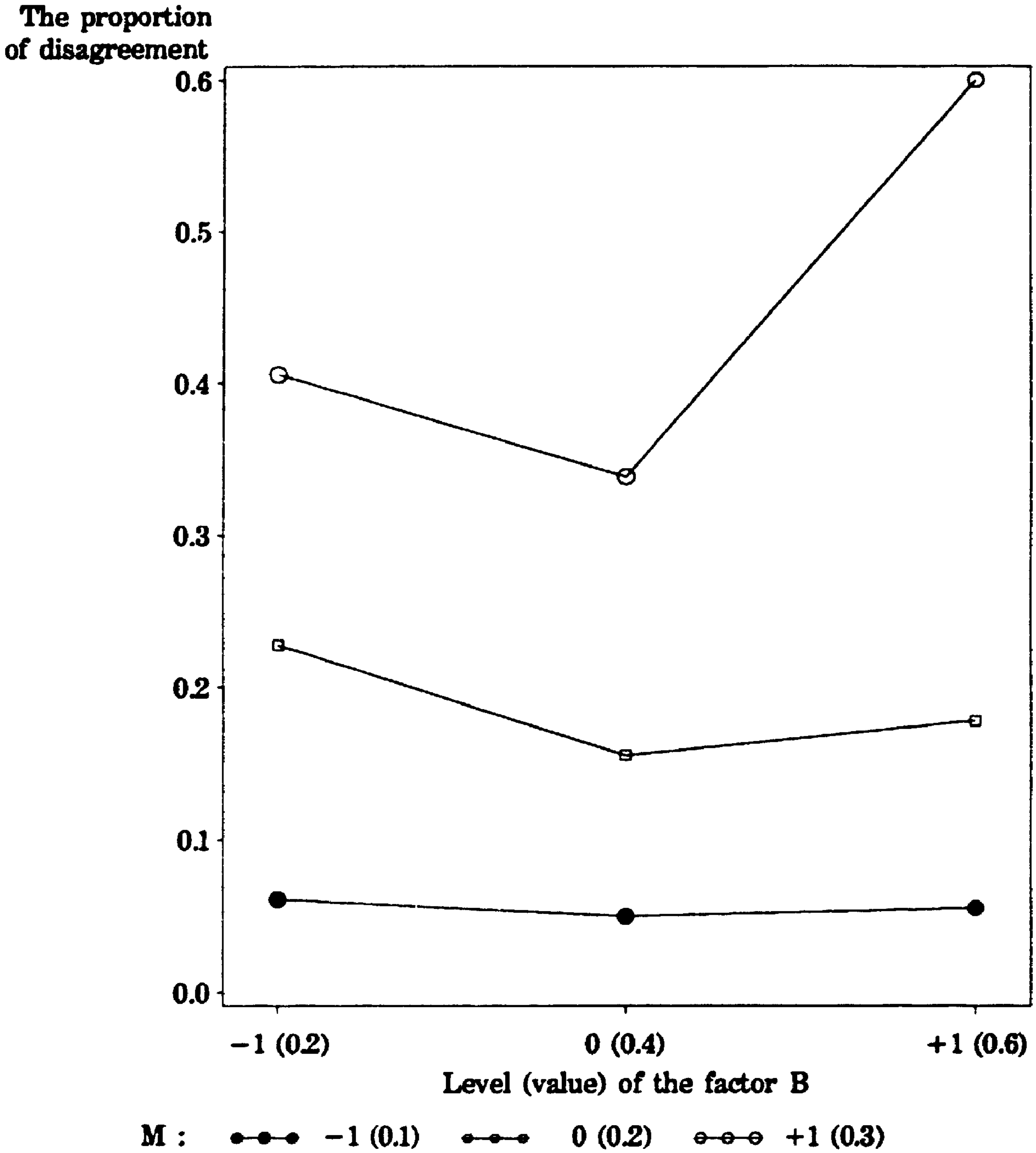
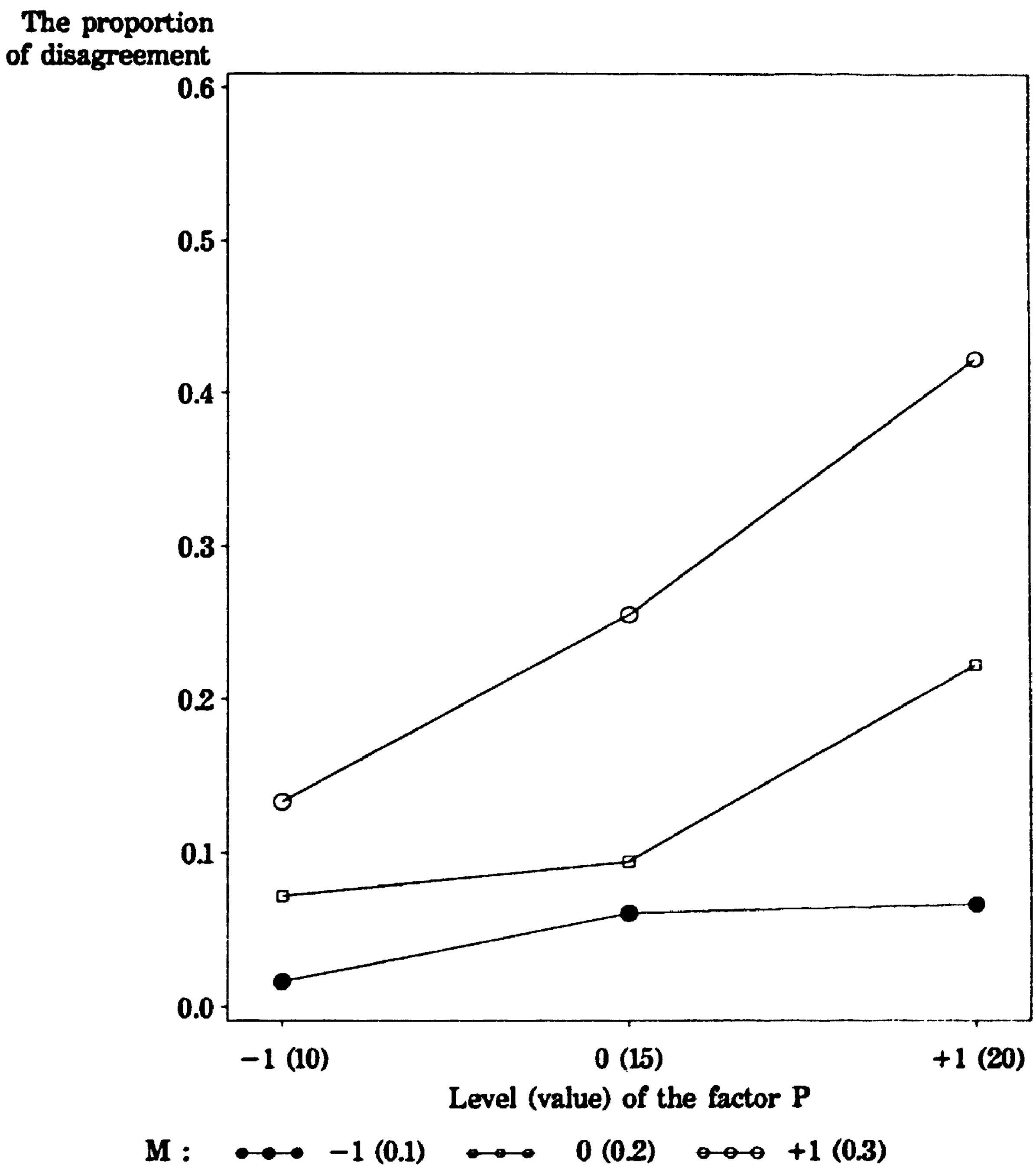


Figure 5.2.2.8

Effect of PM interaction on the proportion of disagreement for the SC selection criterion



5.3 Effects of different structures of predictor variables on ARR, Bias, and $\hat{ERR}$

A variety of structures, that is combinations of values of predictor variables, was examined in the multivariate normal and multivariate binary cases as discussed in sections 4.1 and 4.2, respectively. For the multivariate normal case, the structures of predictor variables depend on the five factors P, V, Δ^2, D, and N. For the multivariate binary case, the structures of the predictor variables depend on the four factors P, B, M, and N. This section examines the effects of these factors on the apparent error rate (ARR), its bias (Bias), and the estimated true error rate ($\hat{ERR}$).

We note that the arc-sine transformation was made on ARR and $\hat{ERR}$ to stabilize the variance. The arc-sine transformation is recommended when responses are proportions, because proportional-type data typically do not have a uniform variance (Draper and Smith, 1981, p238). The arc-sine of $ARR^{1/2}$ and $\hat{ERR}^{1/2}$, denoted by $\sin^{-1}ARR^{1/2}$ and $\sin^{-1}\hat{ERR}^{1/2}$, respectively, have uniform variance $\frac{1}{4N}$, where N is the number of observations.

5.3.1 Multivariate normal case

Table 5.3.1.1 gives the means of ARR, Bias, and $\hat{ERR}$ over the levels of the five factors P, V, Δ^2, D, and N. The means of ARR, Bias, and $\hat{ERR}$ range from 0.052 to 0.276, from 0.013 to 0.058, and from 0.078 to 0.327, respectively, for all five factors. The results of table 5.3.1.1 are graphically presented in figures 5.3.1.1 through 5.3.1.5 for the five factors P, V, Δ^2, D, and N, respectively.

Response surface analyses were used to assess the effects of the five factors P, V, Δ^2, D, and N on $\sin^{-1}ARR^{1/2}$, Bias, and $\sin^{-1}\hat{ERR}^{1/2}$. Tables 5.3.1.2, 5.3.1 3, and 5.3.1.4 give the results of the response surface analyses for $\sin^{-1}ARR^{1/2}$, Bias, and

$\sin^{-1}\hat{ERR}^{1/2}$, respectively.

Before embarking on the major discussions of the results of the response surface analyses, one remark with regard to the lack-of-fit tests may be made. The lack-of-fit tests for tables 5.3.1.2, 5.3.1.3, and 5.3.1.4 are found to be significant. This means that the quadratic surfaces do not fit the data very well. However, we partially justify the use of the results of the response surface analyses with the following reasons. First, the critical value of the lack-of-fit test is $F_{22,917} = 1.55$ at $\alpha = 0.05$ in these cases. Thus a small value of the lack-of-fit test would be highly significant. Second, as indicated in Box and Draper (1987, p74), when there is a large amount of data, the deficient model may be used with proper caution: the magnitude as well as the significance of the t-values will be considered in the interpretation of the results. Third, the results of the analyses of variance of $\sin^{-1}ARR^{1/2}$, Bias, and $\sin^{-1}\hat{ERR}^{1/2}$ for the full $2^5 = 32$ factorial points are given in tables 5.3.1.5, 5.3.1.6, and 5.3.1.7, respectively, to assist the validation of the results of the response surface analyses. The values of R^2 for tables 5.3.1.5, 5.3.1.6, and 5.3.1.7 are 0.82, 0.90, and 0.81, respectively, which can be considered large values. Both the results of response surface analyses and of the analyses of variance do not appear to be radically different.

Mindful of these justifications, now refer to tables 5.3.1.2, 5.3.1.3, and 5.3.1.4 for the discussions.

The effect of the factor P is moderately significant for $\sin^{-1}ARR^{1/2}$ (t = -8.44), highly significant for Bias (t = 63.74), and not significant for $\sin^{-1}\hat{ERR}^{1/2}$ (t = 0.53, p > 0.59). Although the P^2 term is statistically significant for Bias, the magnitude of the t-value for the P^2 term is less than 10 percent of that of the t-value for the P term (i.e, -4.62 versus 63.74). Figure 5.3.1.1 shows that ARR and Bias are decreasing and increasing functions of P, respectively, and that $\hat{ERR}$ is a slight increasing function of P. However, the non-significant effect of P on $\hat{ERR}$ implies that the trade-offs between decreases in ARR and increases in Bias are well-balanced in the ranges of the simulation parameters in this study.

The effect of the factor V is not significant for any of $\sin^{-1}ARR^{1/2}$, Bias, and $\sin^{-1}\hat{ERR}^{1/2}$ (t = -0.42, -0.51, and -0.40, respectively) Figure 5.3.1.2 confirms these findings.

The effect of the factor M is highly significant for all of $\sin^{-1}ARR^{1/2}$, Bias, and $\sin^{-1}\hat{ERR}^{1/2}$ (t = - 52.16, -38.24, and -54.27, respectively). Although the M^2 term is statistically significant for Bias, the magnitude of the t-value for the M^2 term is much smaller than that of the t-value for the M term (i.e, 3.94 versus -38.24). Figure 5.3.1.3 shows that all of ARR, Bias, and $\hat{ERR}$ are decreasing functions of M.

The effect of the factor D is moderately significant for all of $\sin^{-1}ARR^{1/2}$, Bias, and $\sin^{-1}\hat{ERR}^{1/2}$ (t = 4.76, 3.46, and 4.98, respectively). The D^2 term is also moderately significant for $\sin^{-1}ARR^{1/2}$ and $\sin^{-1}\hat{ERR}^{1/2}$ (t = 4.38 and 4.55, respectively). Figure 5.3.1.4 shows that all of ARR, Bias, and ERR are increasing functions of D. However, as we will see below in the results of the VD interaction, the effect of the factor D largely depends on the factor V.

The effect of the factor N is moderately significant for $\sin^{-1}ARR^{1/2}$ (t = 7.10), highly significant for Bias (t = -50.00), and not significant for $\sin^{-1}\hat{ERR}^{1/2}$ (t = -0.12). It is observed that the N^2 is moderately significant for $\sin^{-1}ARR^{1/2}$ and Bias (t = -3.54 and 8.11, respectively) and that weakly significant for $\sin^{-1}\hat{ERR}^{1/2}$ (t = -2.16). Figure 5.3.1.5 shows that ARR is an increasing function of N and that Bias is a decreasing function of N. Accordingly, there are not much differences in $\hat{ERR}$ among the five levels of N.

Before considering the interactions of the five factors P, V, Δ^2, D, and N on

$\sin^{-1}ARR^{1/4}$, Bias, and $\sin^{-1}\hat{ERR}^{1/4}$, two types of interaction must be distinguished: 'quantitative' interaction and 'qualitative' interaction. In the results of clinical trials Peto (1982) called variation in the magnitude, but not the direction, of a treatment effect 'quantitative' interaction, and he used the term 'qualitative' interaction to describe reversals in direction of a treatment effect among some subsets of patients. He stressed the medical importance of 'qualitative' interactions. For instance, 'qualitative' interactions arise when a new treatment, compared with a control treatment, is beneficial in some subset of patients and harmful in other subsets. Gail and Simon (1985) used the terms 'non-crossover' and 'crossover' interactions as synonyms for 'quantitative' and 'qualitative' interactions, respectively.

Table 5.3.1.2 for $\sin^{-1}ARR^{1/4}$ shows that the PN and VD interactions are moderately significant (t = 3.51 and -2.02, respectively). Figures 5.3.1.6 and 5.3.1.7 show the behaviors of the PN and VD interactions, respectively, on ARR at the factorial levels. The PN interaction shows that there is a larger decrease in ARR for the smaller sample size (N=150) than the larger sample size (N=250) as P increases from 10 to 20. The VD interaction shows that the effect of D on ARR largely depends on the level of V. Specifically, ARR is a increasing function of D for the smaller value of V, while ARR is a slight decreasing function of D for the larger value of V. Because VD is a crossover interaction, whereas PN is a non-crossover interaction the VD interaction is more important than the PN interaction.

Table 5.3.1.3 for Bias shows that the four interactions PM, PN, VD, and MN are moderately significant (t = -9.98, - 7.53, 2.78, and 8.70, respectively). Figures 5.3.1.8 through 5.3.1.11 show the behaviors of the PM, PN, VD, and MN interactions, respectively, on Bias at the factorial levels. Although the p-value for the VD interaction is largest among these four interactions, VD is a crossover interaction and the other three interactions are noncrossover interactions. The VD interaction shows that Bias is a increasing function of D for the smaller value of V, while Bias is a slight decreasing function of D for the larger value of V.

Table 5.3.1.4 for $\sin^{-1}\hat{ERR}^{1/2}$ shows that only the VD interaction is moderately significant (t = -2.05). The graph of the VD interaction is presented in figure 5.3.1.12. This figure shows that VD is a crossover interaction; $\hat{ERR}$ is a increasing function of D for the smaller value of V, while $\hat{ERR}$ is a slight decreasing function of D for the larger value of V.

5.3.2 Multivariate binary case

The tables and graphs for the multivariate binary case are analogous to those in the multivariate normal case.

Table 5.3.2.1 gives the means of ARR, Bias, and $\hat{ERR}$ over the levels of the four factors P, B, M, and N. The means of ARR, Bias, and $\hat{ERR}$ range from 0.124 to 0.317, from 0.034 to 0.060, and from 0.158 to 0.376, respectively, for all four factors. The results of table 5.3.2.1 are graphically presented in figures 5.3.2.1 through 5.3.2.4 for the four factors P, B, M, and N, respectively.

Analyses of variance of $\sin^{-1}ARR^{1/2}$, Bias, and $\sin^{-1}\hat{ERR}^{1/2}$ were used to assess effects of the four factors P, B, M, and N. Tables 5.3.2.2 through 5.3.2.4 give the results of the analyses of variances of $\sin^{-1}ARR^{1/2}$, Bias, and $\sin^{-1}\hat{ERR}^{1/2}$, respectively. It is noted that the values of R^2 for tables 5.3.2.2 through 5.3.2.4 are 0.89, 0.88, and 0.89, respectively, which indicate that these ANOVA fit the data well.

The effect of the factor P is highly significant for all of $\sin^{-1}ARR^{1/2}$, Bias, and $\sin^{-1}\hat{ERR}^{1/2}$ (F-ratio = 1259.65, 865.40, and 819.06, respectively). Figure 5.3.2.1 shows that ARR and $\hat{ERR}$ are decreasing functions of P and that Bias is an increasing function of P. The result of $\hat{ERR}$ means that the trade-offs between decreases in ARR and increases in Bias are not well-balanced; the decrease in ARR are much larger than the increase in Bias as P increases.

The effect of the factor B is highly significant for all of $\sin^{-1}ARR^{½}$, Bias, and $\sin^{-1}E\hat{R}R^{½}$ (F-ratio = 369.51, 154.97, and 370.75, respectively). Figure 5.3.2.2 shows that all of ARR, Bias, and $E\hat{R}R$ are increasing functions of B.

The effect of the factor M is highly significant for all of $\sin^{-1}ARR^{½}$, Bias, and $\sin^{-1}E\hat{R}R^{½}$ (F-ratio = 4445.90, 3348.82, and 4843.10, respectively). Figure 5.3.2.3 shows that all of ARR, Bias, and $E\hat{R}R$ are decreasing functions of M. These results are the same as those of the factor Δ^2 in the multivariate normal case.

The effect of the factor N is moderately significant for $\sin^{-1}ARR^{½}$ (F-ratio = 51.46), highly significant for Bias (F-ratio = 1389.60), and not significant for $\sin^{-1}E\hat{R}R^{½}$ (F-ratio = 1.93). Figure 5.3.2.3 shows that ARR is an increasing function of N and that Bias is a decreasing function of N. Accordingly, there are not much differences in $E\hat{R}R$ among the three levels of N. The same behavior for the factor N was observed in the multivariate normal case.

Consider the interactions of the four factors P, B, M, and N on $\sin^{-1}ARR^{½}$, Bias, and $\sin^{-1}E\hat{R}R^{½}$. Although the ANOVA tables include the second, third, and fourth order interactions, because of the complexity of interpreting the third and fourth order interactions discussion is restricted to the second order interactions.

There are many interactions which are statistically significant for $\sin^{-1}ARR^{½}$, Bias, and $\sin^{-1}E\hat{R}R^{½}$. Since all of these were found to be non-crossover interactions, the PM, MN, and BM interactions which are most significant for $\sin^{-1}ARR^{½}$, Bias, and $\sin^{-1}E\hat{R}R^{½}$, respectively, were chosen to depict the behaviors of these interactions. Figures 5.3.2.5, 5.3.2.6, and 5.3.2.7 show that the PM, MN, and BM interactions are slightly nonparallel. As in the multivariate normal case, tables

5.3.2.2, 5.3.2.3, and 5.3.2.4 show that the F-ratios of the significant interactions are much smaller than those of the significant main effects.

5.3.3 Conclusions

1. ARR is a decreasing function of P and Δ^2 in the multivariate normal case, and of P and M in the multivariate binary case.

2. ARR is an increasing function of N in the multivariate normal case, and of B and N in the multivariate binary case.

3. Bias is a decreasing function of Δ^2 and N in the multivariate normal case, and of M and N in the multivariate binary case.

4. Bias is an increasing function of P in the multivariate normal case, and of P and B in the multivariate binary case.

5. $E\hat{R}R$ is a decreasing function of Δ^2 in the multivariate normal case, and of P and M in the multivariate binary case.

6. $E\hat{R}R$ is a increasing function of D in the multivariate normal case, and of B in the multivariate binary case.

7. In the multivariate normal case, Δ^2 is the most significant factor with largest t-value for ARR and $E\hat{R}R$, while P is for Bias.

8. In the multivariate binary case, M is the most significant factor with largest F-ratio for all of ARR, Bias, and $E\hat{R}R$.

Table 5.3.1.1
Means of ARR, Bias, and $\hat{ERR}$ over the levels of the five factors P, V, Δ^2, D, and N

Factor	[a]Level	Value	$\overline{ARR}$(s.e.)[b]	$\overline{Bias}$(s.e.)[b]	$\overline{\hat{ERR}}$(s.e.)[b]
	-2	5	0.155 (0.50)	0.013 (0.03)	0.168 (0.52)
	-1	10	0.151 (0.36)	0.026 (0.04)	0.177 (0.38)
P	0	15	0.144 (0.32)	0.036 (0.05)	0.180 (0.35)
	+1	20	0.133 (0.36)	0.045 (0.07)	0.178 (0.40)
	+2	25	0.135 (0.67)	0.052 (0.10)	0.187 (0.74)
	-2	0.2	0.139 (0.47)	0.035 (0.05)	0.174 (0.50)
	-1	0.4	0.142 (0.36)	0.036 (0.08)	0.178 (0.39)
V	0	0.6	0.144 (0.33)	0.036 (0.07)	0.180 (0.36)
	+1	0.8	0.141 (0.37)	0.036 (0.08)	0.177 (0.40)
	+2	1.0	0.142 (0.46)	0.035 (0.06)	0.176 (0.49)
	-2	1.0	0.276 (0.85)	0.051 (0.12)	0.327 (0.95)
	-1	1.5	0.199 (0.19)	0.041 (0.08)	0.240 (0.19)
Δ^2	0	2.0	0.141 (0.20)	0.035 (0.06)	0.176 (0.20)
	+1	2.5	0.084 (0.15)	0.030 (0.06)	0.114 (0.16)
	+2	3.0	0.052 (0.41)	0.026 (0.06)	0.078 (0.44)
	-2	0.2	0.131 (0.59)	0.034 (0.05)	0.165 (0.62)
	-1	0.4	0.140 (0.37)	0.035 (0.08)	0.175 (0.40)
D	0	0.6	0.140 (0.31)	0.036 (0.07)	0.176 (0.33)
	+1	0.8	0.143 (0.36)	0.036 (0.08)	0.179 (0.39)
	+2	1.0	0.209 (0.74)	0.041 (0.08)	0.250 (0.79)
	-2	100	0.112 (0.64)	0.058 (0.10)	0.171 (0.71)
	-1	150	0.135 (0.37)	0.043 (0.08)	0.178 (0.41)
N	0	200	0.146 (0.32)	0.035 (0.06)	0.181 (0.35)
	+1	250	0.148 (0.35)	0.028 (0.05)	0.177 (0.38)
	+2	300	0.142 (0.55)	0.024 (0.04)	0.167 (0.56)

[a]The number of sampling situations for each level is 1 for levels ±2, 16 for levels ±1, and 14 for level 0.
[b]The standard error of mean is multiplied by 10^{-2}.

Table 5.3.1.2

Response surface analysis of $\sin^{-1}ARR^{1/2}$ for the five factors P, V, M, D, and N

Factor	d.f	$\hat{\beta}$	s.e	t-value	p-value
[a]P	1	-0.0138	0.0016	-8.44	0.00
V	1	-0.0007	0.0016	-0.42	0.68
[a]M	1	-0.0851	0.0016	-52.16	0.00
[a]D	1	0.0078	0.0016	4.76	0.00
[a]N	1	0.0116	0.0016	7.10	0.00
P^2	1	0.0000	0.0019	0.01	0.99
V^2	1	-0.0014	0.0019	-0.76	0.45
M^2	1	-0.0001	0.0019	-0.05	0.96
[a]D^2	1	0.0082	0.0019	4.38	0.00
[a]N^2	1	-0.0066	0.0019	-3.54	0.00
PV	1	-0.0000	0.0018	-0.02	0.99
PM	1	-0.0027	0.0018	-1.50	0.14
PD	1	0.0008	0.0018	0.41	0.68
[a]PN	1	0.0064	0.0018	3.51	0.00
VM	1	-0.0010	0.0018	-0.54	0.59
[b]VD	1	-0.0037	0.0018	-2.02	0.04
VN	1	-0.0001	0.0018	-0.06	0.95
MD	1	0.0009	0.0018	0.48	0.63
MN	1	0.0014	0.0018	0.75	0.45
DN	1	-0.0009	0.0018	-0.51	0.61

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

Lack-of-fit test: $F_{22,917} = 3.888$ ($p < 0.001$)

Table 5.3.1.3

Response surface analysis of Bias for the five factors P, V, M, D, and N

Factor	d.f	$\hat{\beta}$	s.e	t-value	p-value
[a]P	1	0.00952	0.00015	63.74	0.00
V	1	-0.00008	0.00015	-0.51	0.61
[a]M	1	-0.00571	0.00015	-38.24	0.00
[a]D	1	0.00052	0.00015	3.46	0.00
[a]N	1	-0.00747	0.00015	-50.00	0.00
[a]P^2	1	-0.00079	0.00017	-4.62	0.00
V^2	1	-0.00029	0.00017	-1.69	0.09
[a]M^2	1	0.00068	0.00017	3.94	0.00
D^2	1	0.00030	0.00017	1.74	0.08
[a]N^2	1	0.00139	0.00017	8.11	0.00
PV	1	-0.00004	0.00017	-0.24	0.81
[a]PM	1	-0.00167	0.00017	-9.98	0.00
PD	1	0.00005	0.00017	0.27	0.78
[a]PN	1	-0.00126	0.00017	-7.53	0.00
VM	1	-0.00011	0.00017	-0.65	0.52
[b]VD	1	-0.00046	0.00017	-2.78	0.01
VN	1	0.00014	0.00017	0.85	0.39
MD	1	0.00027	0.00017	1.61	0.11
[a]MN	1	0.00145	0.00017	8.70	0.00
DN	1	0.00004	0.00017	0.24	0.81

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

Lack-of-fit test: $F_{22,917} = 4.452$ ($p < 0.001$)

Table 5.3.1.4

Response surface analysis of $\sin^{-1}\hat{ERR}^{1/2}$ for the five factors P, V, M, D, and N

Factor	d.f	$\hat{\beta}$	s.e	t-value	p-value
P	1	0.0008	0.0016	0.53	0.59
V	1	-0.0006	0.0016	-0.40	0.69
[a]M	1	-0.0842	0.0016	-54.27	0.00
[a]D	1	0.0077	0.0016	4.98	0.00
N	1	-0.0002	0.0016	-0.12	0.91
P^2	1	-0.0009	0.0018	-0.49	0.62
V^2	1	-0.0016	0.0018	-0.88	0.38
M^2	1	0.0016	0.0018	0.93	0.36
[a]D^2	1	0.0081	0.0018	4.55	0.00
[b]N^2	1	-0.0038	0.0018	-2.16	0.03
PV	1	0.0001	0.0017	0.04	0.97
PM	1	-0.0018	0.0017	-1.05	0.29
PD	1	0.0005	0.0017	0.31	0.76
PN	1	0.0033	0.0017	1.90	0.06
VM	1	-0.0009	0.0017	-0.52	0.60
[b]VD	1	-0.0035	0.0017	-2.05	0.04
VN	1	-0.0001	0.0017	-0.03	0.97
MD	1	0.0011	0.0017	0.61	0.54
MN	1	0.0008	0.0017	0.46	0.64
DN	1	-0.0006	0.0017	-0.34	0.73

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

Lack-of-fit test: $F_{22,917} = 4.045$ ($p < 0.001$)

Table 5.3.1.5
Analysis of variance of $\sin^{-1}ARR^{1/2}$ for the 2^5 factorial design

Source of Variation	Degrees of Freedom	Mean Square	F-ratio	P-value
[a]P	1	0.15105	66.22	0.00
V	1	0.00078	0.34	0.56
[a]M	1	4.73394	2075.32	0.00
D	1	0.00642	2.81	0.09
[a]N	1	0.08642	37.89	0.00
PV	1	0.00000	0.00	0.99
PM	1	0.00477	2.09	0.15
PD	1	0.00036	0.16	0.69
[a]PN	1	0.02630	11.53	0.00
VM	1	0.00062	0.27	0.60
[b]VD	1	0.00866	3.80	0.05
VN	1	0.00001	0.00	0.95
MD	1	0.00050	0.22	0.64
MN	1	0.00120	0.52	0.47
DN	1	0.00055	0.24	0.63
PVM	1	0.00038	0.17	0.68
[b]PVD	1	0.00902	3.96	0.05
PVN	1	0.00392	1.72	0.19
PMD	1	0.00007	0.03	0.86
[b]PMN	1	0.00877	3.85	0.05
PDN	1	0.00191	0.84	0.36
VMD	1	0.00004	0.02	0.89
VMN	1	0.00086	0.38	0.54
VDN	1	0.00401	1.76	0.19
MDN	1	0.00051	0.22	0.64
[b]PVMD	1	0.01267	5.56	0.02
PVMN	1	0.00242	1.06	0.30
PVDN	1	0.00591	2.59	0.11
PMDN	1	0.00451	1.98	0.16
VMDN	1	0.00045	0.20	0.66
PVMDN	1	0.00379	1.66	0.20
Error	608	0.00228		

[a]significant with $p < 0.001$
[b]significant with $p < 0.05$
$R^2 = 0.82$

Table 5.3.1.6
Analysis of variance of Bias for the 2^5 factorial design

Source of Variation	Degrees of Freedom	Mean Square	F-ratio	P-value
[a]P	1	0.056922	2798.00	0.00
V	1	0.000003	0.15	0.70
[a]M	1	0.020114	988.73	0.00
D	1	0.000029	1.45	0.23
[a]N	1	0.033249	1634.38	0.00
PV	1	0.000001	0.05	0.82
[a]PM	1	0.001777	87.33	0.00
PD	1	0.000001	0.07	0.80
[a]PN	1	0.001012	49.75	0.00
VM	1	0.000007	0.37	0.55
[b]VD	1	0.000138	6.76	0.01
VN	1	0.000013	0.64	0.42
MD	1	0.000046	2.28	0.13
[a]MN	1	0.001349	66.31	0.00
DN	1	0.000001	0.05	0.82
PVM	1	0.000003	0.17	0.68
PVD	1	0.000071	3.50	0.06
PVN	1	0.000011	0.53	0.47
PMD	1	0.000005	0.24	0.62
[b]PMN	1	0.000155	7.61	0.01
PDN	1	0.000002	0.11	0.74
VMD	1	0.000010	0.49	0.48
VMN	1	0.000001	0.05	0.83
VDN	1	0.000075	3.69	0.06
MDN	1	0.000014	0.70	0.40
[b]PVMD	1	0.000136	6.69	0.01
PVMN	1	0.000000	0.00	1.00
PVDN	1	0.000055	2.69	0.10
PMDN	1	0.000009	0.43	0.51
VMDN	1	0.000012	0.60	0.44
[b]PVMDN	1	0.000090	4.40	0.04
Error	608	0.000020		

[a]significant with $p < 0.001$
[b]significant with $p < 0.05$
$R^2 = 0.90$

Table 5.3.1.7
Analysis of variance of $\sin^{-1}\hat{ERR}^{1/2}$ for the 2^5 factorial design

Source of Variation	Degrees of Freedom	Mean Square	F-ratio	P-value
P	1	0.00016	0.08	0.78
V	1	0.00060	0.30	0.59
[a]M	1	4.59645	2272.73	0.00
D	1	0.00602	2.98	0.09
N	1	0.00001	0.00	0.96
PV	1	0.00000	0.00	0.97
PM	1	0.00212	1.05	0.31
PD	1	0.00018	0.09	0.76
PN	1	0.00694	3.43	0.06
VM	1	0.00052	0.26	0.61
[b]VD	1	0.00807	3.99	0.05
VN	1	0.00000	0.00	0.98
MD	1	0.00072	0.35	0.55
MN	1	0.00041	0.20	0.65
DN	1	0.00022	0.11	0.74
PVM	1	0.00026	0.13	0.72
PVD	1	0.00733	3.63	0.06
PVN	1	0.00346	1.71	0.19
PMD	1	0.00009	0.04	0.83
PMN	1	0.00590	2.92	0.09
PDN	1	0.00100	0.49	0.48
VMD	1	0.00000	0.00	0.98
VMN	1	0.00099	0.49	0.48
VDN	1	0.00344	1.70	0.19
MDN	1	0.00043	0.21	0.64
[b]PVMD	1	0.01174	5.81	0.02
PVMN	1	0.00176	0.87	0.35
PVDN	1	0.00043	0.21	0.64
PMDN	1	0.00353	1.74	0.19
VMDN	1	0.00022	0.11	0.74
PVMDN	1	0.00351	1.73	0.19
Error	608	0.00202		

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

$R^2 = 0.81$

Figure 5.3.1.1

Mean of ARR, Bias, and estimated ERR for the factor P

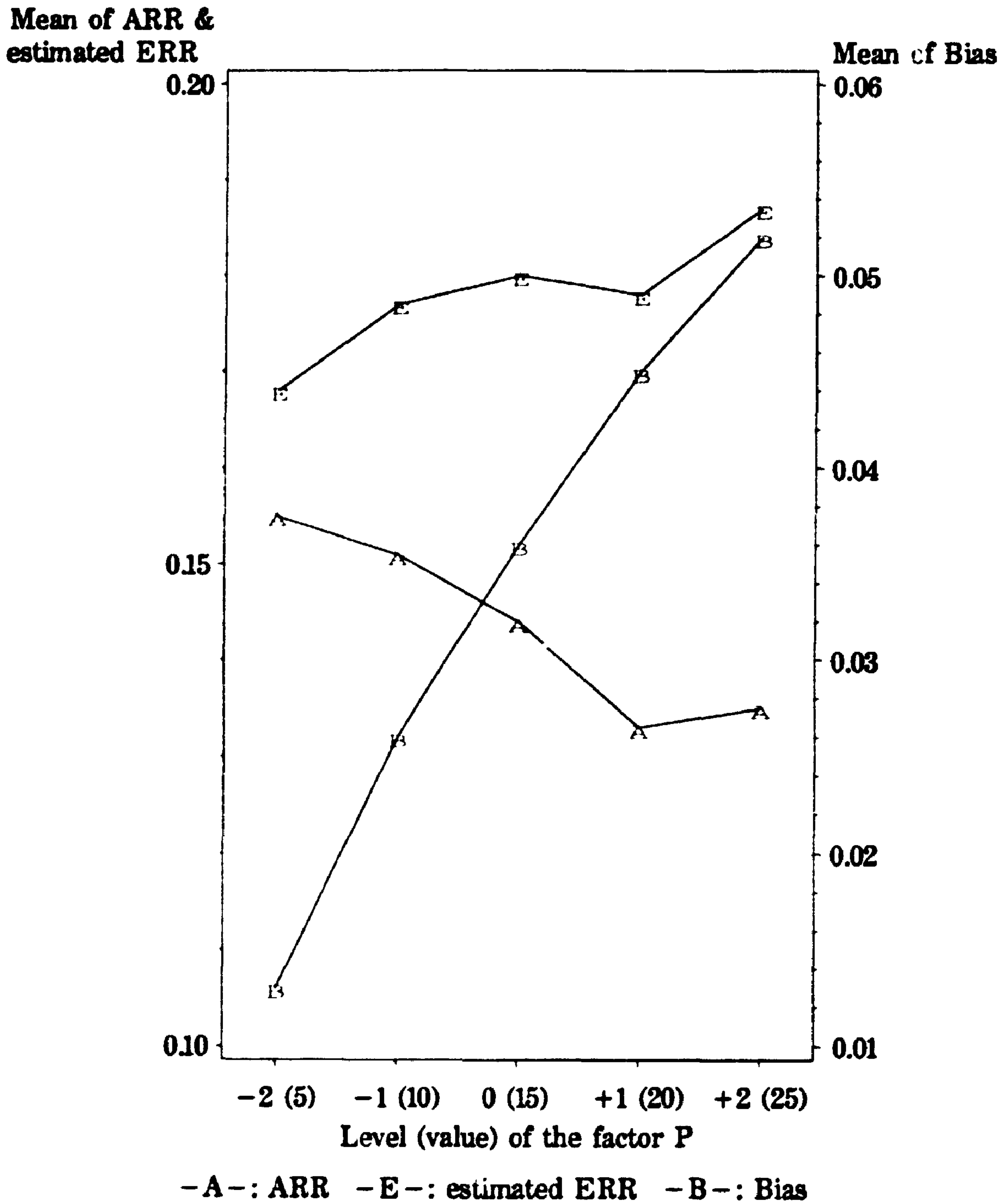


Figure 5.3.1.2

Mean of ARR, Bias, and estimated ERR for the factor V

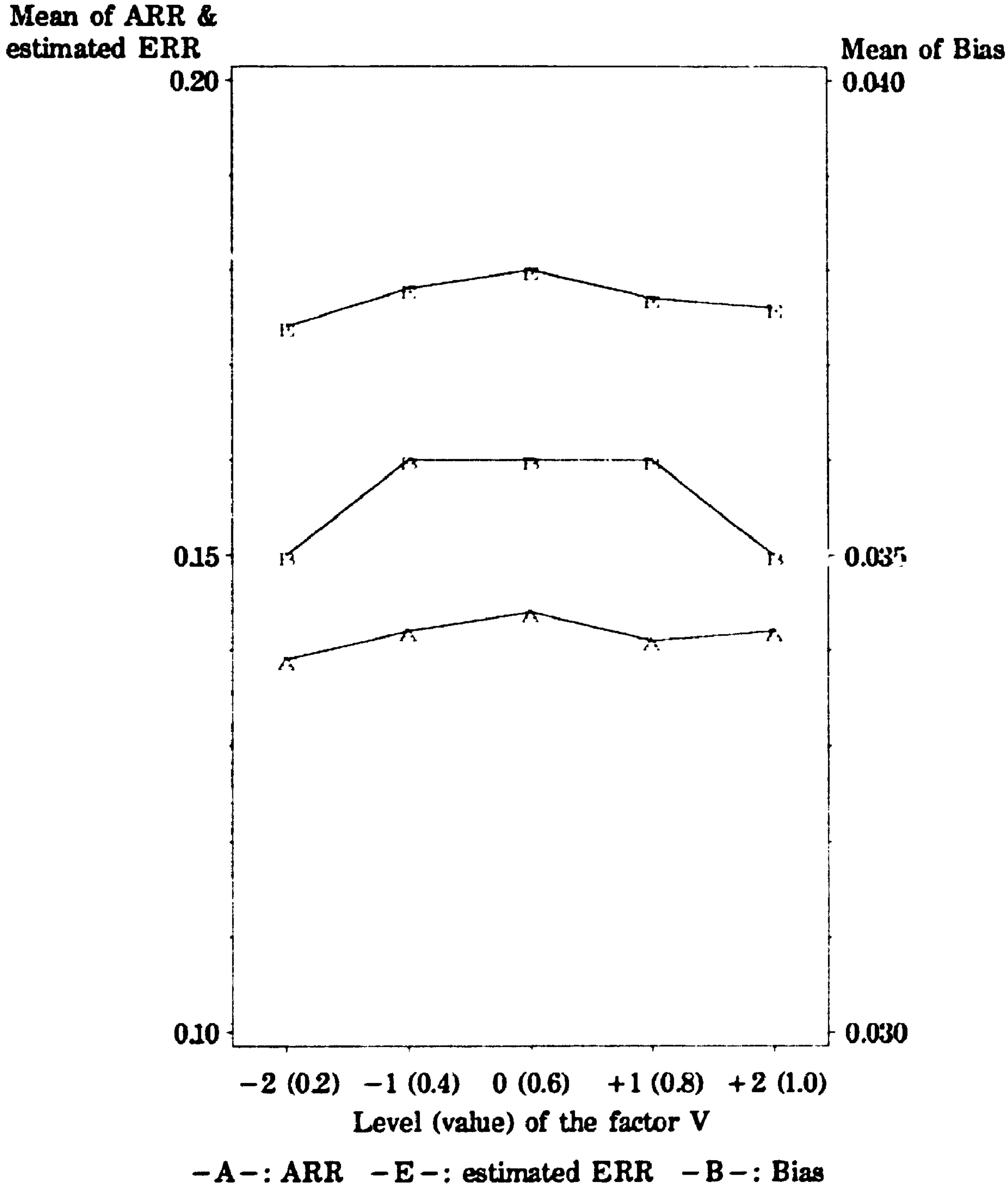


Figure 5.3.1.3

Mean of ARR, Bias, and estimated ERR for the factor M

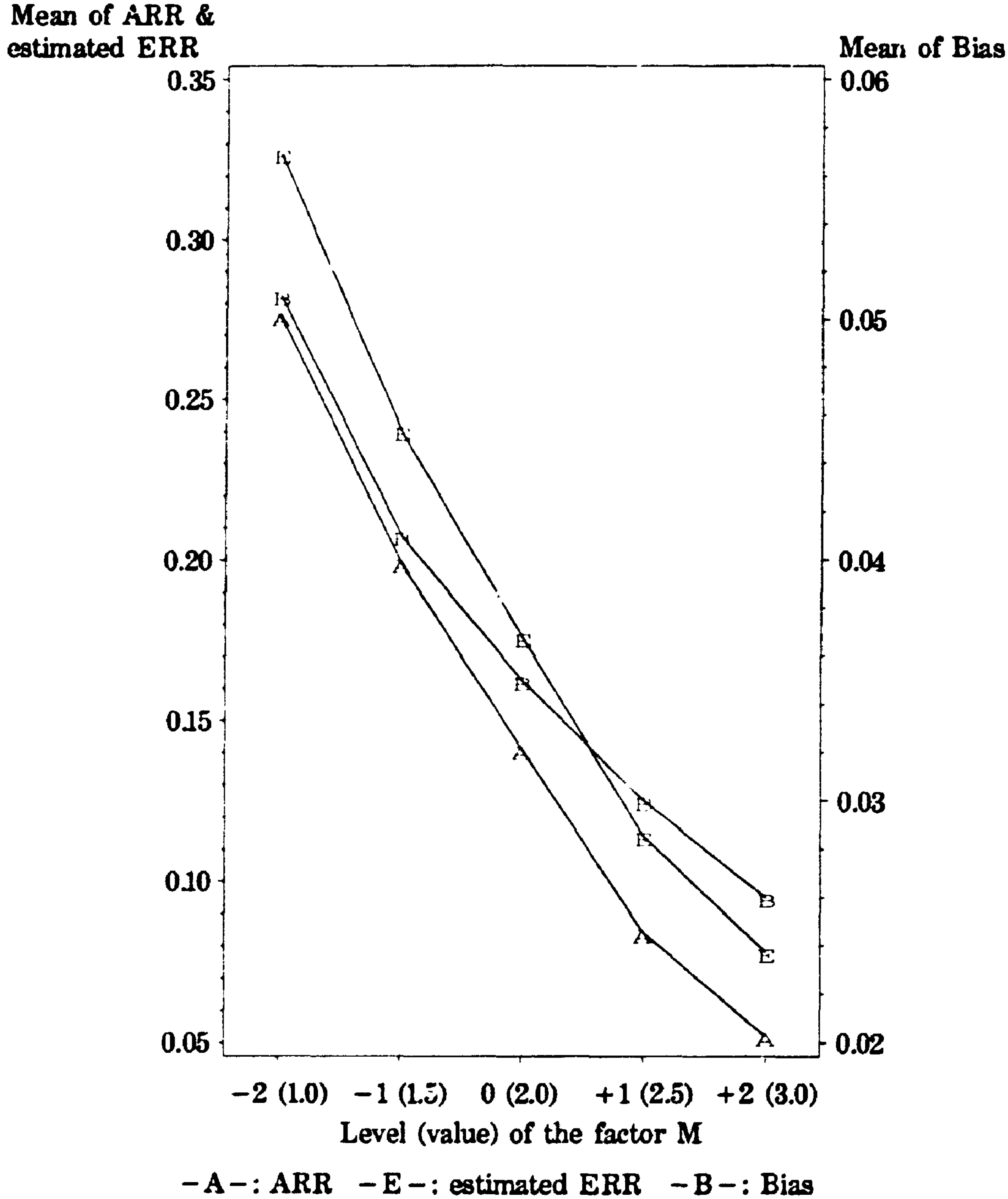


Figure 5.3.1.4

Mean of ARR, Bias, and estimated ERR for the factor D

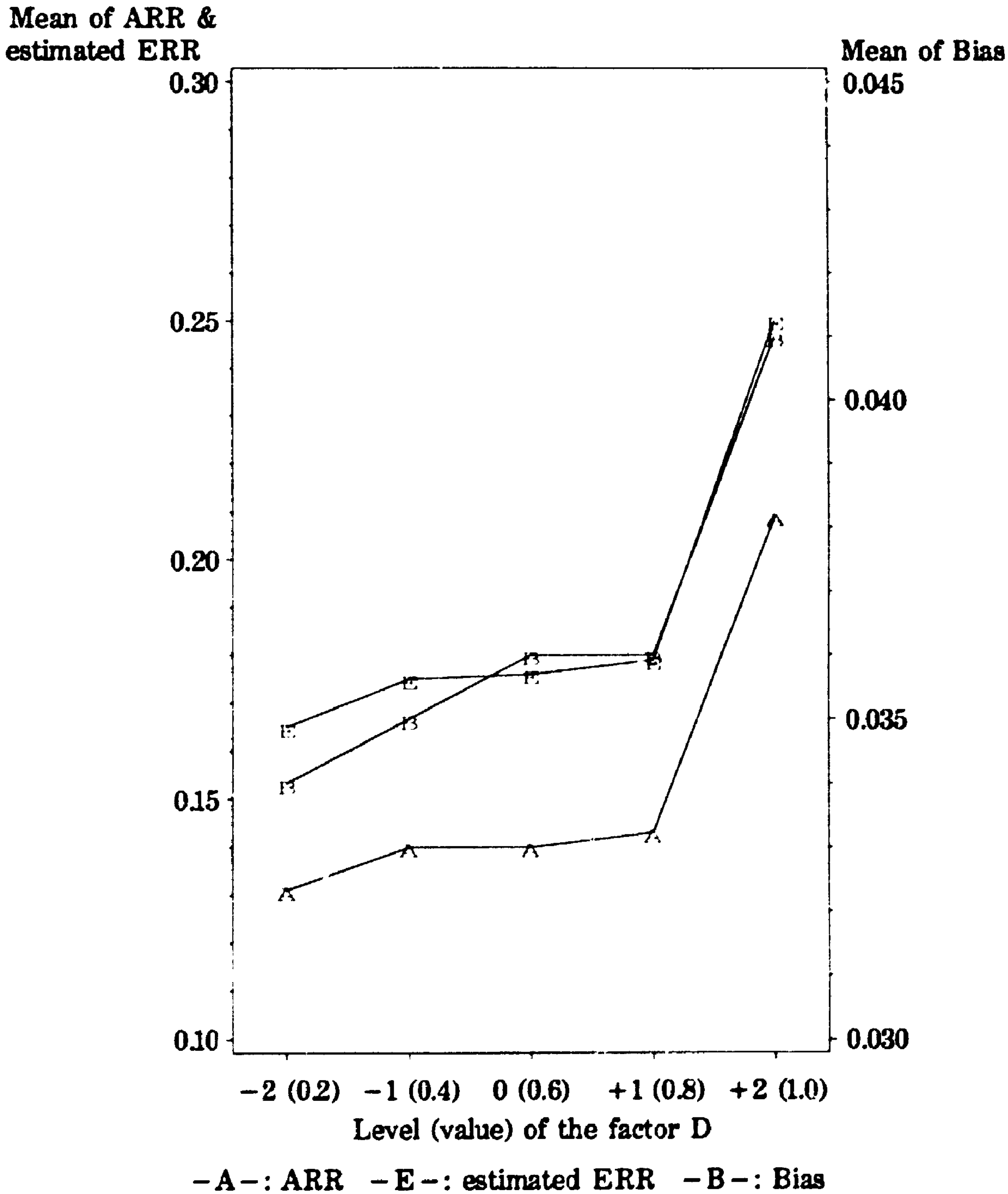


Figure 5.3.1.5

Mean of ARR, Bias, and estimated ERR for the factor N

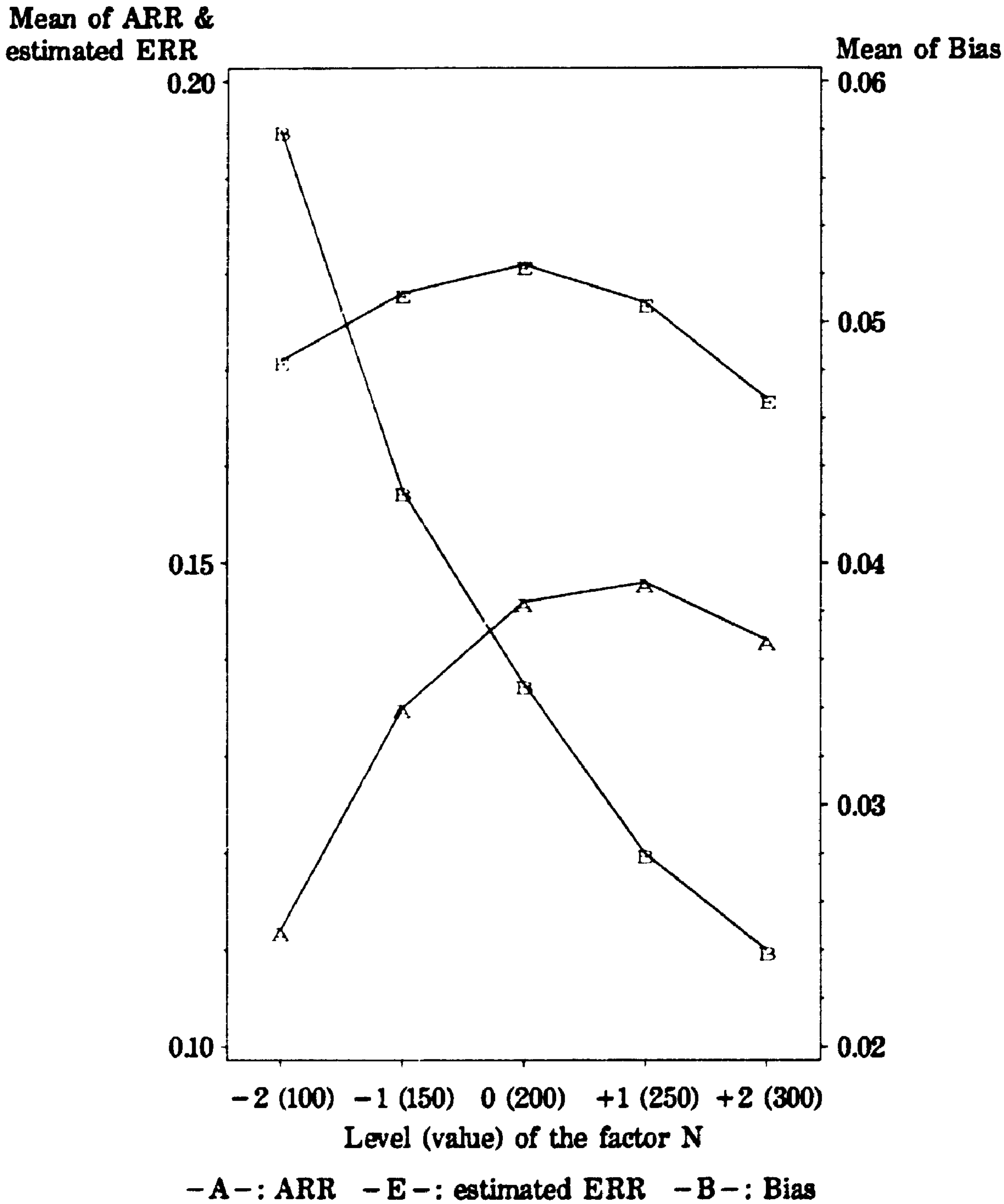


Figure 5.3.1.6

Effect of PN interaction on ARR at the factorial levels

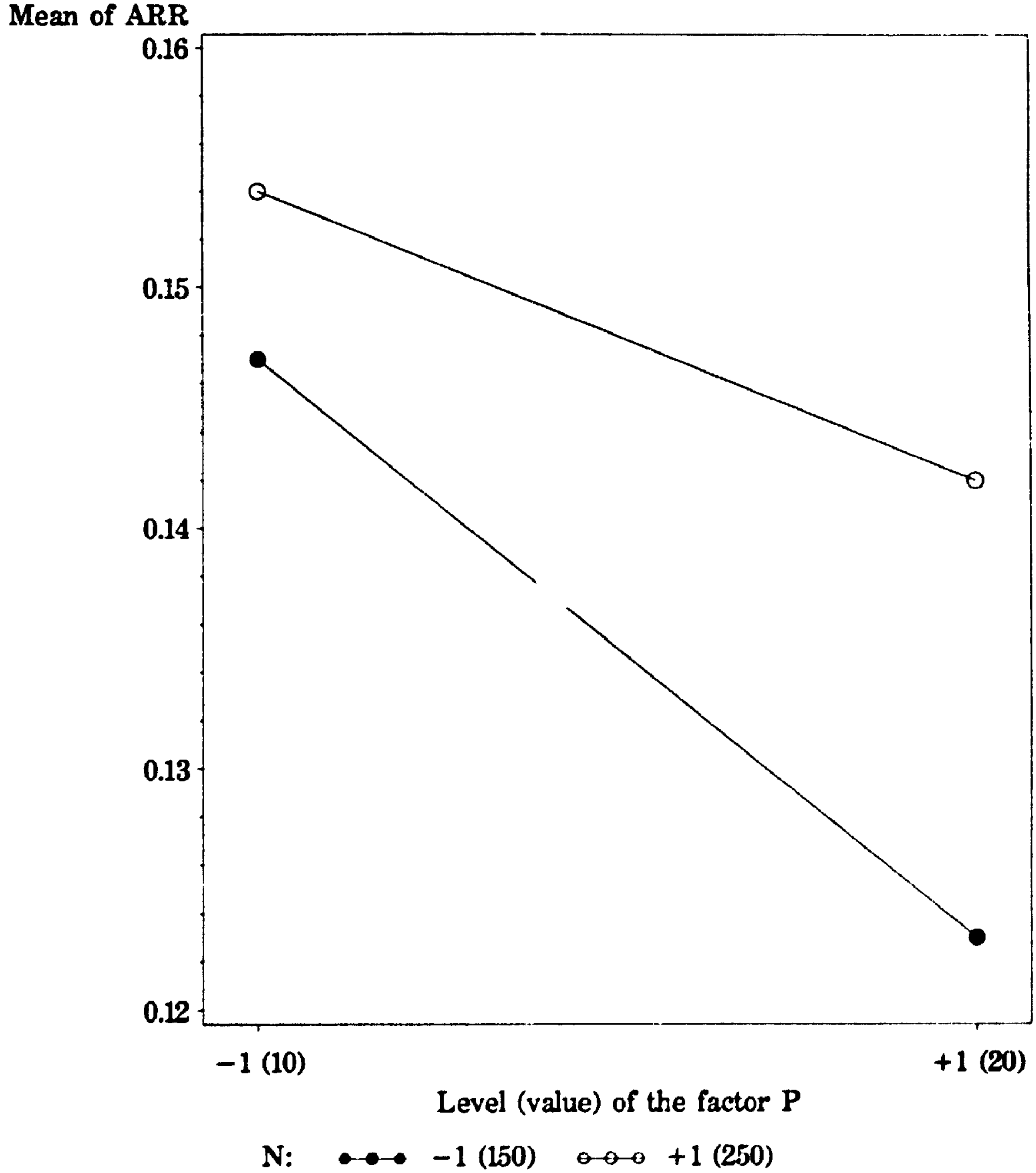


Figure 5.3.1.7

Effect of VD interaction on ARR at the factorial levels

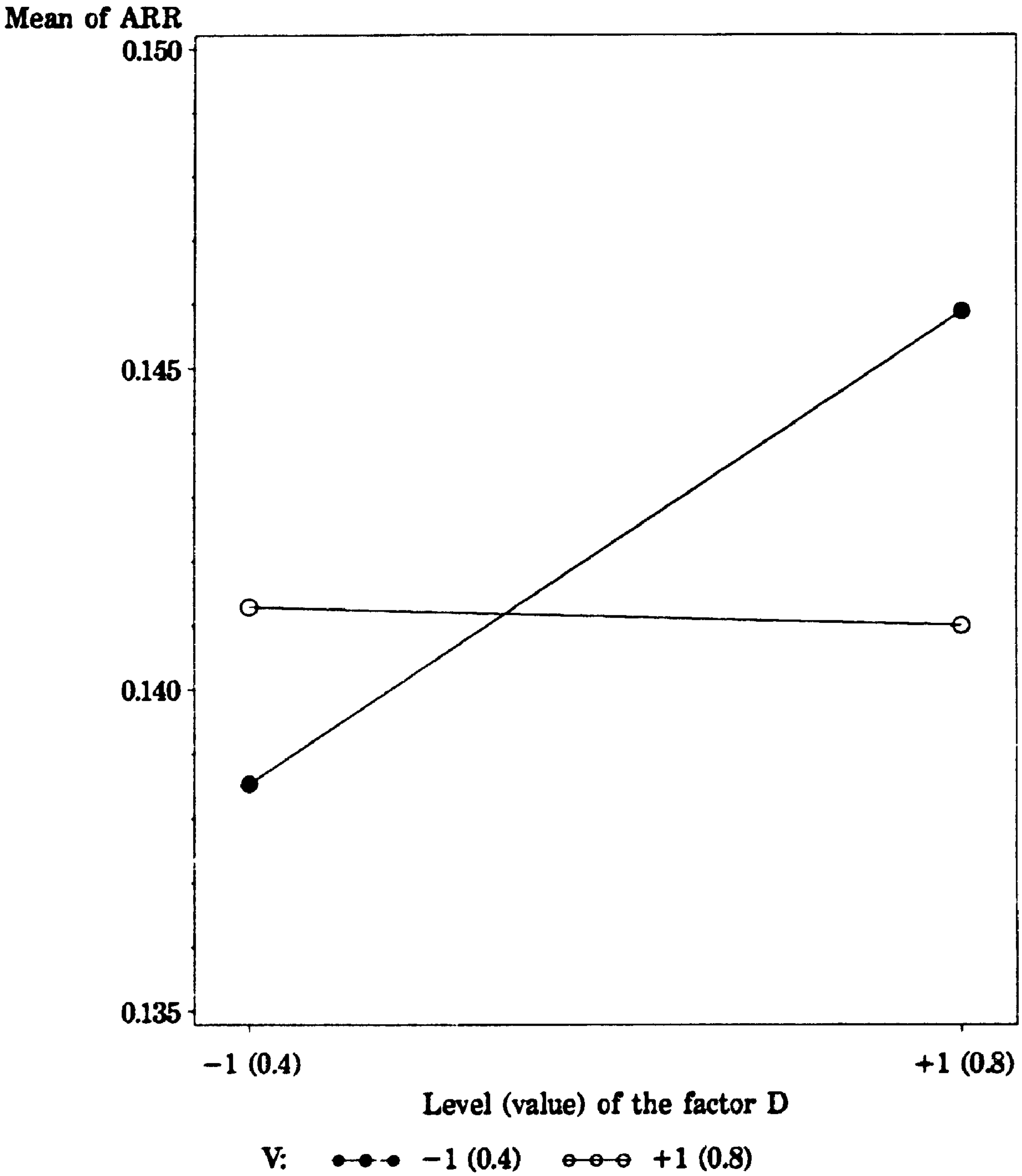


Figure 5.3.1.8
Effect of PM interaction on Bias at the factorial levels

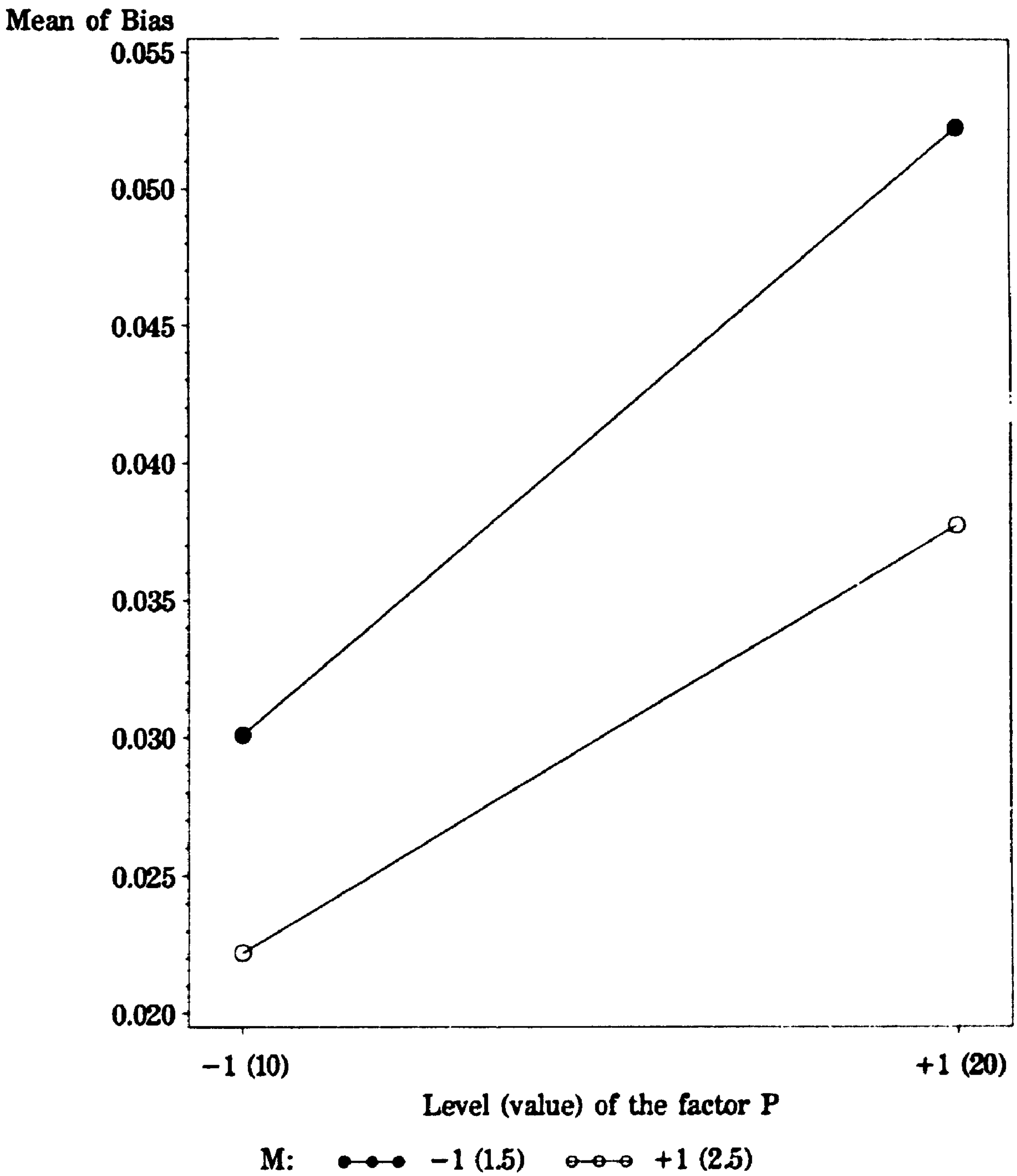


Figure 5.3.1.9
Effect of PN interaction on Bias at the factorial levels

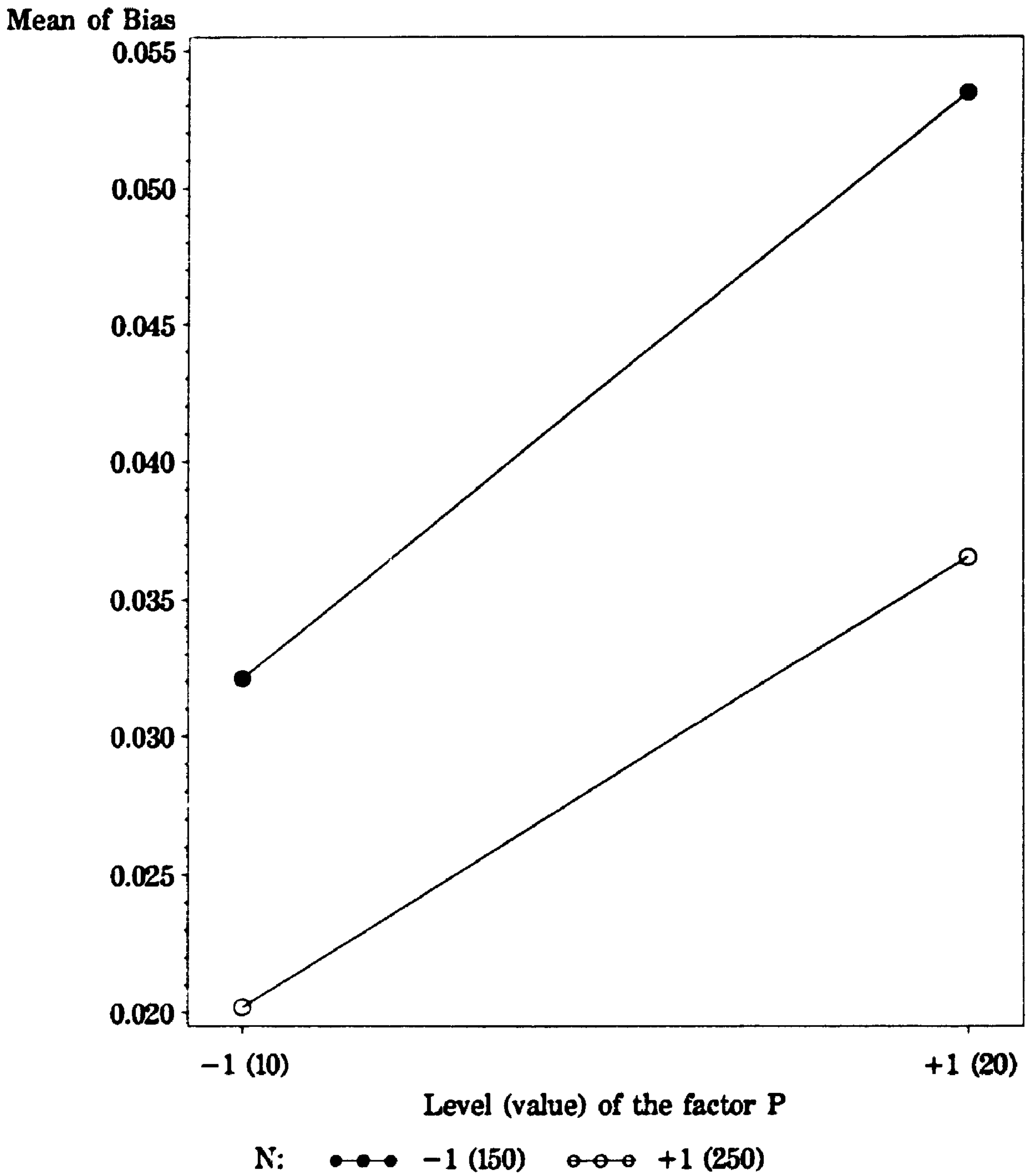


Figure 5.3.1.10

Effect of VD interaction on Bias at the factorial levels

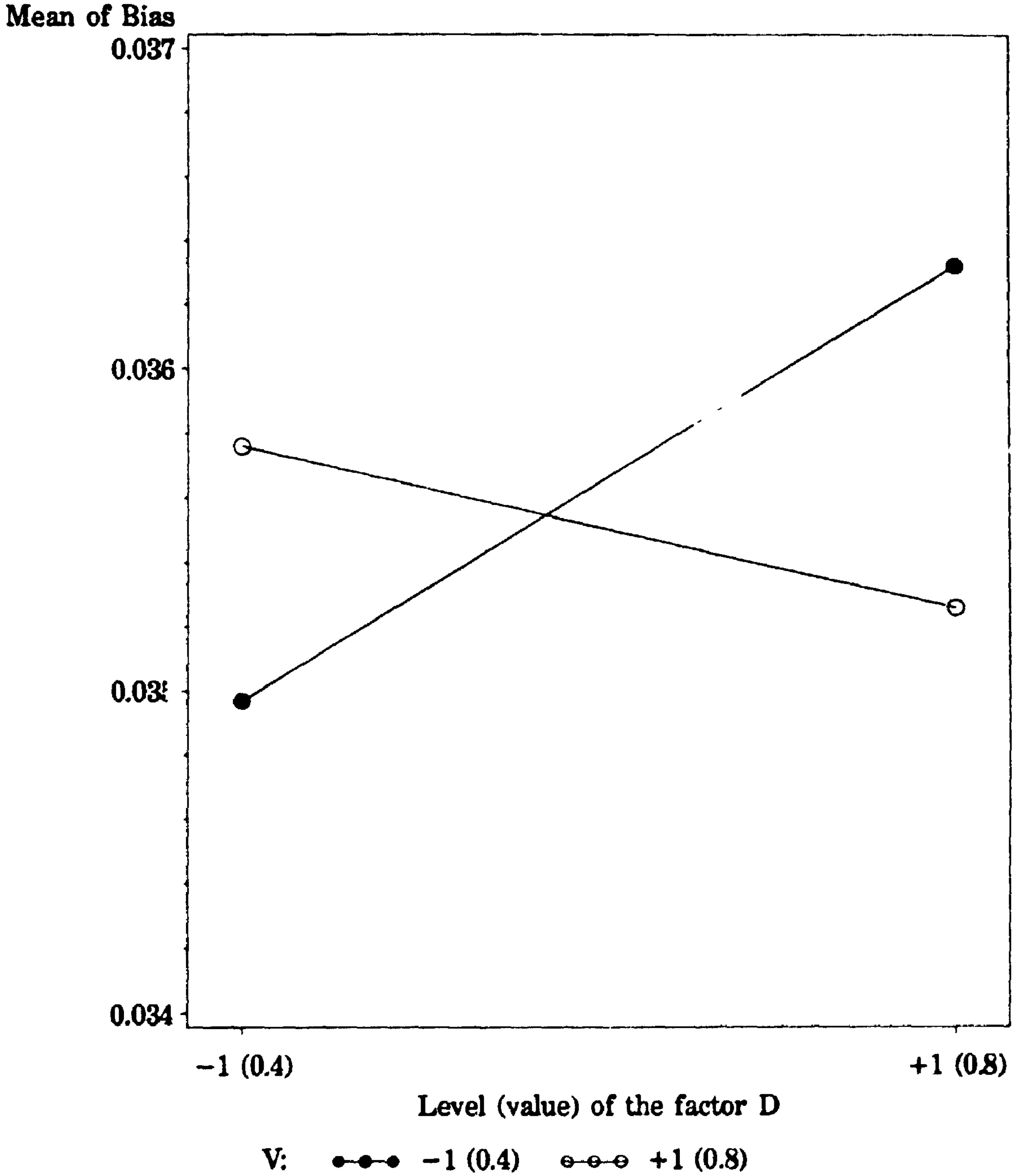


Figure 5.3.1.11

Effect of MN interaction on Bias at the factorial levels

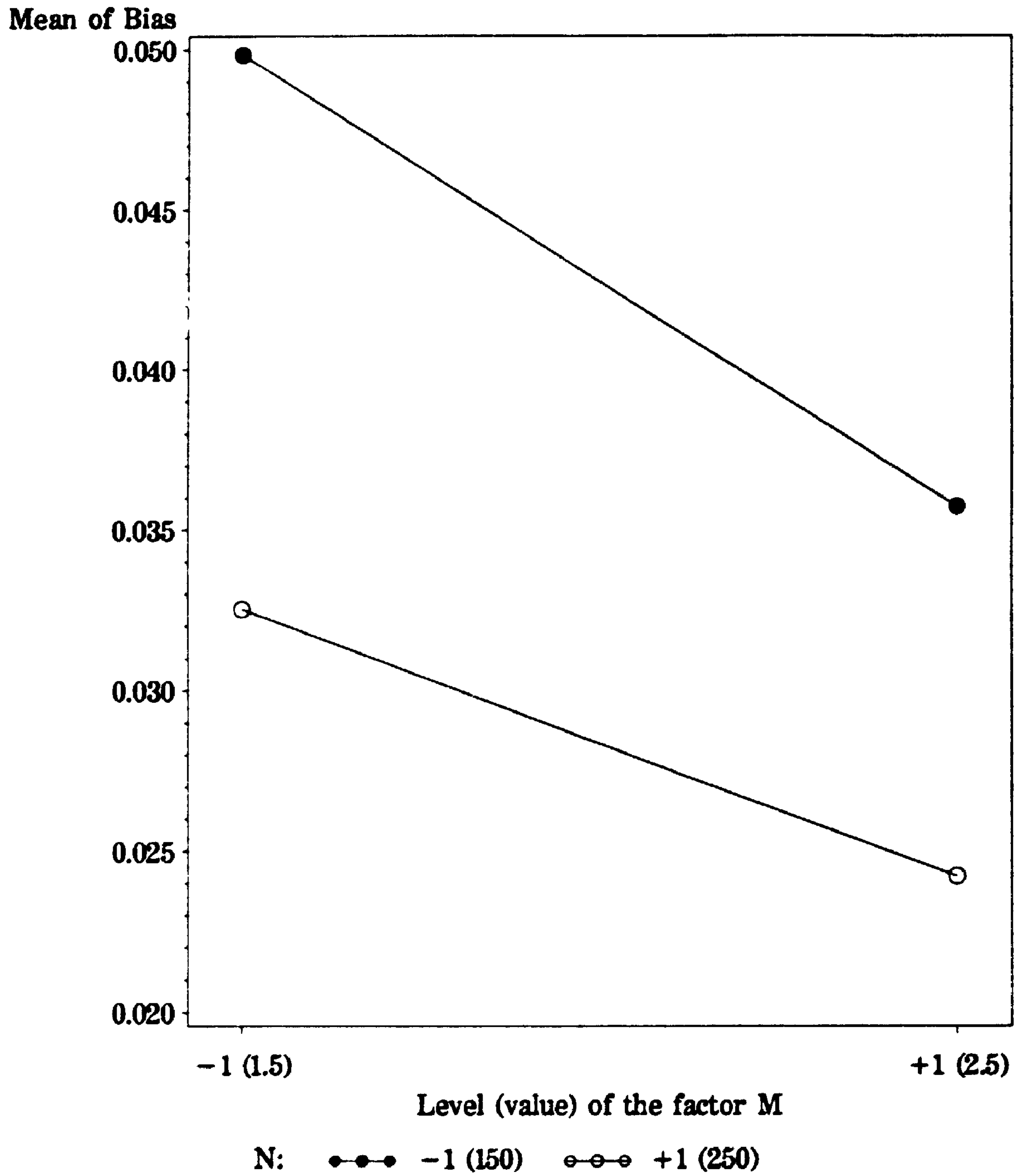


Figure 5.3.1.12
Effect of VD interaction on estimated ERR at the factorial levels

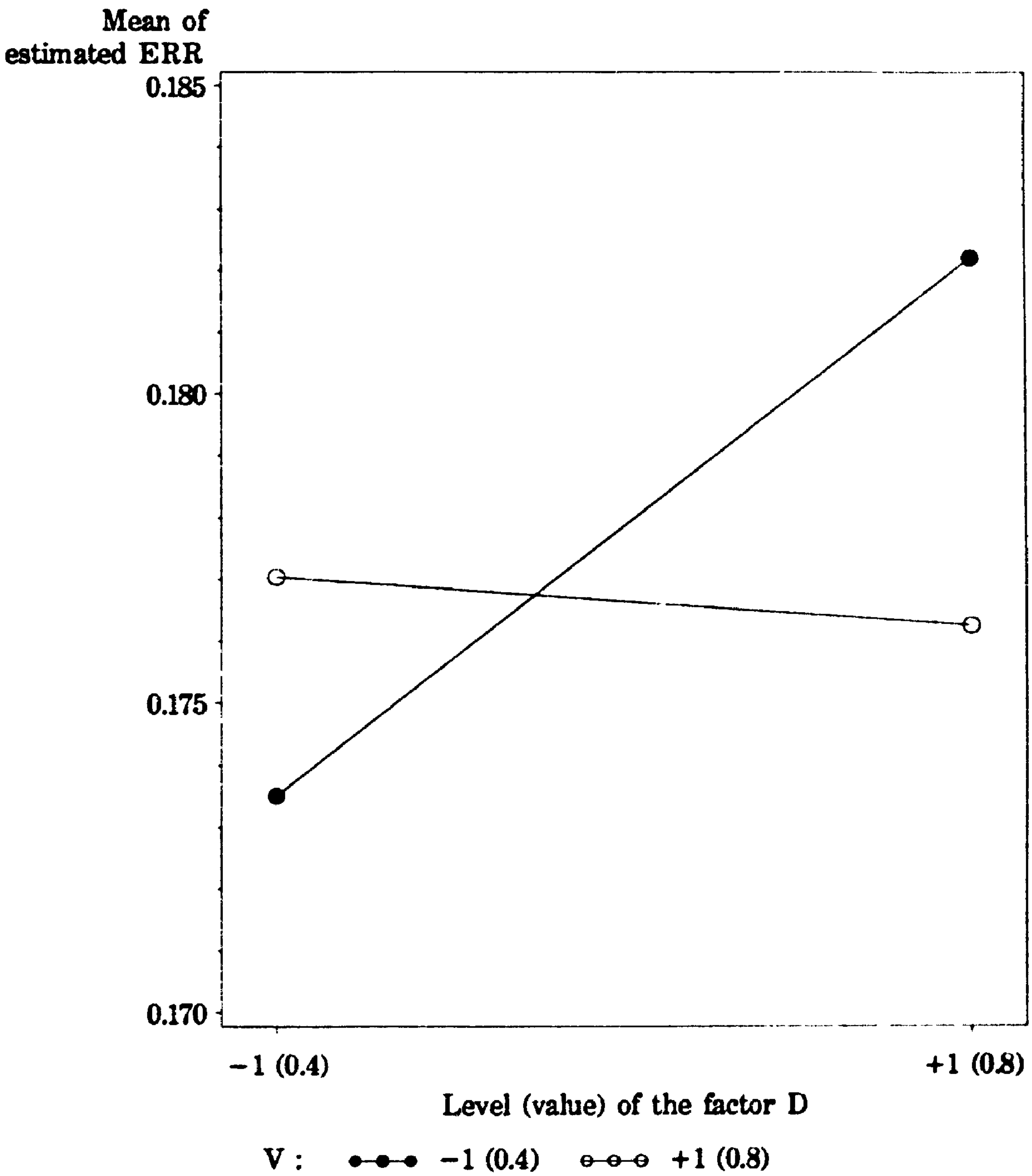


Table 5.3.2.1
Means of ARR, Bias, and $\hat{ERR}$ over the levels of the four factors P, B, M, and N

Factor	[a]Level	Value	$\overline{ARR}$ (s.e.)[b]	$\overline{Bias}$ (s.e.)[b]	$\overline{E\hat{R}R}$ (s.e.)[b]
P	-1	10	0.271 (0.35)	0.039 (0.05)	0.310 (0.39)
	0	15	0.218 (0.39)	0.047 (0.06)	0.265 (0.43)
	+1	20	0.172 (0.40)	0.052 (0.07)	0.224 (0.46)
B	-1	0.2	0.194 (0.44)	0.043 (0.06)	0.237 (0.48)
	0	0.4	0.224 (0.39)	0.047 (0.06)	0.271 (0.42)
	+1	0.6	0.244 (0.39)	0.048 (0.07)	0.292 (0.42)
M	-1	0.1	0.317 (0.21)	0.060 (0.05)	0.376 (0.21)
	0	0.2	0.220 (0.26)	0.044 (0.04)	0.265 (0.25)
	+1	0.3	0.124 (0.26)	0.034 (0.04)	0.158 (0.26)
N	-1	150	0.211 (0.42)	0.055 (0.07)	0.265 (0.47)
	0	200	0.223 (0.41)	0.045 (0.05)	0.268 (0.45)
	+1	250	0.228 (0.42)	0.038 (0.05)	0.266 (0.45)

[a]The number of sampling situations for each level is 27.

[b]The standard error of mean is multiplied by 10^{-2}.

Table 5.3.2.2
Analysis of variance of $\sin^{-1}ARR^{1/2}$
for the four factors P, B, M, and N

Source of Variation	Degrees of Freedom	Mean Square	F-ratio	p-value
[a]P	2	2.41848	1259.65	0.0001
[a]B	2	0.70944	369.51	0.0001
[a]M	2	8.53594	4445.90	0.0001
[a]N	2	0.09880	51.46	0.0001
[a]PB	4	0.02208	11.50	0.0001
[a]PM	4	0.12068	62.85	0.0001
[a]PN	4	0.01650	8.59	0.0001
[a]BM	4	0.10162	52.93	0.0001
BN	4	0.00031	0.16	0.9582
MN	4	0.00268	1.40	0.2322
[a]PBM	8	0.00635	3.31	0.0009
PBN	8	0.00139	0.72	0.6711
[b]PMN	8	0.00490	2.55	0.0092
BMN	8	0.00160	0.83	0.5746
PBMN	16	0.00227	0.2734	
Error	1539	0.00192		

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

$R^2 = 0.89$

Table 5.3.2.3
Analysis of variance of Bias
for the four factors P, B, M, and N

Source of Variation	Degrees of Freedom	Mean Square	F-ratio	p-value
[a]P	2	0.022999	865.40	0.0001
[a]B	2	0.004119	154.97	0.0001
[a]M	2	0.089002	3348.82	0.0001
[a]N	2	0.036932	1389.60	0.0001
[a]PB	4	0.000715	26.92	0.0001
[a]PM	4	0.000712	26.78	0.0001
PN	4	0.000047	1.76	0.1338
[a]BM	4	0.000242	9.12	0.0001
[b]BN	4	0.000096	3.61	0.0061
[a]MN	4	0.001243	46.78	0.0001
[a]PBM	8	0.000173	6.50	0.0001
PBN	8	0.000045	1.69	0.0950
[a]PMN	8	0.000109	4.10	0.0001
[b]BMN	8	0.000061	2.31	0.0185
[b]PBMN	16	0.000057	2.15	0.0052
Error	1539	0.000027		

[a]significant with $P < 0.001$

[b]significant with $P < 0.05$

$R^2 = 0.88$

Table 5.3.2.4
Analysis of variance of $\sin^{-1}\hat{ERR}^{1/2}$
for the four factors P, B, M, and N

Source of Variation	Degrees of Freedom	Mean Square	F-ratio	p-value
[a]P	2	1.53748	819.06	0.0001
[a]B	2	0.69595	370.75	0.0001
[a]M	2	9.09113	4843.10	0.0001
N	2	0.00362	1.93	0.1459
[a]PB	4	0.02299	12.25	0.0001
[a]PM	4	0.08541	45.50	0.0001
[b]PN	4	0.00786	4.19	0.0022
[a]BM	4	0.08579	45.70	0.0001
BN	4	0.00048	0.26	0.9052
MN	4	0.00192	1.02	0.3942
[a]PBM	8	0.00669	3.56	0.0004
PBN	8	0.00130	0.69	0.6994
[b]PMN	8	0.00365	1.95	0.0496
BMN	8	0.00173	0.92	0.4966
PBMN	16	0.00230	1.23	0.2392
Error	1539	0.00188		

[a]significant with $P < 0.001$

[b]significant with $P < 0.05$

$R^2 = 0.89$

Figure 5.3.2.1
Mean of ARR, Bias, and estimated ERR for the factor P

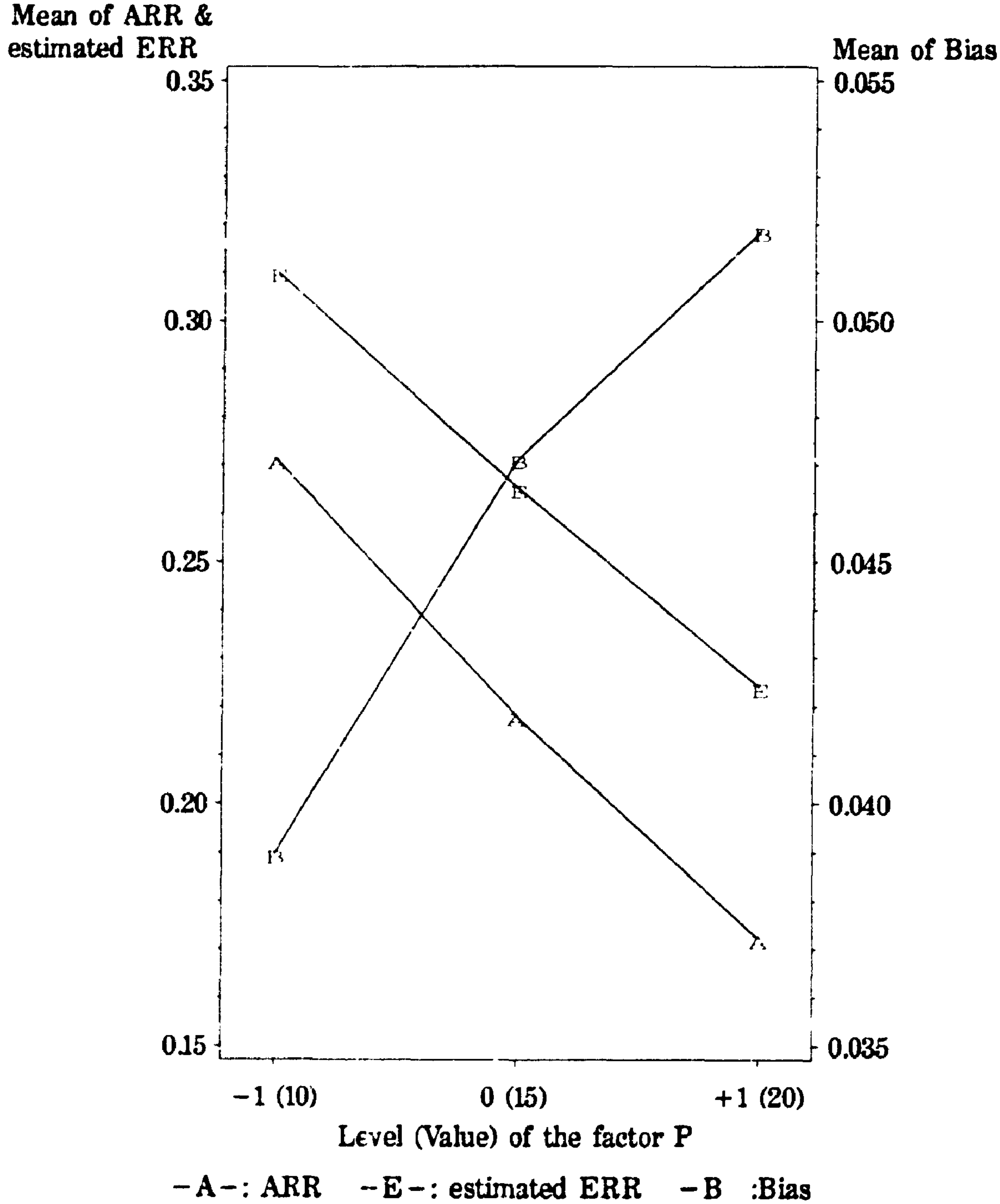


Figure 5.3.2.2

Mean of ARR, Bias, and estimated ERR for the factor B

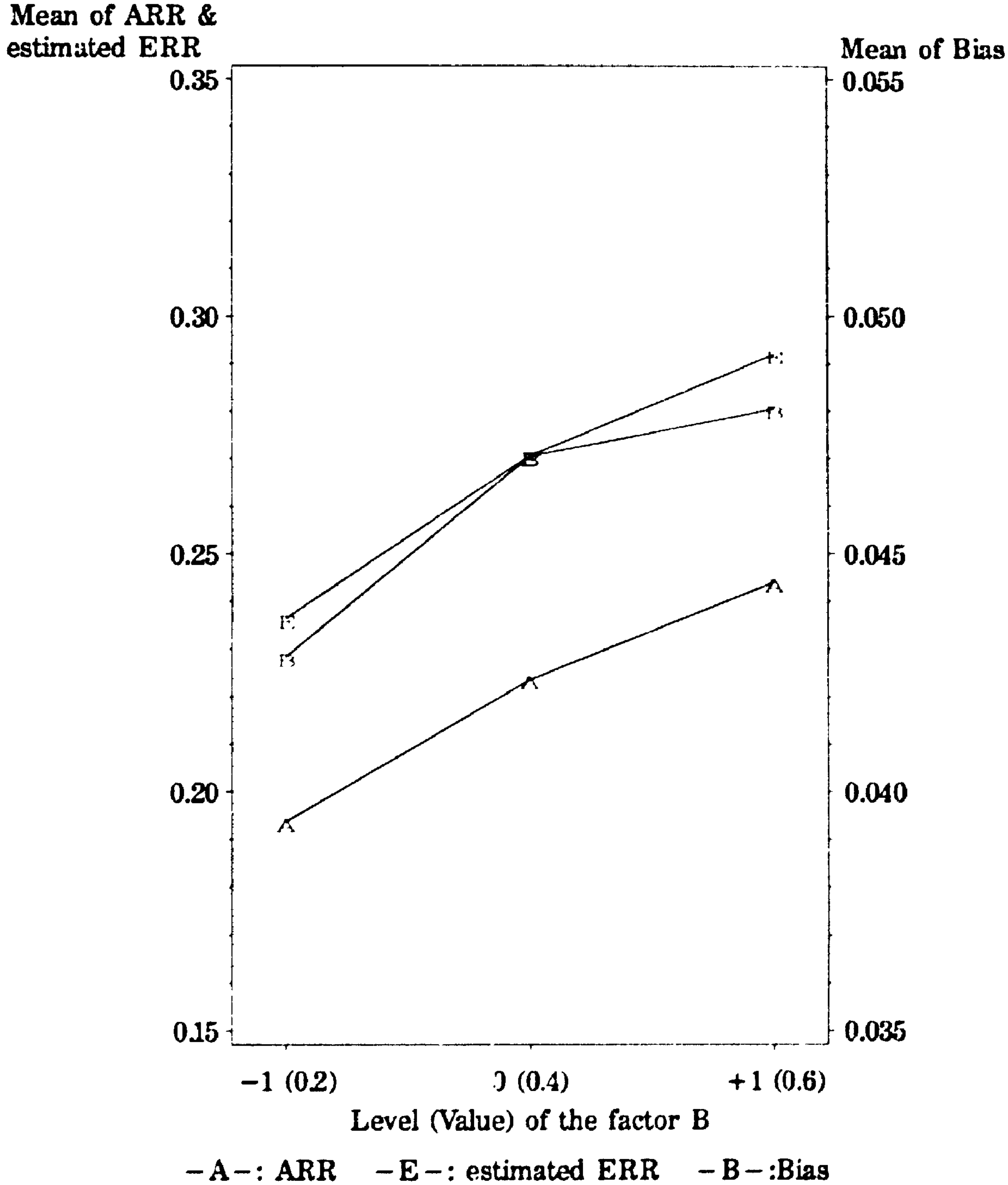


Figure 5.3.2.3

Mean of ARR, Bias, and estimated ERR for the factor M

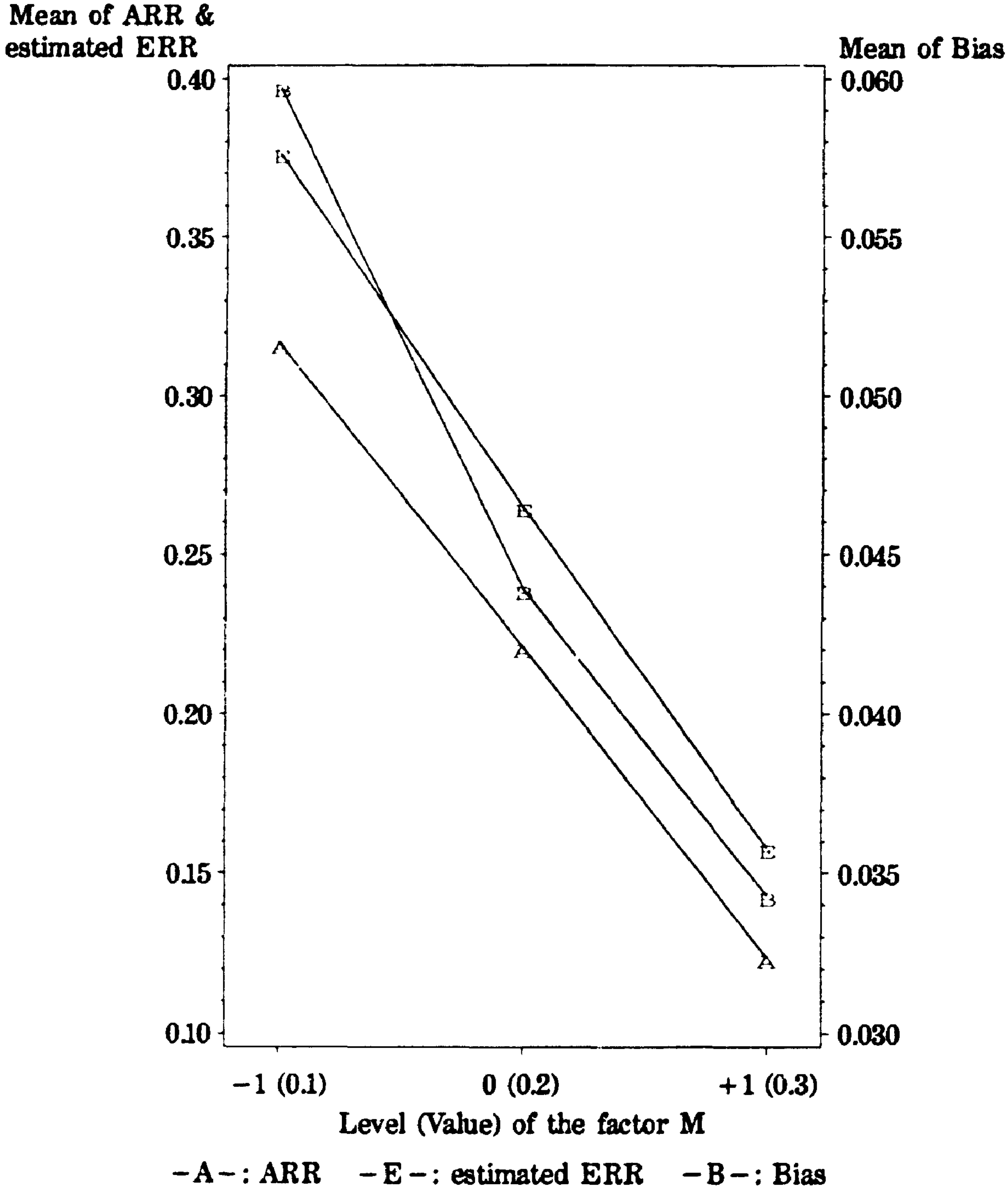


Figure 5.3.2.4

Mean of ARR, Bias, and estimated ERR for the factor N

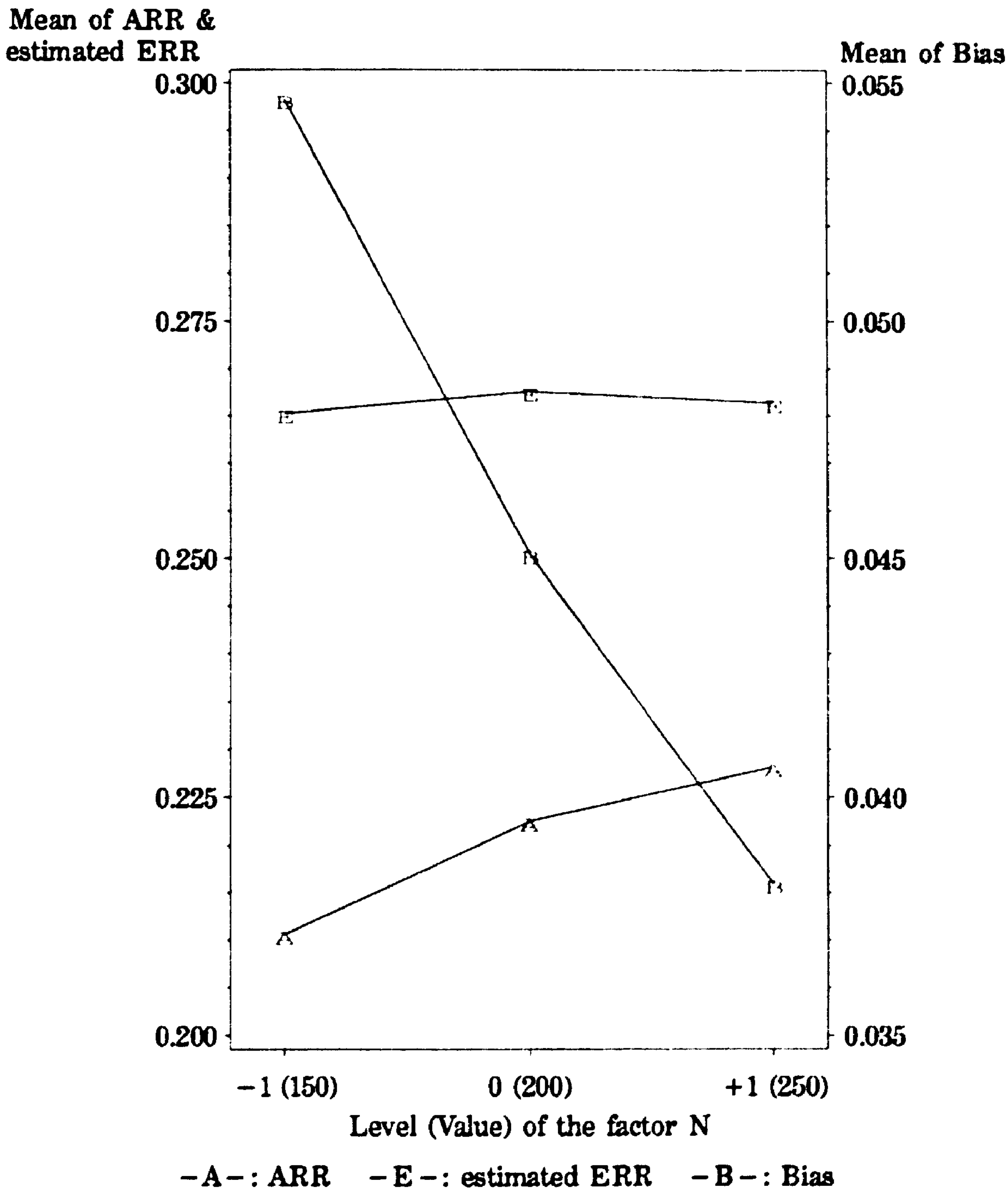


Figure 5.3.2.5

Effect of PM interaction on ARR

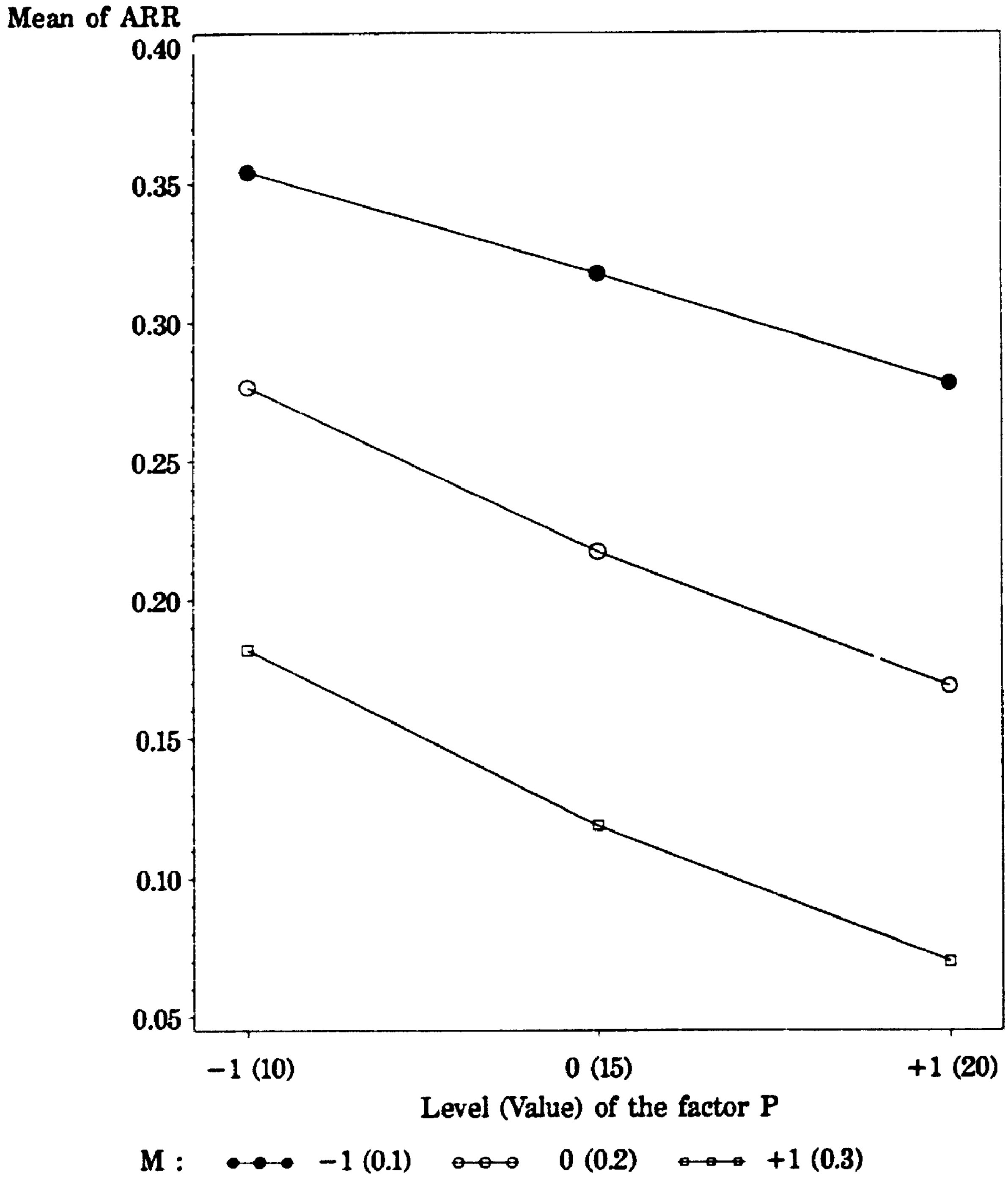


Figure 5.3.2.6

Effect of MN interaction on Bias

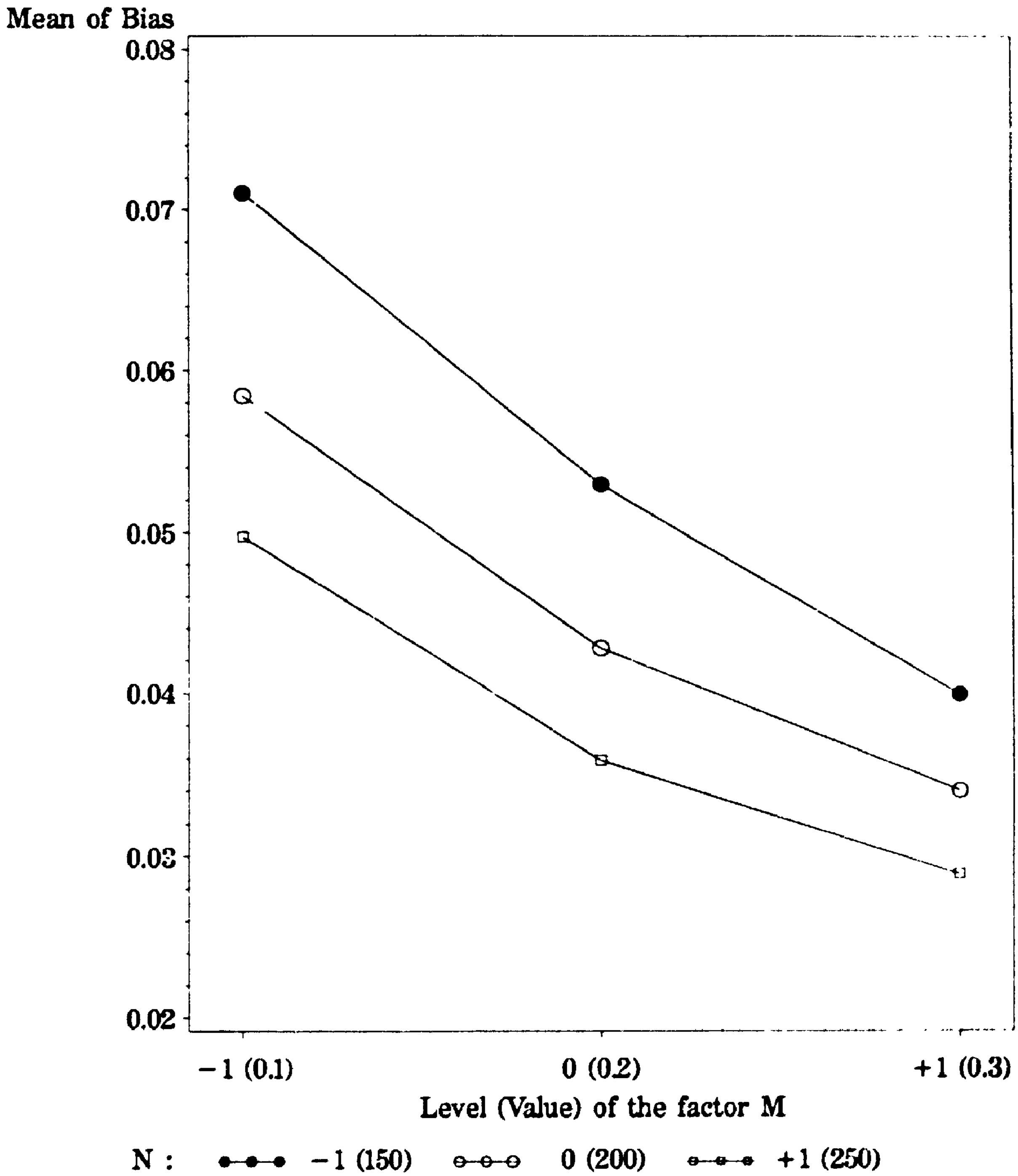
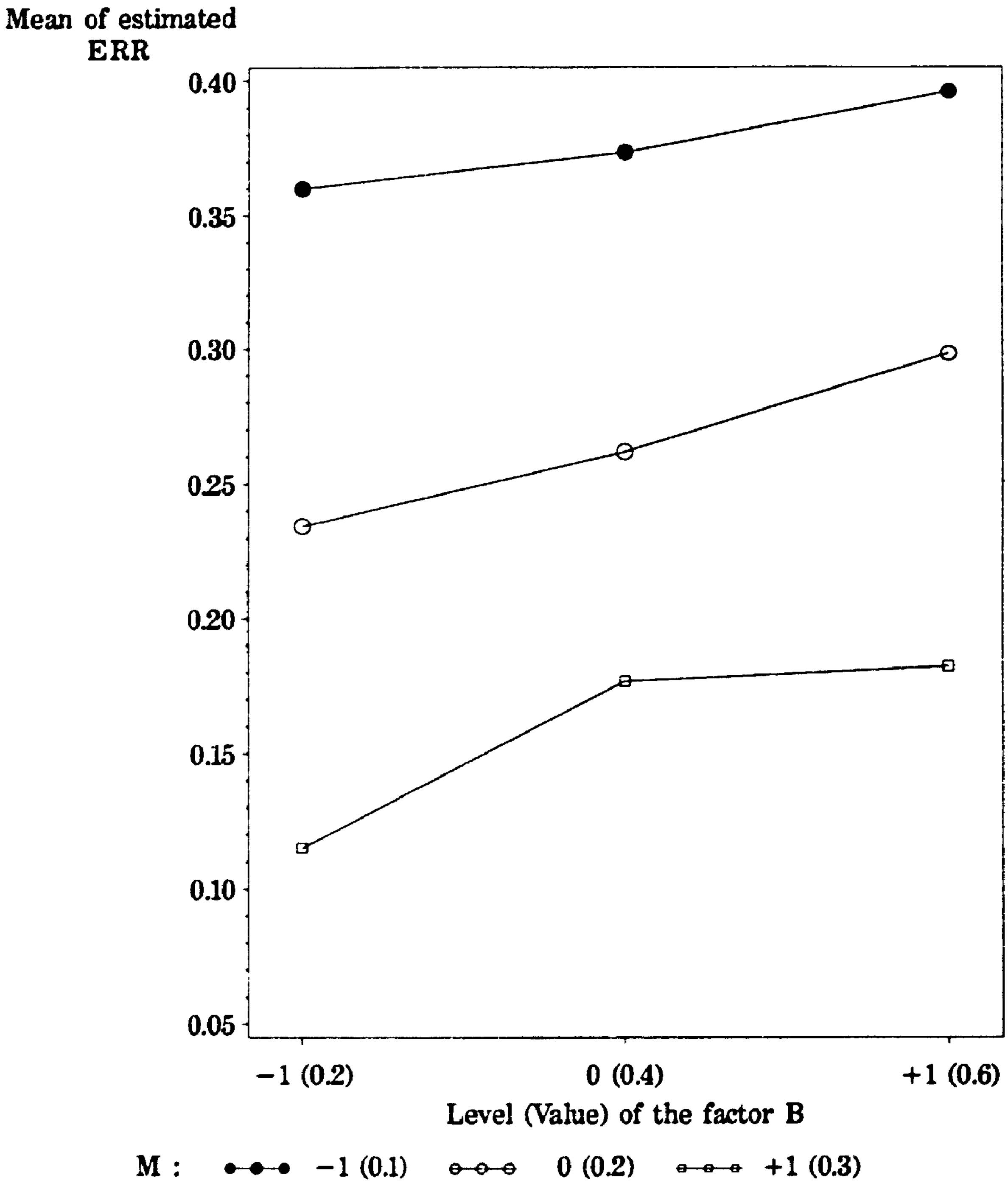


Figure 5.3.2.7
Effect of BM interaction on estimated ERR



5.4 Sizes of subset models determined by the stopping criteria

The five stopping criteria, $\chi^2_{(\alpha)}$, E_m, C_{pm}, AIC_m, and SCH_m, have been defined in section 3.3. Each of these stopping criteria determines the final number of predictor variables in subset models.

This section has two main purposes: 1) to compare the sizes of subset models determined by the five stopping criteria with the LR selection criterion in forward stepwise logistic regression; and 2) to understand the effects of the five factors P, V, Δ^2, D, and N in the multivariate normal case, and of the four factors P, B, M, and N in the multivariate binary case on the sizes of subset models for the five stopping criteria.

For convenience, let the symbol 'S' denotes the stopping criteria, and q and P refer to the sizes of subset and full models, respectively.

5.4.1 Multivariate normal case

In order to assess the effects of the six factors P, V, Δ^2, D, N, and S on q, a repeated measures analysis of variance was employed for the 32 factorial sampling situations. The reason for restricting attention to the 32 factorial sampling situations is that they comprise a fully balanced 2^5 factorial design in which the five factors appear at all combinations of the low and high (-1 and +1) factorial levels, making a repeated measures analysis of variance possible.

In the repeated measures analysis of variance (ANOVA) terminology, the five factors P, V, Δ^2, D, and N are often called 'between-suoject factors', and the five stopping criteria $\chi^2_{(0.20)}$, E_m, C_{pm}, AIC_m, and SCH_m are viewed as constituting 5 levels of the S 'within-subject factor'. In this study, all of six factors P, V, M, D, N and S are assumed to be fixed factors.

The factorial levels (-1 and +1) of the five between-subject factors determine 32 unique cells, and 100 observations are nested within each cell. They are grouped in 20 observations of 5 'repeated measures': namely, 20 replications of q for the 5 levels of the factor S. For a given replication, the 5 measures are obtained from the same sample data and hence are correlated; this is the reason a repeated measures analysis must be used.

In a repeated measures ANOVA, the total sum of squares (SS_{total}) is partitioned into two basic parts; one part is sum of squares for between-subject factors (SS_b) and a second part is sum of squares for within-subject factors (SS_w).

Depending on the complexity of the design, one or both of these partitions may then be further partitioned. For this study, SS_b can be further partitioned into two parts; one of these parts is the main effects and all interactions of the five factors P, V, M, D, and N, and the other part is sum of squares for the error term which is sometimes referred to as $Error_{between}$. Thus the partition of SS_b can be seen as essentially a separate analysis of variance, with its own error term $Error_{between}$. In other words, we have the usual analysis of variance for the between-subject factors by collapsing the levels of the factor S.

SS_w can be partitioned into three parts; the sum of squares for the main effect S, the sum of squares for the interactions of S and between-subject terms, and the sum of squares for the error term which is sometimes referred to as $Error_{within}$.

The two error terms, $Error_{between}$ and $Error_{within}$, are used separately to test the significance of the main effects and interactions of the six factors. Specifically, $Error_{between}$ is used as the denominator in F-ratio tests involving any of the main or interaction effects of the five factors P, V, M, D, and N; whereas $Error_{within}$ is used as the denominator in F-ratio tests involving the main effect of the factor S, as well as the interactions of the factor S with the other five factors.

One of the assumptions required for any F-ratio to be distributed as the central F is that of compound symmetry of the variance-covariance matrix. The conditions of compound symmetry are that the measures have the same variance and the correlation between the measurements for any two levels of the within factor is equal to the correlation between any other two levels. Compound symmetry is a sufficient but not necessary condition. A less restrictive condition is the sphericity assumption. The sphericity assumption is satisfied if the variance of the difference between any two repeated measures is the same for any two levels of the within factor. Compound symmetry is one of many variance-covariance structures which satisfy this assumption.

A test on the sphericity assumption has been developed by Mauchly (1940) and GLM procedure in SAS uses this test. The sphericity test lacks power for small sample sizes, and may reject for minor departures from sphericity when N is large. Box (1954) developed a measure ε, of the degree to which the variance-covariance matrix departs from the sphericity assumption. If the sphericity assumption is violated, an adjustment to numerator and denominator degrees of freedom in F-ratio cab be used. Two such adjustments (methods of Greenhouse and Geisser (1959) and Huynh and Feldt (1976)) based on Box's ε are provided in GLM procedure in SAS. Since the sphericity assumption was violated in this study, Greenhouse and Geisser's method, which is more conservative than Huynh and Feldt's method, was used. Of course, this 'adjustment' is only applicable to the tests for the effects which include the factor S. However, very large values of F-ratio are significant regardless of the adjustments of degrees of freedom.

Now refer to table 5.4.1.1 which gives the results of the repeated measures analysis of variance of q. As mentioned previously, the results are separated into two parts. In the first page of table 5.4.1.1 are the results for the effects which do not include the factor S (test of between-subject effect), and in the second page are the results for the effects which include the factor S (test of within-subject effect).

The first page of table 5.4.1.1 shows that the main effects for the factors P, V, Δ^2, and D are highly significant (F-ratio = 152.40, 206.30, 39.74, and 65.18, respectively). In fact, only the factor N is not significant. The magnitudes of the F-ratio for the interaction terms except for the VD interaction (F-ratio = 122.31), though some of them are statistically significant, are quite small compared to those of the F-ratio for the main effects. However, the results for the effects of between-subject factors described in the previous paragraph do not provide any indications of possible differences between the individual stopping criterion. They merely indicate the effects of the five factors P, V, M, D, and N on q for the 'overall' five stopping criteria.

On the other hand, the results for the effects which include the factor S reflect differences in q among the five stopping criteria. The second page of table 5.4.1.1 shows that the effect of the factor S is highly significant (F-ratio = 506.39). This provides a strong evidence of overall differences in q among the five stopping criteria.

Table 5.4.1.2 gives the mean of q for the five stopping criteria over the levels of the five factors P, V, Δ^2, D, and N. Figures 5.4.1.1 through 5.4.1.5 depict the effect of the five factors P, V, Δ^2, D, and N on q, respectively, for the five stopping criteria $\chi^2_{(0.20)}$, E_m, C_{pm}, AIC_m, and SCH_m. These figures illustrate the previous findings in the repeated ANOVA in table 5.4.1.1.

A response surface analysis of q was employed to assess the effects of the five factors P, V, Δ^2, D, and N for each of the five stopping criteria $\chi^2_{(0.20)}$, E_m, C_{pm}, AIC_m, and SCH_m.

Table 5.4.1.3 gives the results for the response surface analysis of q for the $\chi^2_{(0.20)}$ stopping criterion. Since the lack-of-fit test is not significant, the quadratic surface fits the data well. It is observed that variables P and V are highly significant (t = 10.47 and 15.12, respectively) and that variables M and D are moderately significant (t = 3.90 and 4.43, respectively). It is also observed that the V^2 term is highly significant (t = 7.10) which means non-linear effect on q in variable V (see figure 5.4.1.2). The three interaction terms PD, VD, and MD are statistically significant (t = -3.78, 7.63, and 2.66, respectively). In order to examine graphically the behaviors of these interactions, three graphs for PD, VD, and MD are presented in figures 5.4.1.6 through 5.4.1.8, respectively. These figures show that PD and MD are non-crossover interactions, while VD is a crossover interaction. This time all three graphs were presented to give a general idea on the shape of 'significant' interactions. However, in the subsequent discussions of the other stopping criteria, only the crossover or the most significant interaction will be presented. Figure 5.4.1.7 for the VD interaction shows that at the small value (0.4) of D q does not depend on V, while at the large value (0.8) of D q depends on V. As will be seen in the results of the other stopping criteria, essentially the same phenomena for the VD interactions are observed in the cases of the E_m, C_{pm}, AIC_m, and SCH_m stopping criteria.

Table 5.4.1.4 gives the results for the response surface analysis of q for the E_m stopping criterion. The lack-of-fit test is not significant which indicates the quadratic surface fits the data well. It is observed that variable P is highly significant (t = 11.00) and that variables V and D are moderately significant (t = 5.47 and 3.60, respectively). It is also noted that the V^2 term is moderately significant (t = 2.23). The two interaction terms VD and VN are moderately significant (t = 3.25 and 2.49, respectively). The graph for the VD interaction is presented in figure 5.4.1.9

and the same comment can be made as for figure 5.4.1.7 in the case of the $\chi^2_{(0.20)}$ stopping criterion.

Table 5.4.1.5 gives the results for the response surface analysis of q for the C_{pm} stopping criterion. The lack-of-fit test is significant which means that the quadratic surface does not fit the data very well. Hence remedial measures, perhaps involving transformation of dependent variable or of one or more of independent variables or possibly the use of a radically different model, would be taken. However, as indicated in Box and Draper (1987, p.74), when there is a very large amount of data, the deficient model may be used with proper caution: the magnitude as well as the significance of t-values will be considered in the following discussions of the results. The results of the analysis of variance of q for a 2^5 factorial design are given in table 5.4.1.8 to assist the validation of the results from the response surface analysis. We note that the value of R^2 is only 0.31 in table 5.4.1.8, which indicates that this ANOVA does not fit the data well. The results from both analyses are identical except that the response surface analysis picked the PD interaction as significant, while the analysis of variance did not. However, the PD interaction in the response surface analysis is weakly significant ($p = 0.034$), while the PD interaction in the analysis of variance is not significant. The most striking result in table 5.4.1.5 is the fact that although P is not significant, $\hat{\beta}$ is negative. This means that as P increases C_{pm} tends to pick up smaller subset models (see figure 5.4.1.1). The other main findings are: 1) variables V, M, and D are highly significant (t = 11.27, 6.49, and 7.77, respectively); 2) the V^2 term is highly significant ($t = 6.03$) and the D^2 term is moderately significant ($t = 2.81$); and 3) there are many interaction terms which are statistically significant, but only the VD interaction is highly significant ($t = 11.00$). The graph for the VD interaction is presented in figure 5.4.1.10 and the same comment can be made as previously.

Table 5.4.1.6 gives the results for the response surface analysis of q for the AIC_m stopping criterion. The lack-of-fit test is not significant which means that the quadratic surface fits the data well. It is very interesting to see the results in table 5.1.4.6 are almost the same as those in table 5.1.4.3. This is due to the fact that, as will be seen from the results of Bonferroni's multiple comparisons among the five stopping criteria, the difference of the overall sizes of subset models for the $\chi^2_{(0.20)}$ and AIC_m stopping criteria are not statistically significant. For this reason, we do not discuss more in detail for the AIC_m stopping criterion except that the graph for the VD interaction is presented in figure 5.4.1.11 and the same comment can be made as previously.

Table 5.4.1.7 gives the results for the response surface analysis of q for the SCH_m stopping criterion. The lack-of-fit test is significant which means that the quadratic surface does not fit the data very well. As for the same purpose mentioned in table 5.1.4.5, the results of the analysis of variance of q for a 2^5 factorial design are given in table 5.4.1.9. We note that the value of R^2 is 0.71 in table 5.4.1.9, which indicates that this ANOVA fairly fits the data. The results from both analyses are identical except that the response surface analysis indicates that variable P is significant (t = -3.75), while the analysis of variance does not. However, the trends are the same in that the sign of β is negative in the response surface analysis. Figure 5.4.1.1 shows that the mean of q goes down as P increases. The negative sign of β was also found in the case of C_{pm}, although the latter was not significant. It is noted that variables V and M are highly significant (t = 25.52 and 15.14, respectively) and that variables M and D are moderately significant (t = 6.06 and 7.79, respectively). It is also noted that the V^2 term is highly significant (t = 17.70). Even though there are many statistically significant interaction terms, only the VD interaction is highly significant (t = 22.73). The graph for the VD interaction is presented in figure 5.4.1.12. The same comment can be made as previously. It is interesting to see that the VD interactions are the most significant among all interaction terms for all five stopping criteria.

Table 5.4.1.2 gives a summary of the results for the mean of q for the five stopping criteria by level of the five factors P, V, Δ^2, D, and N. The overall mean of q for all 48 sampling situations for the five stopping criteria $\chi^2_{(0.20)}$, E_m, C_{pm}, AIC_m, and SCH_m are 4.7, 5.6, 3.6, 4.8, and 1.8, respectively. The value $\alpha = 0.20$ for the $\chi^2_{(\alpha)}$ stopping criterion was used as discussed in section 5.1.1. The overall size of subset models for the five stopping criteria may be written in descending order as $E_m > AIC_m > \chi^2_{(0.20)} > C_{pm} > SCH_m$. Pairwise comparisons of the overall mean of q were obtained by using Bonferroni method of multiple-comparisons procedure. The minimum significant difference, MSD, which any pair of means must differ by is:

$$t_{(\alpha,k,v)}\left[MSE\left(\frac{1}{N_1}+\frac{1}{N_2}\right)\right]^{1/2}$$

where

α is the overall significance level,

k is the number of pairwise comparisons,

v is the degrees of freedom for the error σ^2,

$t_{(\alpha,k,v)}$ is the upper $\frac{\alpha}{2k}$ point of the t distribution with v degrees of freedom,

MSE is the estimate for the error σ^2 and

N_1 and N_2 are the number of observations involved in the means which are being compared.

In our case, the minimum significant difference, MSD, was 0.282 with $\alpha=0.05$, $k=10$, $v=4795$, $t_{(0.05,10,4795)}=2.81$, $MSE=4.838511$, and $N_1=N_2=960$. All pairwise comparisons except for the pair of the $\chi^2_{(0.20)}$ and AIC_m stopping criteria are statistically significant at $\alpha = 0.05$ level. This result means that minimizing the AIC is almost equivalent to the $\chi^2_{(0.20)}$ stopping criterion in the size of subsets. The E_m stopping criterion tends to pick the largest size of subset models. The C_{pm} and SCH_m stopping criteria tend to pick very small size of subset models.

Finally, table 5.4.1.10 gives the mean of absolute size ($\bar{q}$) and proportional size ($\frac{\bar{q}}{P}$) of the model over the α level of significance for different values of P. The LR selection criterion was used with the $\chi^2_{(\alpha)}$ stopping criterion. The results are also presented in figure 5.4.1.13 which exhibits two interesting results. First, the smaller the value of P, the larger is the proportion of P which is significant at $\alpha = 0.05$. Second, the smaller the model, that is the smaller the value of P, the more quickly the procedure selects a full model. The homogeneity of these five curves was tested by the log-rank test. The log-rank test is highly significant ($\chi^2_4 = 34.08$ and $p < 0.0001$).

5.4.2 Multivariate binary case

The tables and graphs for the multivariate binary case are analogous to those in the multivariate normal case.

In the repeated ANOVA, the four factors P, B,M, and N are between-subject factors, and the five stopping criteria $\chi^2_{(0.15)}$, E_m, C_{pm}, AIC_m, and SCH_m are five levels of the S within-subject factor. In this study, all of the five factors P, B, M, N, and S are assumed to be fixed factors.

Table 5.4.2.1 gives the results of the repeated measures analysis of variance of q for the five factors P, B, M, N, and S. The results are separated into two parts: in the first page of the table are the results for the effects which do not include the factor S (test of between-subject effect), while in the second page of the table are the results for the effects which include the factor S (test of within-subject effect).

Consider first the results for the effects which do not include the factor S. Table 5.4.2.1 shows that the main effects for the factors P, B, M, and N are highly significant (F-ratio = 2350.94, 313.44, 2643.79, and 183.12, respectively), although the magnitudes for the F-ratio for the factors B and N are much smaller than those for the factors P and M. Only the PM interaction is moderately significant (F-ratio = 83.55) and the magnitudes of the rest of interactions are quite small compared to those of the main effects.

The second page of table 5.4.2.1 shows the results for the effects which include the factor S. The very large F-ratio (4921.70) for the effect of the factor S provides a strong evidence of overall differences in q among the five stopping criteria. It is noticed that all interaction terms are statistically significant, however only the SP interaction has a large F-ratio (388.83).

Figures 5.4.2.1 through 5.4.1.4 illustrate graphically the previous findings in the repeated ANOVA in table 5.4.2.1. These figures depict the effects of the four factors P, B, M, and N on q, respectively, for the five stopping criteria $\chi^2_{(0.15)}$, E_m, C_{pm}, AIC_m, and SCH_m. Comparing figure 5.4.2.1 with figure 5.4.1.1, one result can be noticed; that is, for the C_{pm} and SCH_m stopping criteria the mean of q increases as P increases in the multivariate binary case, while the opposite phenomenon was observed in the multivariate normal case.

An analysis of variance of q was employed to assess the effects of the four factors P, B, M, and N for the individual stopping criterion. The results of the analysis of variance of q for the five stopping criteria $\chi^2_{(0.15)}$, E_m, AIC_m, C_{pm}, and SCH_m are presented in tables 5.4.2.3 through 5.4.2.7, respectively. We note that the values of R^2 for tables 5.4.2.3 through 5.4.2.7 are 0.84, 0.66, 0.85, 0.57, and 0.85, respectively, which indicate that the ANOVAs for $\chi^2_{(0.15)}$, AIC_m, and SCH_m fit the data well and that the ANOVAs for E_m and C_{pm} do not fit the data well.

Due to many similarities of the results of the analysis of variances from the five stopping criteria, we do not discuss each stopping criterion separately. The results for the three stopping criteria, namely $\chi^2_{(0.15)}$, E_m, and AIC_m, are very similar, whereas the results for the other two stopping criteria, namely C_{pm} and SCH_m, are very similar.

Tables 5.4.2.3 through 5.4.2.5 for the $\chi^2_{(0.15)}$, E_m, and AIC_m stopping criteria, respectively, show that all four main effects, namely P, B, M, and N, are highly significant. The magnitudes of the F-ratio for the four main effects may be written in descending order as P > M > B > N. This means that P is the most influential factor on the sizes of subset models for these three stopping criteria. It is noted that the magnitudes of the F-ratio for all interaction terms are negligible compared to those of for the main effect terms. For instance, for the $\chi^2_{(0.15)}$ stopping criterion in table 5.4.2.3 the magnitude of the F-ratio for the PM interaction which is most significant among all interaction terms is much smaller than those for the main effect terms (21.65, 1969.58, and 1612.60 for PM, P, and M, respectively). For this reason, two most significant interaction terms, namely PM and BM, for the $\chi^2_{(0.15)}$, E_m, and AIC_m stopping criteria are presented in figures 5.4.2.5 through 5.4.2.10, respectively, to illustrate the general shape of all interactions. These figures show that the PM and BM interactions are slightly non-parallel.

Tables 5.4.2.6 and 5.4.2.7 for the C_{pm} and SCH_m stopping criteria, respectively show that all four main effects, namely P, B, M, and N, are statistically significant. However, the magnitudes of the F-ratio for the three factors P, B, and N are much smaller than that for the factor M. For instance, for the C_{pm} stopping criterion the F-ratios are 22.45, 14.65, 708.80, and 15.41 for the four factors P, B, M, and N, respectively. The magnitudes of the F-ratio for the four main effects may be written in descending order as M > P > N > B for the C_{pm} stopping criterion, and as M > P > B > N for the SCH_m stopping criterion. This means that M is the most influential factor on the sizes of subset models for these two stopping criteria. This result is different from that in the cases of the $\chi^2_{(0.15)}$, E_m, and AIC_m stopping criteria; P was the most influential factor for those three stopping criteria. It is noted that the magnitudes of the F-ratio for all interaction terms are very small compared to those of for the main effect terms. For this reason, two most significant interaction terms for the C_{pm} and SCH_m stopping criteria, namely PM and BM, are presented in figures 5.4.2.11 through 5.4.2.14, respectively. These figures show that the PM and BM interactions for the C_{pm} stopping criterion are slightly crossed and the PM and MB interactions for the SCH_m stopping criterion are merely non-parallel.

Table 5.4.2.2 gives a summary of the results for the mean of q for the five stopping criteria by level of the four factors P, B, M, and N. The overall mean of q for all 81 sampling situations for the five stopping criteria $\chi^2_{(0.15)}$, E_m, C_{pm}, AIC_m, and SCH_m are 9.2, 9.0, 4.1, 9.3, and 3.4, respectively. The value $\alpha = 0.15$ for the $\chi^2_{(\alpha)}$ stopping criterion was used as discussed in section 5.1.2. The overall size of subset models for the five stopping criteria may be written in descending order as $AIC_m > \chi^2_{(0.15)} > E_m > C_{pm} > SCH_m$. A striking result from this order is the change of order for the E_m stopping criterion; E_m was first in the order of stopping criterion in the multivariate normal case. Pairwise comparisons of the overall mean of q were obtained by using Bonferroni method of multiple-comparisons procedure. The minimum significant difference, MSD, which any pair of means must differ by was 0.3562 with α=0.05, k=10, v=8095, $t_{(0.05,10,8095)}$=2.81, MSE=13.03625, and N_1=N_2=1620. The results of Bonferroni multiple comparisons show that the five stopping criteria make up three groups; $\{AIC_m, \chi^2_{(0.15)}, E_m\}$, $\{C_{pm}\}$, and $\{SCH_m\}$.

Any pair within a group are not statistically significant, while any pair between groups are statistically significant.

Table 5.4.2.1 gives the mean of absolute size $(\bar{q})$ and proportional size $(\frac{\bar{q}}{P})$ of the model over the α level of significance for different value of P. The LR selection criterion was used with the $\chi^2_{(\alpha)}$ stopping criterion. The results are also presented in figure 5.4.2.15. This figure shows that: 1) the proportions of P which are significant at $\alpha = 0.05$ are almost equal for P = 10, 15, and 20; and 2) the smaller model tends to grow more quickly to its full model, but this is not statistically significant (Log-rank test: $\chi^2_2 = 4.00$, $p > 0.13$). These results are contrasted with those in the multivariate normal case.

5.4.3 Conclusions

1. For the multivariate normal case, the descending order for the size of subset models for the five stopping criteria is $E_m > AIC_m > \chi^2_{(0.20)} > C_{pm} > SCH_m$.
2. For the multivariate binary case, the descending order for the size of subset models for the five stopping criteria is $AIC_m > \chi^2_{(0.15)} > E_m > C_{pm} > SCH_m$.
3. For the multivariate normal case, V is the most significant factor for the $\chi^2_{(0.20)}$, C_{pm}, and SCH_m stopping criteria, whereas P is for the E_m and AIC_m stopping criteria.
4. For the multivariate binary case, P is the most significant factor for the $\chi^2_{(0.15)}$, E_m, and AIC_m stopping criteria; whereas M is for the C_{pm} and SCH_m stopping criteria.

Table 5.4.1.1
Repeated measures analysis of variance of q
Test of Between Subject Effect

Effect	d.f.	Mean Square	F-ratio	P-value
[a]P	1	996.81	152.40	0.00
[a]V	1	1349.40	206.30	0.00
[a]M	1	259.92	39.74	0.00
[a]D	1	426.32	65.18	0.00
N	1	8.00	1.22	0.27
PV	1	14.31	2.19	0.14
[b]PM	1	43.25	6.61	0.01
[a]PD	1	108.05	16.52	0.00
[b]PN	1	26.65	4.07	0.04
[b]VM	1	39.61	6.06	0.01
[a]VD	1	800.00	122.31	0.00
[b]VN	1	36.98	5.65	0.02
MD	1	17.11	2.62	0.11
[b]MN	1	26.28	4.02	0.05
DN	1	7.03	1.07	0.30
PVM	1	1.28	0.20	0.66
[b]PVD	1	28.13	4.30	0.04
PVN	1	0.05	0.01	0.93
PMD	1	0.10	0.02	0.90
[b]PMN	1	31.60	4.83	0.03
PDN	1	9.90	1.51	0.22
VMD	1	15.40	2.35	0.13
VMN	1	13.26	2.03	0.16
VDN	1	0.28	0.04	0.84
MDN	1	13.01	1.99	0.16
PVMD	1	13.78	2.11	0.15
PVMN	1	5.28	0.81	0.37
PVDN	1	9.46	1.45	0.23
PMDN	1	0.02	0.00	0.96
VMDN	1	23.12	3.53	0.06
PVMDN	1	12.01	1.84	0.18
Error(between)	608	6.54		

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

Table 5.4.1.1 (continued) Repeated measures analysis of variance of q Test of Within Subject Effect				
Effect	d.f.	Mean Square	F-ratio	P–value[c]
[a]S	4	1408.77	506.39	0.00
[a]SP	4	173.56	62.39	0.00
[a]SV	4	21.98	7.90	0.00
SM	4	6.52	2.34	0.08
SD	4	3.46	1.24	0.29
SN	4	3.21	1.16	0.32
SPV	4	2.56	0.92	0.42
SPM	4	4.83	1.74	0.17
SPD	4	2.43	0.87	0.44
SPN	4	2.46	0.88	0.43
SVM	4	2.40	0.86	0.44
SVD	4	5.37	1.93	0.13
[b]SVN	4	7.78	2.80	0.05
SMD	4	4.55	1.64	0.19
SMN	4	4.78	1.72	0.17
SDN	4	1.43	0.51	0.64
SPVM	4	0.38	0.14	0.91
SPVD	4	4.64	1.67	0.18
SPVN	4	6.33	2.28	0.09
SPMD	4	2.33	0.84	0.45
SPMN	4	3.04	1.09	0.34
SPDN	4	3.93	1.41	0.24
SVMD	4	7.48	2.69	0.06
SVMN	4	1.46	0.53	0.63
SVDN	4	1.17	0.42	0.70
SMDN	4	2.32	0.83	0.46
SPVMD	4	3.61	1.30	0.27
SPVMN	4	0.67	0.24	0.83
SPVDN	4	1.47	0.53	0.63
SPMDN	4	1.37	0.49	0.65
SVMDN	4	1.56	0.56	0.61
SPVMDN	4	2.20	0.79	0.48
Error(within)	2432	2.78		

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

[c]Greenhouse-Geisser method

Table 5.4.1.2
[a]Mean of q for the five stopping criteria
over the levels of the five factors P, V, Δ^2, D, and N

Factor	[b]Level	Value	Stopping criteria				
			$\chi^2_{(0.20)}$	E_m	C_{pm}	AIC_m	SCH_m
P	-2	5	3.8	3.3	3.9	3.2	2.0
	-1	10	4.1	4.5	3.6	4.0	1.9
	0	15	4.4	5.6	3.3	4.6	1.6
	+1	20	5.5	6.8	3.8	5.8	1.8
	+2	25	5.3	8.4	2.7	6.5	1.2
V	-2	0.2	4.1	5.2	3.5	4.3	1.9
	-1	0.4	3.9	5.1	3.1	4.2	1.4
	0	0.6	4.2	5.5	3.1	4.4	1.4
	+1	0.8	5.8	6.2	4.3	5.6	2.3
	+2	1.0	8.0	8.0	5.1	7.6	3.6
Δ^2	-2	1.0	3.6	5.3	2.7	3.3	1.2
	-1	1.5	4.6	5.4	3.3	4.7	1.7
	0	2.0	4.5	5.7	3.3	4.7	1.6
	+1	2.5	5.1	6.0	4.1	5.3	2.0
	+2	3.0	4.1	4.5	3.5	5.4	1.3
D	-2	0.2	4.1	4.8	2.6	3.9	1.4
	-1	0.4	4.5	5.2	3.3	4.5	1.6
	0	0.6	4.4	5.7	3.2	4.7	1.6
	+1	0.8	5.1	6.1	4.1	5.3	2.1
	+2	1.0	4.7	5.1	4.5	4.8	2.2
N	-2	100	3.8	6.3	3.0	4.9	1.3
	-1	150	4.8	5.6	3.6	4.9	1.7
	0	200	4.4	5.5	3.3	4.6	1.6
	+1	250	4.9	5.7	3.8	4.8	2.0
	+2	300	4.7	6.0	3.3	4.4	1.6
Overall			4.7	5.6	3.6	4.8	1.8

[a]The standard error of mean is in the range of 0.028 and 0.818.
[b]The number of sampling situations for each level is 1 for levels ±2, 16 for levels ±1, and 14 for level 0.

Table 5.4.1.3
Response surface analysis of q
for the $\chi^2_{(0.20)}$ stopping criterion

Variable	d.f	$\hat{\beta}$	s.e	t-value	p-value
[a]P	1	0.654	0.062	10.47	0.000
[a]V	1	0.944	0.062	15.12	0.000
[a]M	1	0.244	0.062	3.90	0.000
[a]D	1	0.276	0.062	4.43	0.000
N	1	0.079	0.062	1.26	0.208
P^2	1	0.140	0.072	1.95	0.051
[a]V^2	1	0.509	0.072	7.10	0.000
M^2	1	-0.035	0.072	-0.49	0.623
D^2	1	0.102	0.072	1.43	0.154
N^2	1	0.059	0.072	0.82	0.414
PV	1	0.042	0.070	0.60	0.546
PM	1	0.095	0.070	1.37	0.172
[a]PD	1	-0.264	0.070	-3.78	0.000
PN	1	-0.039	0.070	-0.56	0.576
VM	1	0.123	0.070	1.77	0.077
[a]VD	1	0.533	0.070	7.63	0.000
VN	1	0.002	0.070	0.02	0.982
[b]MD	1	0.186	0.070	2.66	0.008
MN	1	-0.095	0.070	-1.37	0.172
DN	1	0.039	0.070	0.56	0.576

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

Lack-of-fit test: $F_{22,917} = 1.240$ ($p > 0.205$)

Table 5.4.1.4 Response surface analysis of q for the E_m stopping criterion					
Variable	d.f	$\hat{\beta}$	s.e	t-value	p-value
[a]P	1	1.163	0.106	11.00	0.000
[a]V	1	0.578	0.106	5.47	0.000
M	1	0.188	0.106	1.77	0.076
[a]D	1	0.380	0.106	3.60	0.000
N	1	0.028	0.106	0.26	0.795
P^2	1	0.088	0.121	0.73	0.466
[b]V^2	1	0.270	0.121	2.23	0.026
M^2	1	-0.155	0.121	-1.28	0.201
D^2	1	-0.149	0.121	-1.23	0.219
N^2	1	0.163	0.121	1.35	0.178
PV	1	-0.075	0.118	-0.64	0.526
PM	1	0.194	0.118	1.64	0.101
PD	1	-0.219	0.118	-1.85	0.064
PN	1	-0.116	0.118	-0.98	0.328
VM	1	0.209	0.118	1.77	0.077
[b]VD	1	0.384	0.118	3.25	0.001
[b]VN	1	0.294	0.118	2.49	0.013
MD	1	-0.041	0.118	-0.34	0.731
MN	1	-0.038	0.118	-0.32	0.751
DN	1	0.056	0.118	0.48	0.634

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

Lack-of-fit test: $F_{22,917} = 1.035$ ($p > 0.417$)

Table 5.4.1.5
Response surface analysis of q
for the C_{pm} stopping criterion

Variable	d.f	$\hat{\beta}$	s.e	t-value	p-value
P	1	-0.016	0.051	-0.32	0.749
[a]V	1	0.571	0.051	11.27	0.000
[a]M	1	0.329	0.051	6.49	0.000
[a]D	1	0.394	0.051	7.77	0.000
N	1	0.056	0.051	1.11	0.267
P^2	1	0.107	0.058	1.84	0.066
[a]V^2	1	0.351	0.058	6.03	0.000
M^2	1	0.051	0.058	0.87	0.384
[b]D^2	1	0.163	0.058	2.81	0.005
N^2	1	0.063	0.058	1.09	0.278
PV	1	-0.105	0.057	-1.85	0.065
[b]PM	1	0.148	0.057	2.62	0.009
[b]PD	1	-0.120	0.057	-2.12	0.034
[b]PN	1	-0.186	0.057	-3.28	0.001
VM	1	0.092	0.057	1.63	0.104
[a]VD	1	0.623	0.057	11.00	0.000
VN	1	0.070	0.057	1.24	0.215
MD	1	0.033	0.057	0.58	0.563
[b]MN	1	-0.183	0.057	-3.23	0.001
DN	1	0.098	0.057	1.74	0.083

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

Lack-of-fit test: $F_{22,917} = 2.419$ ($p < 0.003$)

Table 5.4.1.6
Response surface analysis of q
for the AIC_m stopping criterion

Variable	d.f	$\hat{\beta}$	s.e	t-value	p-value
[a]P	1	0.893	0.066	13.63	0.000
[a]V	1	0.728	0.066	11.11	0.000
[a]M	1	0.408	0.066	6.22	0.000
[a]D	1	0.373	0.066	5.69	0.000
N	1	-0.063	0.066	-0.95	0.340
P^2	1	0.120	0.075	1.60	0.109
[a]V^2	1	0.402	0.075	5.35	0.000
M^2	1	-0.011	0.075	-0.14	0.885
D^2	1	-0.005	0.075	-0.06	0.951
N^2	1	0.070	0.075	0.94	0.349
PV	1	-0.081	0.073	-1.11	0.268
[b]PM	1	0.169	0.073	2.30	0.021
[b]PD	1	-0.191	0.073	-2.60	0.009
PN	1	-0.078	0.073	-1.07	0.286
VM	1	0.047	0.073	0.64	0.522
[a]VD	1	0.519	0.073	7.08	0.000
VN	1	0.100	0.073	1.37	0.172
MD	1	0.106	0.073	1.45	0.147
[b]MN	1	-0.163	0.073	-2.22	0.027
DN	1	-0.028	0.073	-0.38	0.701

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

Lack-of-fit test: $F_{22,917} = 0.803$ ($p > 0.724$)

Table 5.4.1.7
Response surface analysis of q
for the SCH_m stopping criterion

Variable	d.f	$\hat{\beta}$	s.e	t-value	p-value
ᵃP	1	-0.065	0.017	-3.75	0.000
ᵃV	1	0.443	0.017	25.52	0.000
ᵃM	1	0.105	0.017	6.06	0.000
ᵃD	1	0.263	0.017	15.14	0.000
ᵃN	1	0.135	0.017	7.79	0.000
ᵇP^2	1	0.052	0.020	2.61	0.009
ᵃV^2	1	0.352	0.020	17.70	0.000
M^2	1	-0.023	0.020	-1.16	0.247
ᵃD^2	1	0.108	0.020	5.44	0.000
N^2	1	0.027	0.020	1.36	0.175
ᵃPV	1	-0.116	0.019	-5.96	0.000
PM	1	-0.025	0.019	-1.29	0.198
ᵃPD	1	-0.125	0.019	-6.45	0.000
PN	1	-0.038	0.019	-1.93	0.053
ᵃVM	1	0.084	0.019	4.35	0.000
ᵃVD	1	0.441	0.019	22.73	0.000
ᵃVN	1	0.072	0.019	3.71	0.000
ᵃMD	1	0.081	0.019	4.19	0.000
MN	1	0.025	0.019	1.29	0.198
ᵃDN	1	0.069	0.019	3.55	0.000

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

Lack-of-fit test: $F_{22,917} = 5.913$ ($p < 0.001$)

Table 5.4.1.8
Analysis of variance of q for the C_{pm} stopping criterion

Effect	d.f.	Mean Square	F-ratio	p-value
P	1	1.914	0.70	0.404
[a]V	1	238.877	87.16	0.000
[a]M	1	84.827	30.95	0.000
[a]D	1	87.764	32.02	0.000
N	1	1.914	0.70	0.404
PV	1	7.014	2.56	0.110
[b]PM	1	14.102	5.15	0.024
PD	1	9.264	3.38	0.067
[b]PN	1	22.127	8.07	0.005
VM	1	5.439	1.98	0.159
[a]VD	1	248.752	90.76	0.000
VN	1	3.164	1.15	0.283
MD	1	0.689	0.25	0.616
[b]MN	1	21.389	7.80	0.005
DN	1	6.202	2.26	0.133
Error	624	2.741		

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

$R^2 = 0.31$

Table 5.4.1.9
Analysis of variance of q for the SCH_{pm} stopping criterion

Effect	d.f.	Mean Square	F-ratio	p-value
P	1	0.625	2.62	0.106
[a]V	1	127.807	536.16	0.000
[a]M	1	10.000	41.95	0.000
[a]D	1	48.400	203.04	0.000
[a]N	1	14.400	60.41	0.000
[a]PV	1	8.556	35.89	0.000
PM	1	0.400	1.68	0.196
[a]PD	1	10.000	41.95	0.000
PN	1	0.900	3.78	0.053
[a]VM	1	4.556	19.11	0.000
[a]VD	1	124.256	521.27	0.000
[a]VN	1	3.306	13.87	0.000
[a]MD	1	4.225	17.72	0.000
MN	1	0.400	1.68	0.196
[a]DN	1	3.025	12.69	0.000
Error	624	0.238		

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

$R^2 = 0.71$

Table 5.4.1.10

The mean of absolute size ($\bar{q}$) and proportional size ($\frac{\bar{q}}{P}$) of the logistic model over the α level of significance for different values of P

P	5	10	15	20	25
α	$\bar{q}$ ($\frac{\bar{q}}{P}$)	$\bar{q}$ ($\frac{\bar{q}}{P}$)	$\bar{q}$ ($\frac{\bar{q}}{P}$)	$\bar{q}$ ($\frac{\bar{q}}{P}$)	$\bar{q}$ ($\frac{\bar{q}}{P}$)
0.05	2.8(0.56)	2.9(0.29)	2.5(0.17)	3.0(0.15)	2.3(0.09)
0.10	4.1(0.82)	3.3(0.33)	3.1(0.21)	3.9(0.20)	3.3(0.13)
0.15	5.0(1.00)	3.5(0.35)	3.9(0.26)	5.0(0.25)	4.7(0.19)
0.20	5.0(1.00)	4.0(0.40)	4.7(0.31)	6.2(0.31)	6.3(0.25)
0.25	5.0(1.00)	4.5(0.45)	5.5(0.36)	7.1(0.36)	7.4(0.29)
0.30	5.0(1.00)	5.5(0.55)	5.8(0.39)	8.0(0.40)	9.0(0.36)
0.35	5.0(1.00)	7.2(0.72)	6.1(0.41)	8.9(0.44)	10.4(0.41)
0.40	5.0(1.00)	9.6(0.96)	6.8(0.45)	9.6(0.48)	12.0(0.48)
0.45	5.0(1.00)	10.0(1.00)	7.5(0.50)	10.3(0.52)	12.6(0.50)
0.50	5.0(1.00)	10.0(1.00)	8.8(0.58)	10.8(0.54)	13.5(0.54)
0.55	5.0(1.00)	10.0(1.00)	10.6(0.71)	11.2(0.56)	13.9(0.56)
0.60	5.0(1.00)	10.0(1.00)	13.0(0.87)	12.1(0.61)	15.0(0.60)
0.65	5.0(1.00)	10.0(1.00)	14.7(0.98)	13.6(0.68)	15.2(0.61)
0.70	5.0(1.00)	10.0(1.00)	15.0(1.00)	15.1(0.76)	16.2(0.65)
0.75	5.0(1.00)	10.0(1.00)	15.0(1.00)	17.0(0.85)	20.2(0.81)
0.80	5.0(1.00)	10.0(1.00)	15.0(1.00)	18.7(0.94)	22.4(0.89)
0.85	5.0(1.00)	10.0(1.00)	15.0(1.00)	19.9(1.00)	23.4(0.93)
0.90	5.0(1.00)	10.0(1.00)	15.0(1.00)	20.0(1.00)	23.4(0.93)
0.95	5.0(1.00)	10.0(1.00)	15.0(1.00)	20.0(1.00)	25.0(1.00)

Figure 5.4.1.1

Mean of q for the five stopping criteria over the levels of the factor P

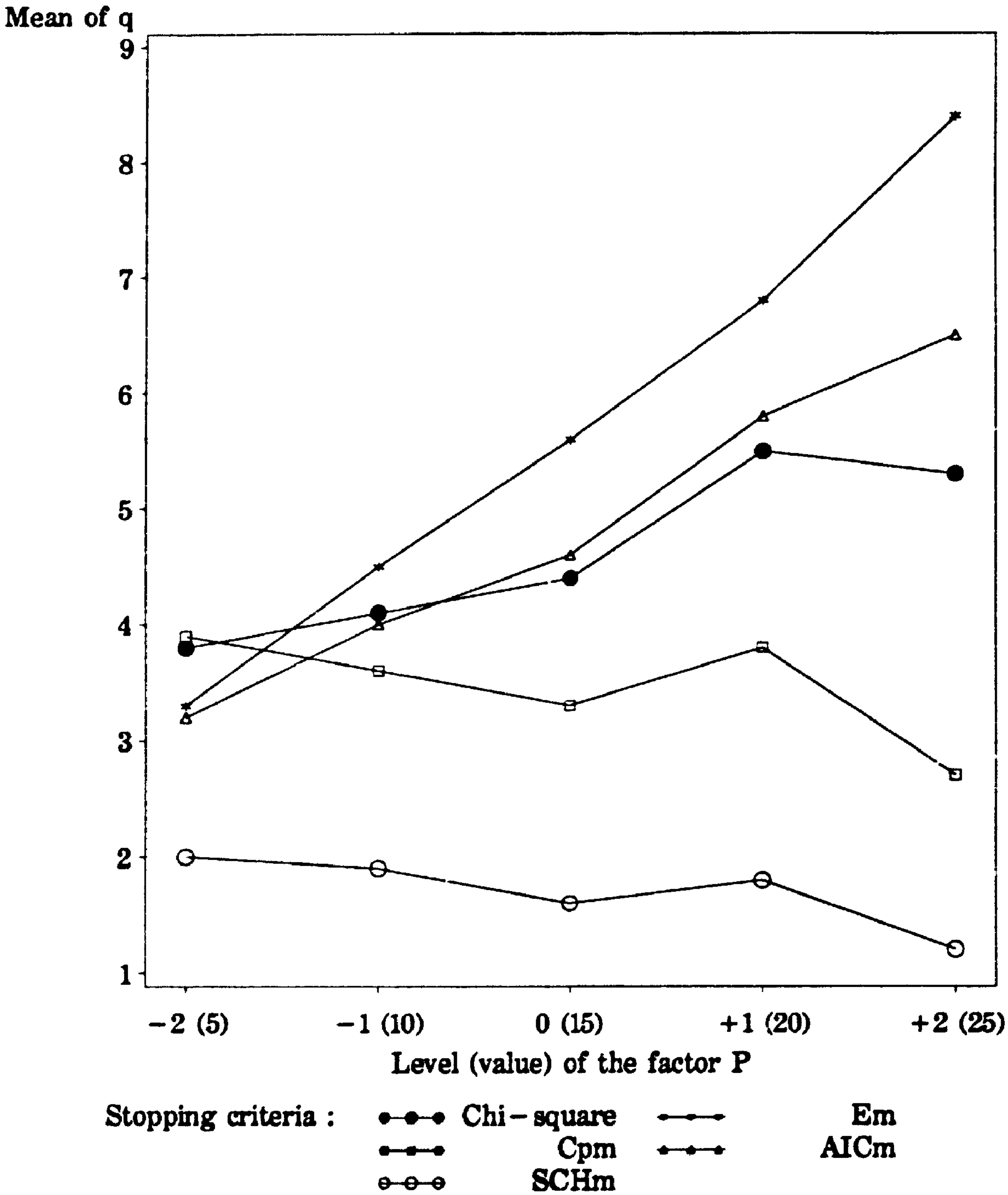


Figure 5.4.1.2

Mean of q for the five stopping criteria over the levels of the factor V

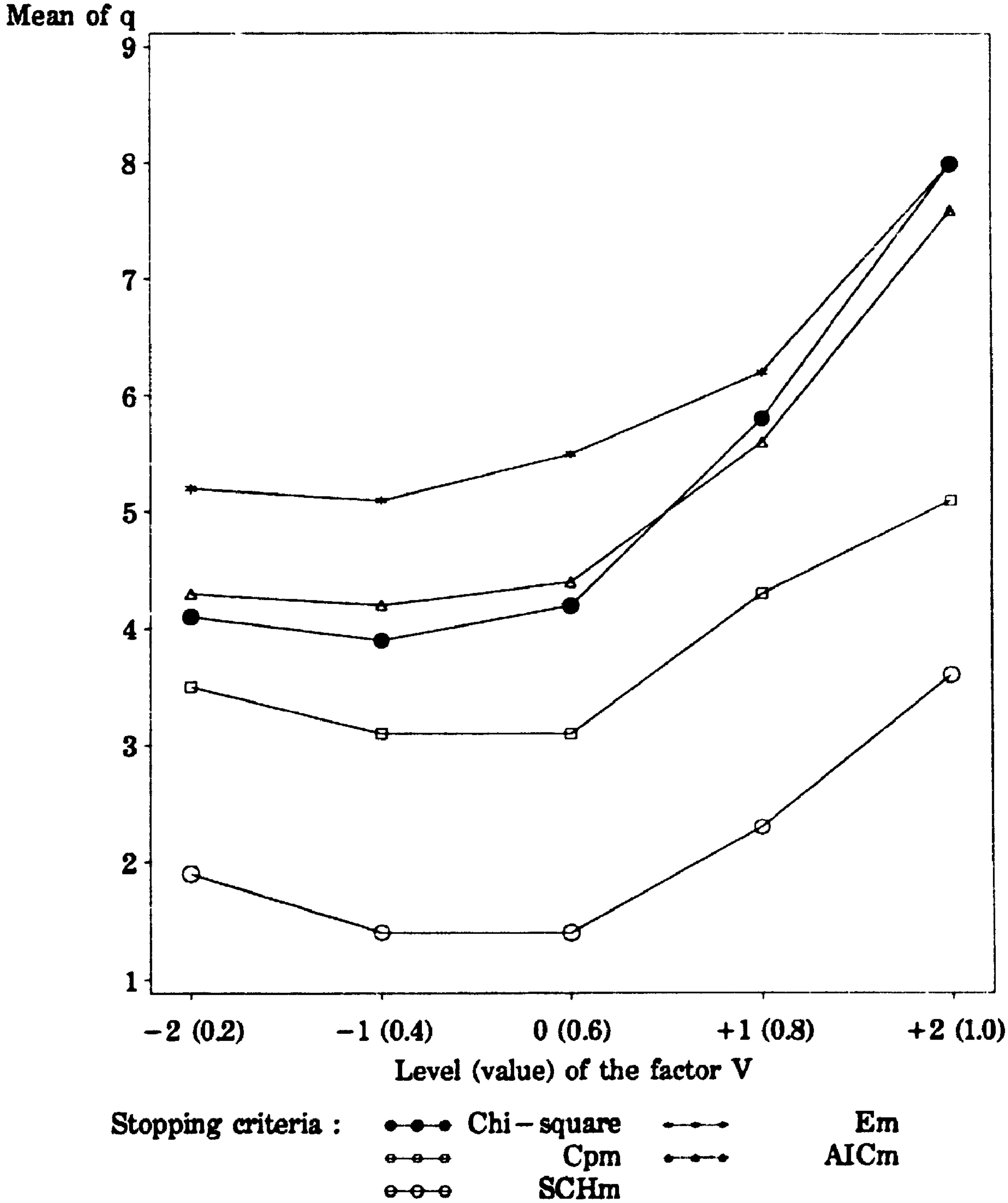


Figure 5.4.1.3

Mean of q for the five stopping criteria over the levels of the factor M

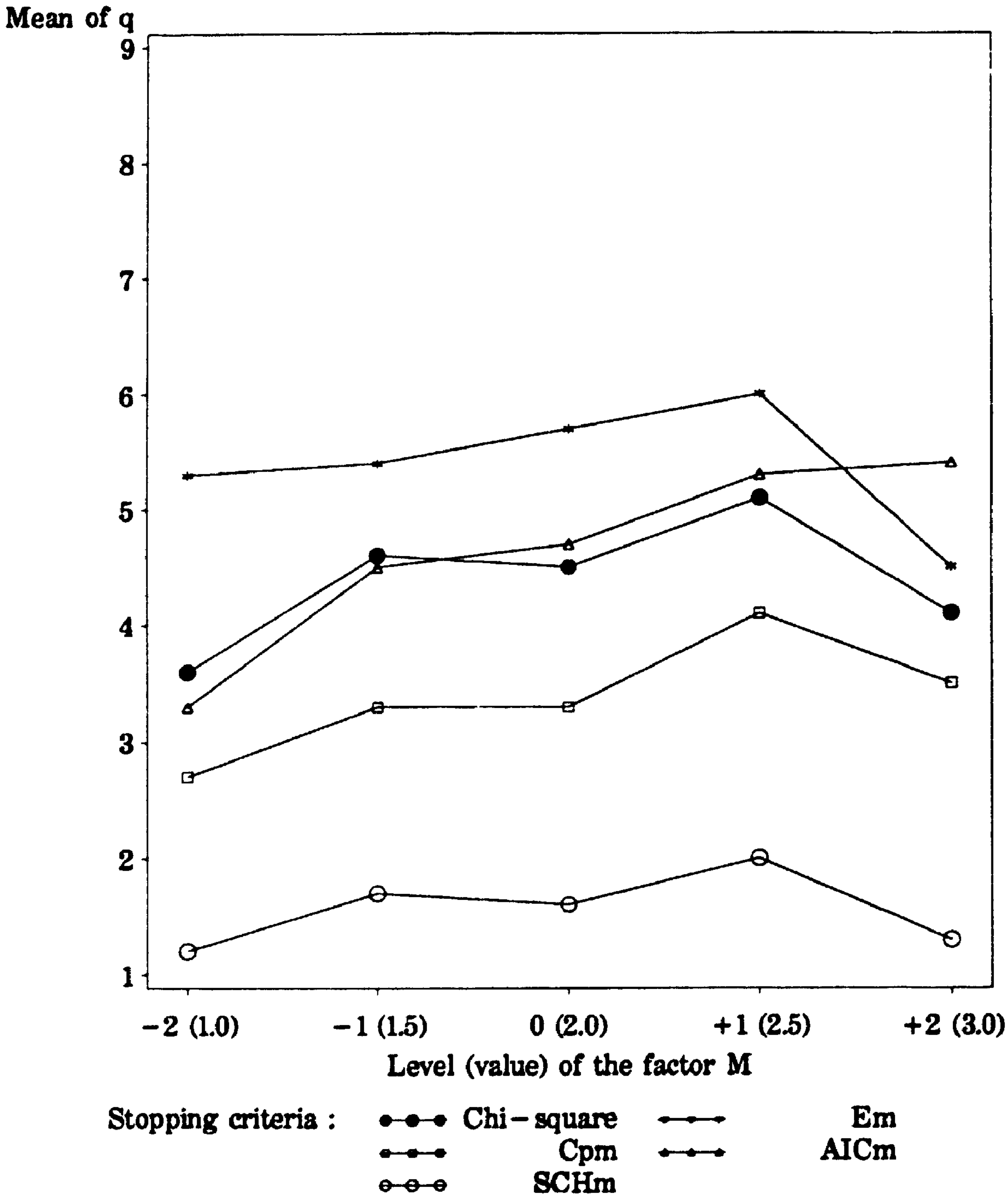


Figure 5.4.1.4

Mean of q for the five stopping criteria over the levels of the factor D

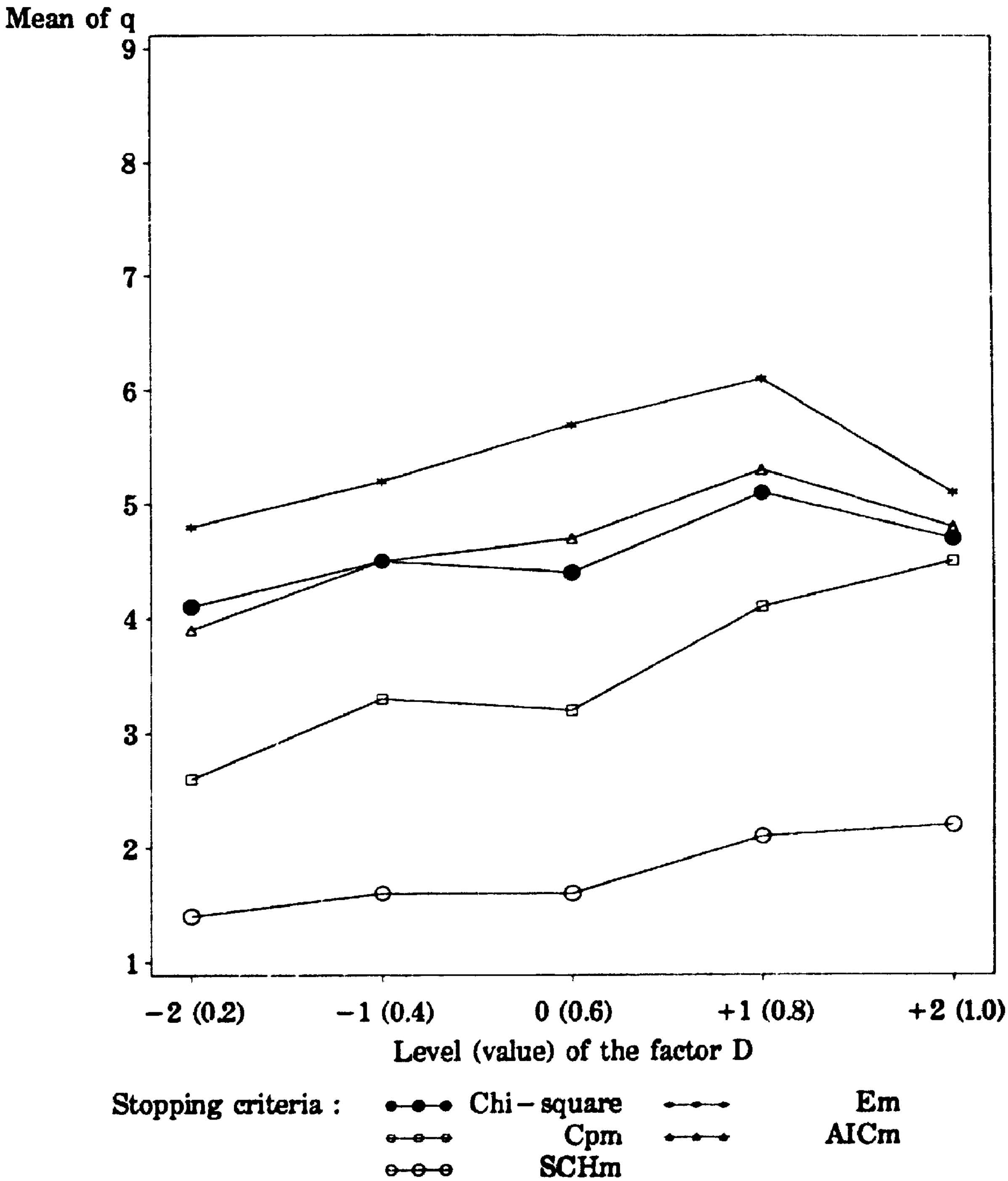


Figure 5.4.1.5

Mean of q for the five stopping criteria over the levels of the factor N

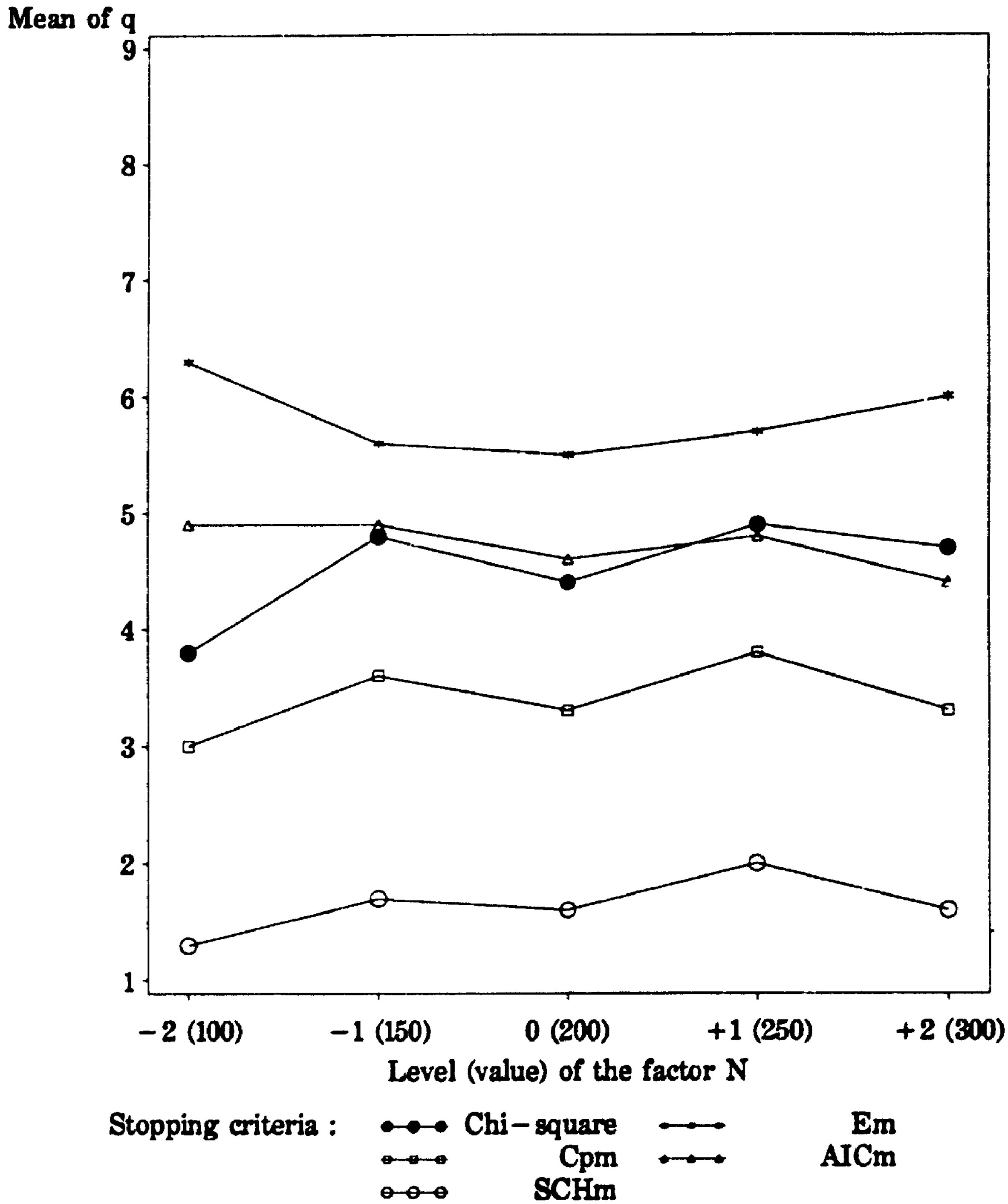


Figure 5.4.1.6

Effect of PD interaction on q for the Chi – square stopping criterion

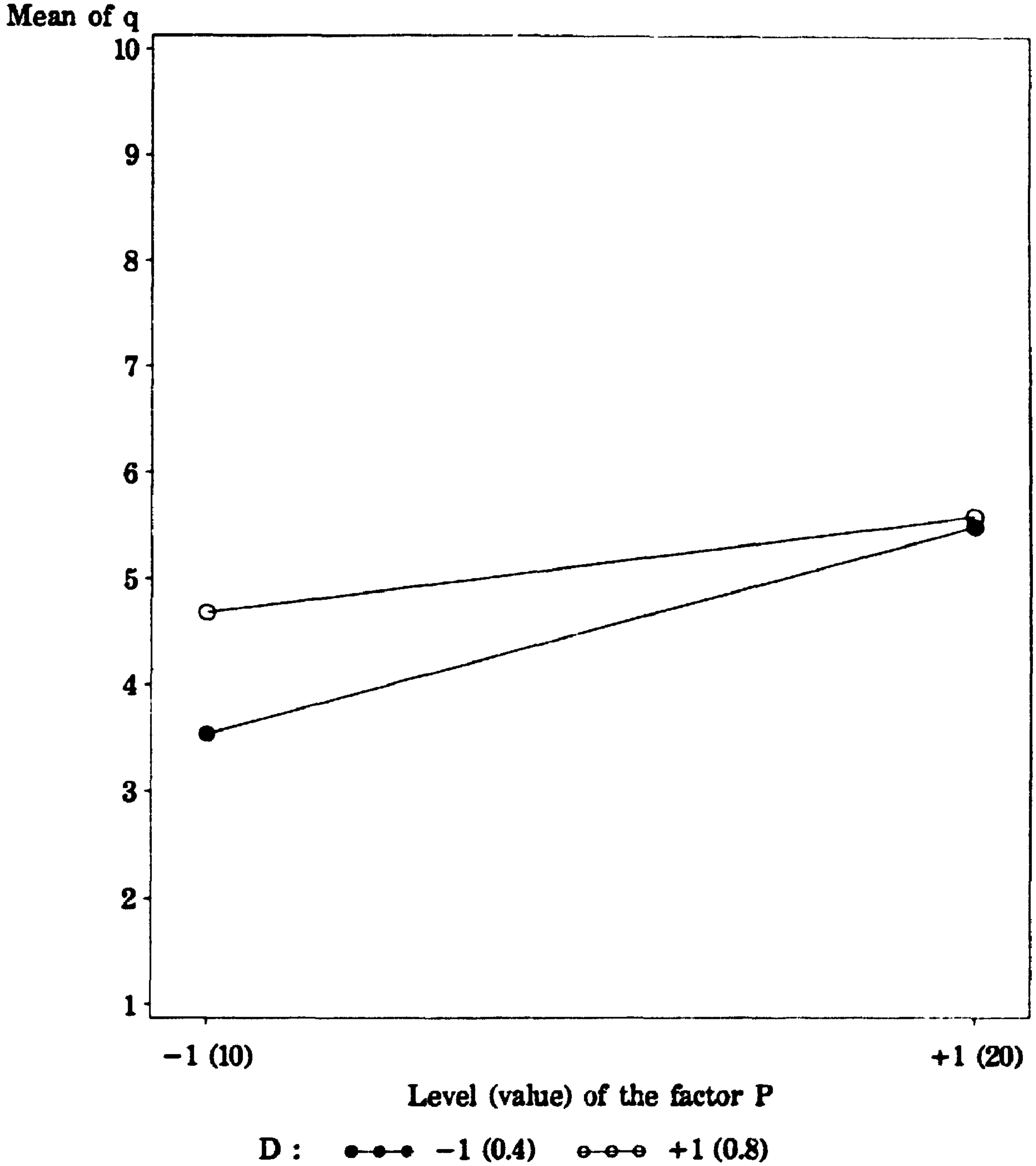


Figure 5.4.1.7

Effect of VD interaction on q for the Chi−square stopping criterion

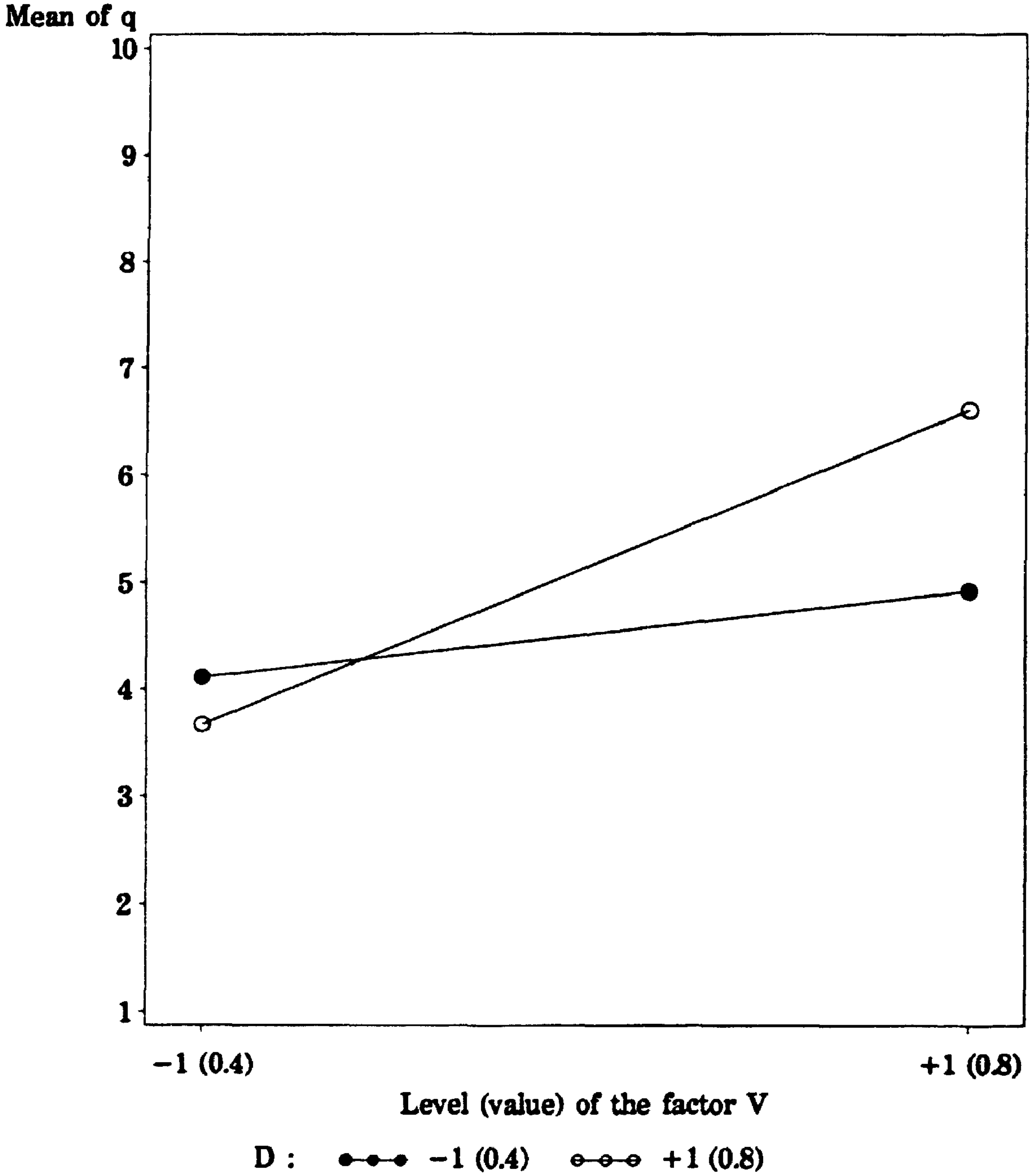


Figure 5.4.1.8
Effect of MD interaction on q for the Chi-square stopping criterion

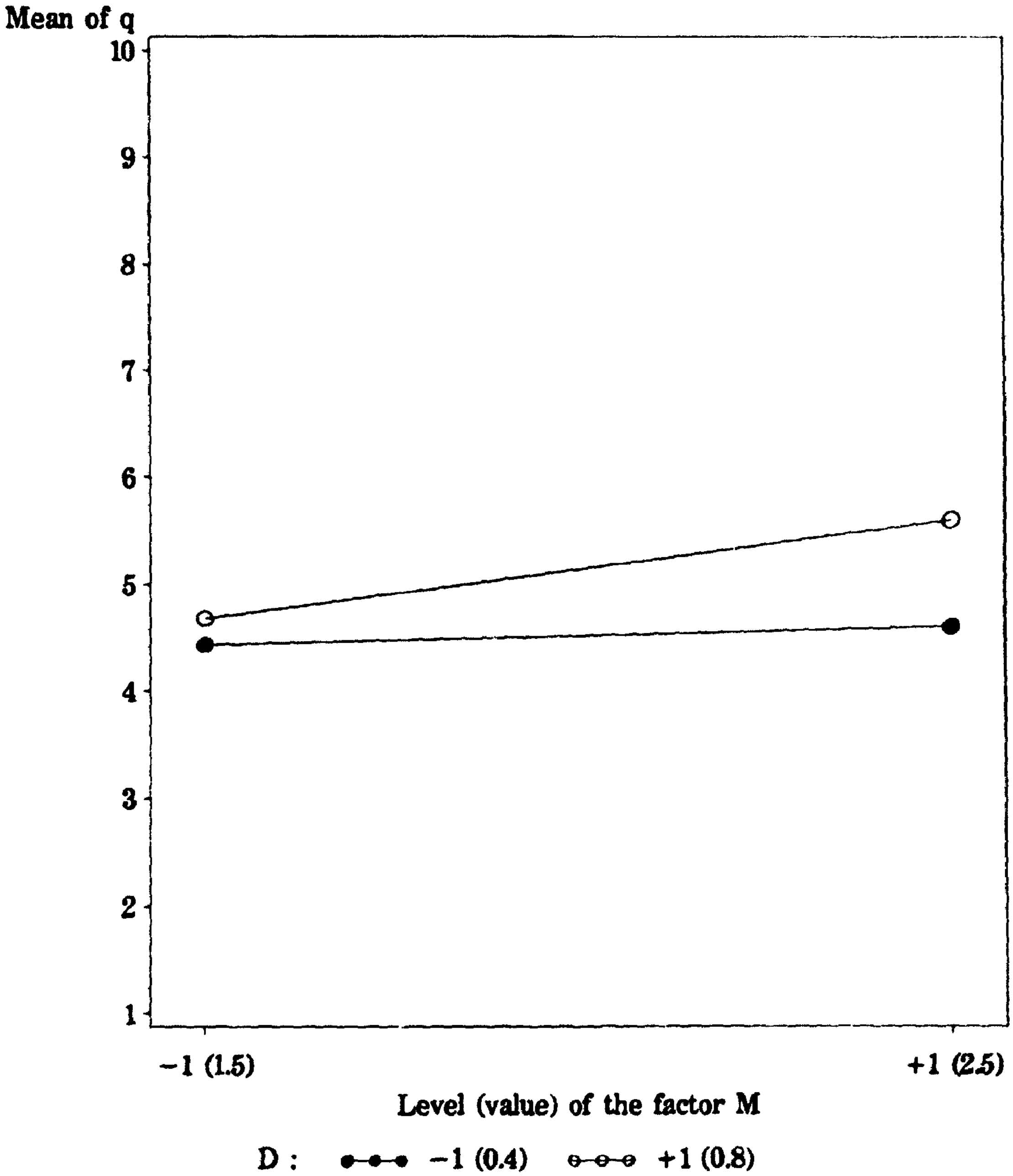


Figure 5.4.1.9

Effect of VD interaction on q for the Em stopping criterion

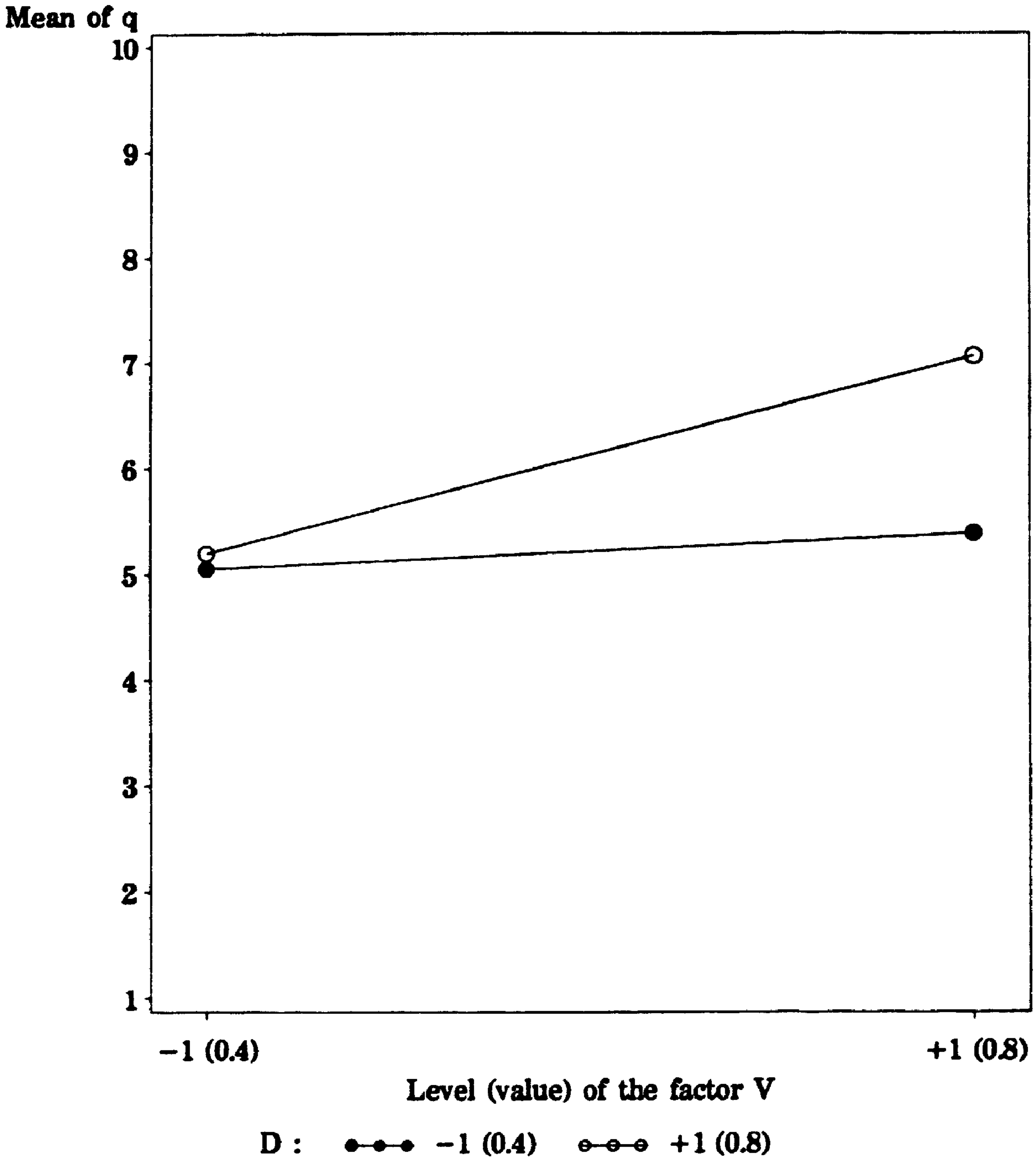


Figure 5.4.1.10

Effect of VD interaction on q
for the Cpm stopping criterion

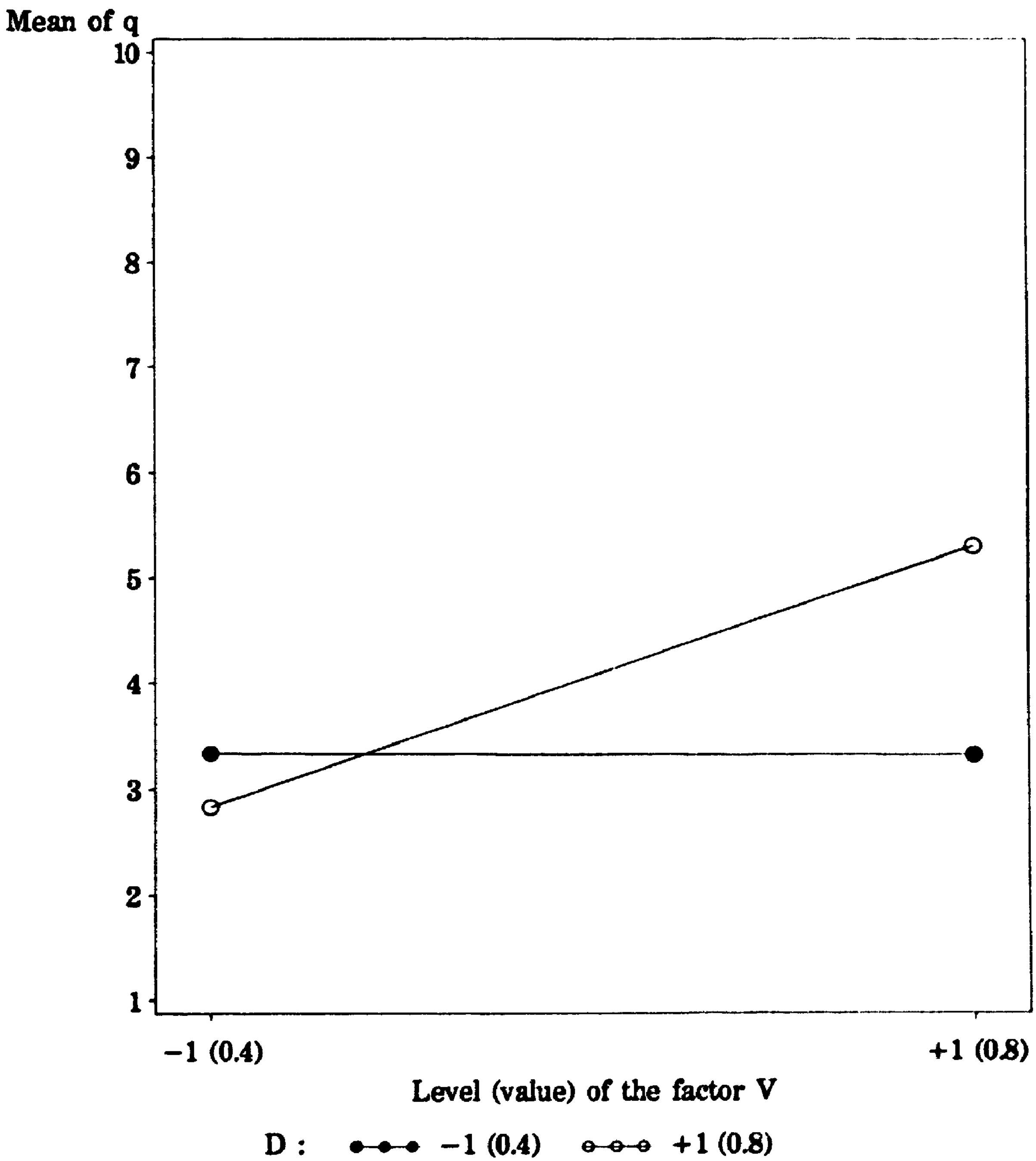


Figure 5.4.1.11

Effect of VD interaction on q for the AICm stopping criterion

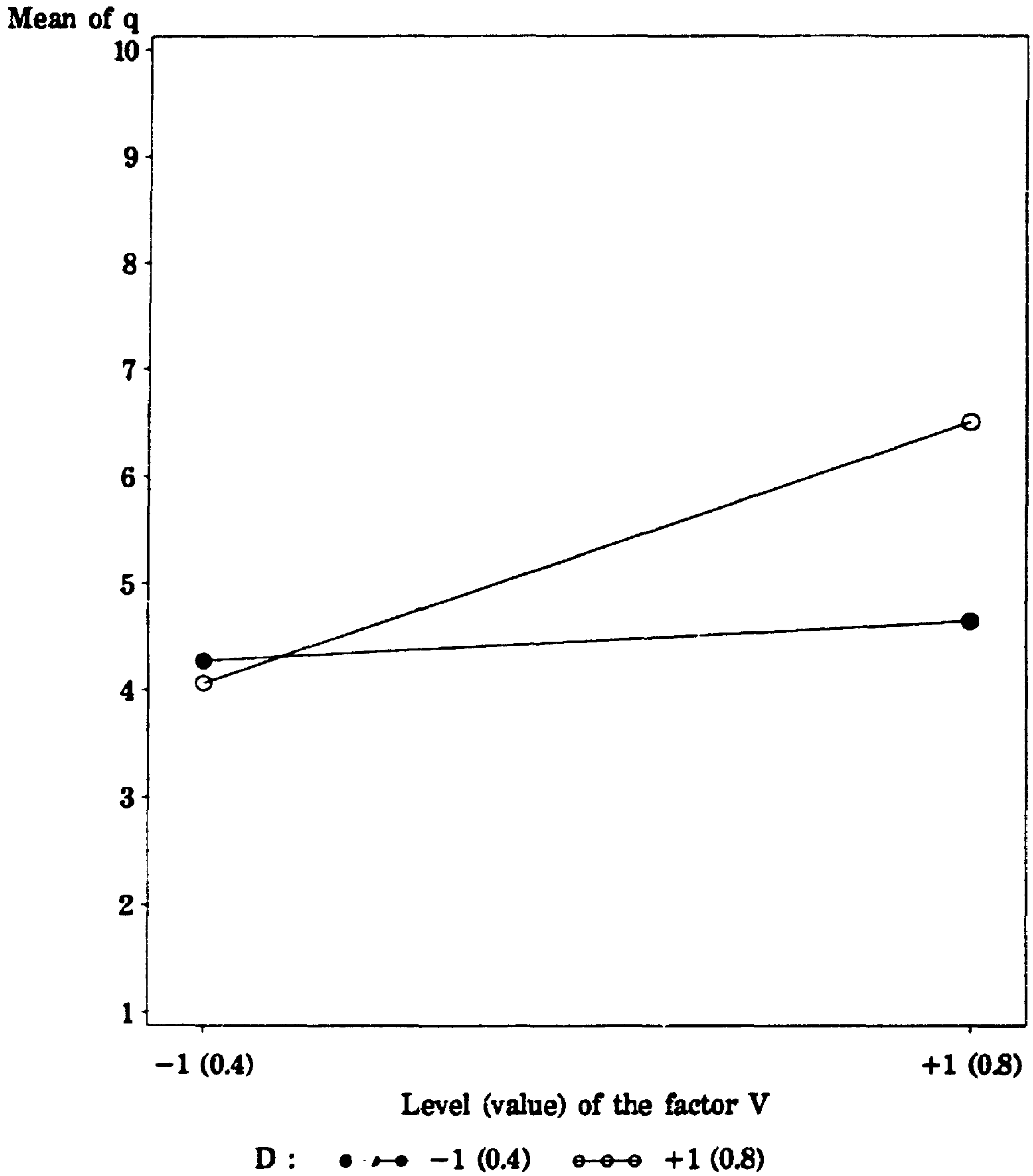


Figure 5.4.1.12

Effect of VD interaction on q for the SCHm stopping criterion

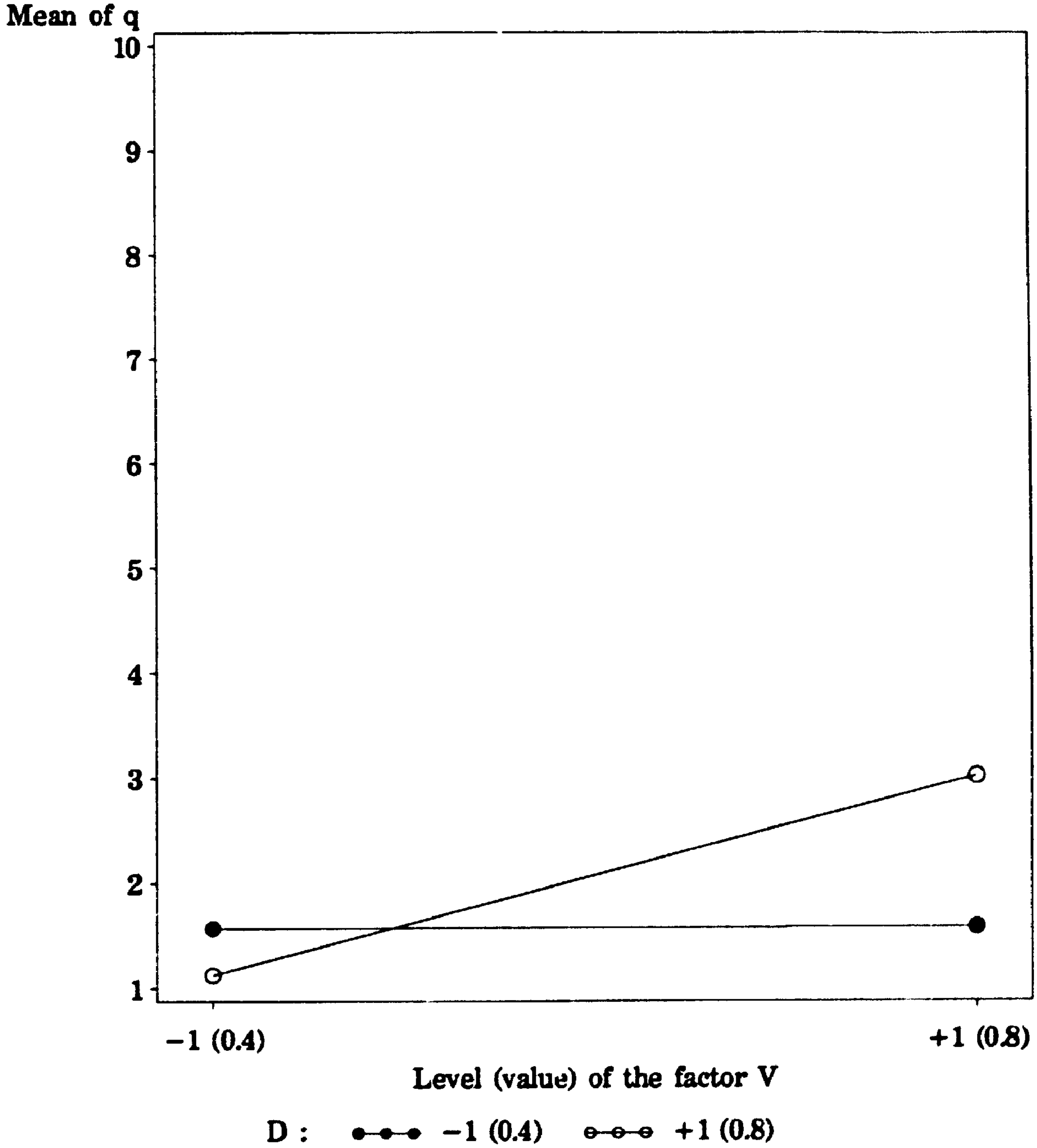


Figure 5.4.1.13

The proportional size of the logistic model over alpha level of significance for different values of P

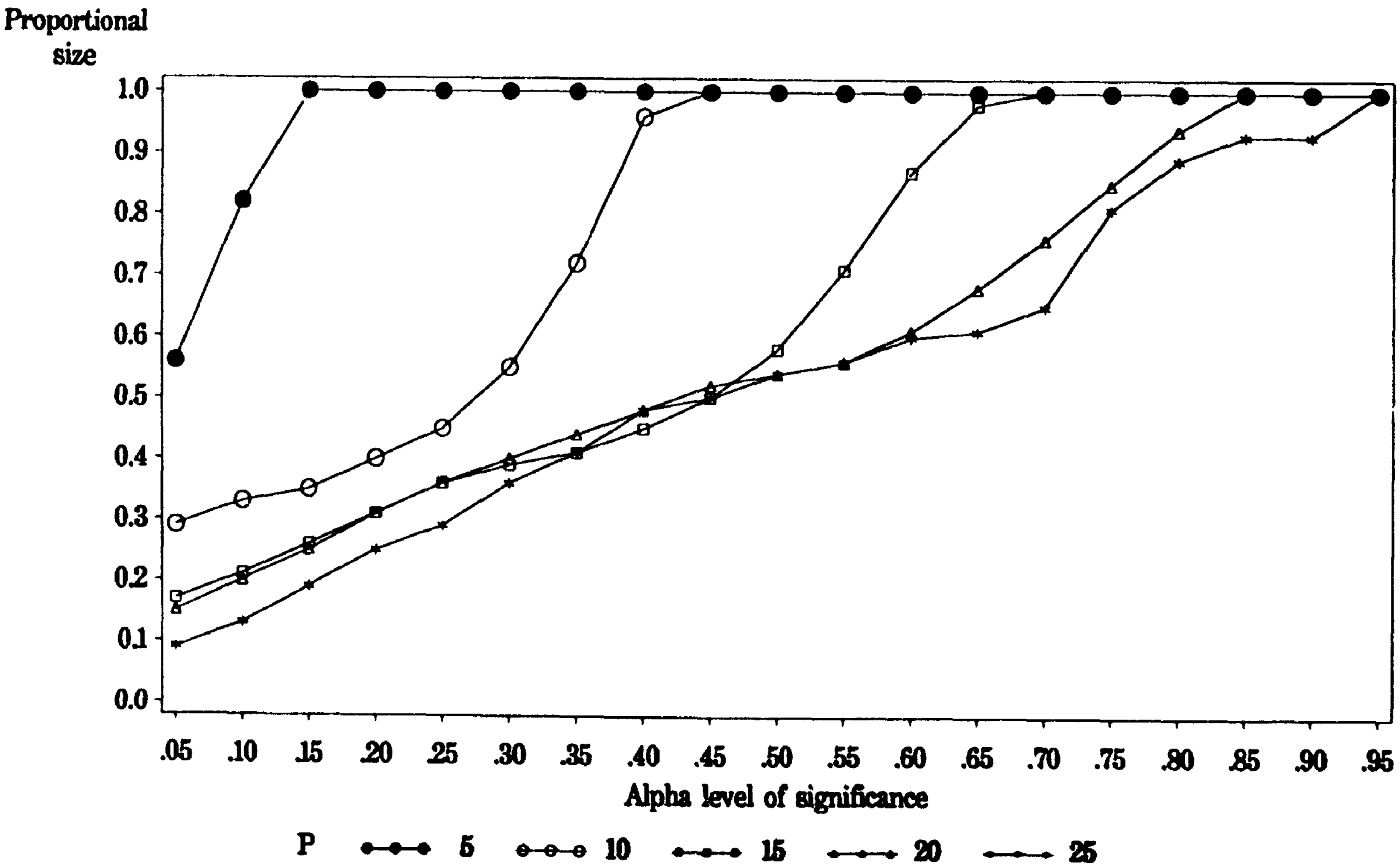


Table 5.4.2.1
Repeated measures analysis of variance of q
Test of Between Subject Effect

Effect	d.f.	Mean Square	F-ratio	p-value
[a]P	2	13005.16	2350.94	0.0001
[a]B	2	1733.89	313.44	0.0001
[a]M	2	14625.16	2643.79	0.0001
[a]N	2	1012.98	183.12	0.0001
[a]PB	4	67.85	12.26	0.0001
[a]PM	4	462.21	83.55	0.0001
[b]PN	4	16.46	2.98	0.0184
[a]BM	4	188.39	34.05	0.0001
BN	4	5.17	0.93	0.4430
[a]MN	4	39.61	7.16	0.0001
PBM	8	3.59	0.65	0.7362
PBN	8	2.76	0.50	0.8570
[b]PMN	8	13.85	2.50	0.0106
BMN	8	7.27	1.31	0.2320
PBMN	16	8.36	1.51	0.0875
Error(between)	1539	5.53		

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

Table 5.4.2.1 (continued)
Repeated measures analysis of variance of q
Test of Within Subject Effect

Effect	d.f.	Mean Square	F-ratio	[c]p–value
[a]S	4	14311.15	4921.70	0.0001
[a]SP	8	1130.62	388.83	0.0001
[a]SB	8	163.09	56.09	0.0001
[a]SM	8	182.71	62.84	0.0001
[a]SN	8	36.68	12.62	0.0001
[a]SPB	16	11.35	3.90	0.0001
[a]SPM	16	25.66	8.82	0.0001
[a]SPN	16	11.00	3.78	0.0001
[a]SBM	16	19.06	6.56	0.0001
[a]SBN	16	9.80	3.37	0.0002
[a]SMN	16	11.65	4.00	0.0001
[a]SPBM	32	11.50	3.96	0.0001
[a]SPBN	32	7.29	2.51	0.0002
[a]SPMN	32	7.48	2.57	0.0001
[a]SBMN	32	6.68	2.30	0.0008
[b]SPBMN	64	4.88	1.68	0.0046
Error(within)	6156	2.91		

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

[c]Greenhouse-Geisser's method

Table 5.4.2.2
[a]Mean of q for the five stopping criteria
over the levels of the four factors P, B, M, and N

Factor	[b]Level	Value	Stopping criteria				
			$\chi^2_{(0.15)}$	E_m	C_{pm}	AIC_m	SCH_m
P	-1	10	6.1	5.5	3.9	6.1	2.5
	0	15	9.3	9.0	3.8	9.4	3.4
	+1	20	12.2	12.5	4.6	12.4	4.3
B	-1	0.2	10.1	10.1	4.4	10.2	4.0
	0	0.4	9.5	9.4	3.7	9.6	3.3
	+1	0.6	8.0	7.6	4.2	8.1	2.9
M	-1	0.1	6.2	7.2	1.9	6.3	1.2
	0	0.2	9.9	9.3	3.7	10.0	3.4
	+1	0.3	11.5	10.6	6.8	11.5	5.6
N	-1	150	8.3	8.6	3.9	8.5	2.9
	0	200	9.2	9.0	4.0	9.3	3.4
	+1	250	10.1	9.5	4.5	10.2	4.0
Overall			9.2	9.0	4.1	9.3	3.4

[a]The standard error of mean is in the range of 0.0196 and 0.2001.

[b]The number of sampling situations for each level is 27.

Table 5.4.2.3
Analysis of variance of q for the $\chi^2_{(0.15)}$ stopping criterion

Effect	d.f.	Mean Square	F-ratio	p-value
[a]P	2	4895.79	1969.58	0.0001
[a]B	2	648.47	260.88	0.0001
[a]M	2	4008.44	1612.60	0.0001
[a]N	2	444.69	178.90	0.0001
[a]PB	4	23.48	9.44	0.0001
[a]PM	4	53.82	21.65	0.0001
[a]PN	4	12.15	4.89	0.0006
[a]BM	4	45.93	18.48	0.0001
BN	4	3.72	1.50	0.2005
[a]MN	4	12.99	5.23	0.0004
PBM	8	1.03	0.41	0.9132
PBN	8	3.19	1.28	0.2485
PMN	8	3.05	1.23	0.2801
BMN	8	1.63	0.66	0.7309
PBMN	16	3.59	1.44	0.1133
Error	1539	2.49		

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

$R^2 = 0.84$

Table 5.4.2.4
Analysis of variance of q for the E_m stopping criterion

Effect	d.f.	Mean Square	F-ratio	p-value
[a]P	2	6671.17	981.77	0.0001
[a]B	2	901.34	132.65	0.0001
[a]M	2	1564.36	230.22	0.0001
[a]N	2	104.16	15.33	0.0001
[b]PB	4	21.86	3.22	0.0122
[a]PM	4	161.33	23.74	0.0001
PN	4	12.97	1.91	0.1065
[a]BM	4	109.55	16.12	0.0001
BN	4	11.35	1.67	0.1543
MN	4	7.55	1.11	0.3495
PBM	8	8.36	1.23	0.2774
[b]PBN	8	13.51	1.99	0.0444
[b]PMN	8	15.30	2.25	0.0216
BMN	8	7.43	1.09	0.3643
PBMN	16	7.21	1.06	0.3874
Error	1539	6.80		

[a]significant with $p < 0.001$
[b]significant with $p < 0.05$
$R^2 = 0.66$

Table 5.4.2.5
Analysis of variance of q for the AIC_m stopping criterion

Effect	d.f.	Mean Square	F-ratio	p-value
[a]P	2	5392.07	2243.51	0.0001
[a]B	2	609.53	253.61	0.0001
[a]M	2	3846.86	1600.58	0.0001
[a]N	2	391.86	163.04	0.0001
[a]PB	4	26.57	11.06	0.0001
[a]PM	4	70.86	29.48	0.0001
[b]PN	4	10.31	4.29	0.0019
[a]BM	4	31.45	13.09	0.0001
BN	4	2.15	0.89	0.4662
[b]MN	4	5.92	2.46	0.0434
PBM	8	1.23	0.51	0.8478
PBN	8	2.93	1.22	0.2825
PMN	8	2.47	1.03	0.4119
BMN	8	1.09	0.45	0.8882
PBMN	16	3.49	1.45	0.1099
Error	1539	2.40		

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

$R^2 = 0.85$

Table 5.4.2.6
Analysis of variance of q for the C_{pm} stopping criterion

Effect	d.f.	Mean Square	F-ratio	p-value
[a]P	2	103.36	22.45	0.0001
[a]B	2	67.43	14.65	0.0001
[a]M	2	3262.72	708.80	0.0001
[a]N	2	70.94	15.41	0.0001
[a]PB	4	35.78	7.77	0.0001
[a]PM	4	174.34	37.87	0.0001
[b]PN	4	19.66	4.27	0.0019
[a]BM	4	43.79	9.51	0.0001
[a]BN	4	23.85	5.18	0.0004
[a]MN	4	29.78	6.47	0.0001
[a]PBM	8	38.15	8.29	0.0001
[b]PBN	8	11.27	2.45	0.0124
[a]PMN	8	21.74	4.72	0.0001
[a]BMN	8	22.97	4.99	0.0001
[a]PBMN	16	12.69	2.76	0.0002
Error	1539	4.60		

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

$R^2 = 0.57$

Table 5.4.2.7
Analysis of variance of q for the SCH_m stopping criterion

Effect	d.f.	Mean Square	F-ratio	p-value
[a]P	2	465.27	531.33	0.0001
[a]B	2	159.51	182.15	0.0001
[a]M	2	2673.65	3053.28	0.0001
[a]N	2	148.50	169.09	0.0001
[a]PB	4	5.56	6.35	0.0001
[a]PM	4	104.50	119.34	0.0001
[a]PN	4	5.37	6.13	0.0001
[a]BM	4	33.92	38.74	0.0001
[b]BN	4	3.31	3.78	0.0046
[a]MN	4	29.95	34.21	0.0001
PBM	8	0.84	0.95	0.4700
PBN	8	1.00	1.14	0.3317
PMN	8	1.19	1.36	0.2085
BMN	8	0.86	0.98	0.4508
PBMN	16	0.91	1.04	0.4131
Error	1539	0.88		

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

$R^2 = 0.85$

Table 5.4.2.8

The mean of absolute size ($\bar{q}$) and proportional size ($\frac{\bar{q}}{P}$) of the logistic model over the α level of significance for different values of P

P	10	15	20
α	$\bar{q}$ ($\frac{\bar{q}}{P}$)	$\bar{q}$ ($\frac{\bar{q}}{P}$)	$\bar{q}$ ($\frac{\bar{q}}{P}$)
0.05	4.8(0.49)	7.1(0.47)	9.4(0.47)
0.10	5.3(0.53)	8.1(0.54)	10.8(0.54)
0.15	5.8(0.58)	8.8(0.59)	11.8(0.59)
0.20	6.5(0.65)	9.3(0.62)	12.6(0.63)
0.25	7.4(0.74)	10.0(0.66)	13.1(0.65)
0.30	8.0(0.80)	10.7(0.71)	13.8(0.69)
0.35	8.7(0.87)	11.5(0.76)	14.7(0.73)
0.40	9.6(0.96)	12.3(0.82)	15.6(0.78)
0.45	10.0(1.00)	13.0(0.86)	16.2(0.81)
0.50	10.0(1.00)	13.8(0.92)	17.0(0.85)
0.55	10.0(1.00)	14.4(0.96)	17.6(0.88)
0.60	10.0(1.00)	14.8(0.99)	18.4(0.92)
0.65	10.0(1.00)	14.9(1.00)	19.0(0.95)
0.70	10.0(1.00)	15.0(1.00)	19.5(0.98)
0.75	10.0(1.00)	15.0(1.00)	19.9(0.99)
0.80	10.0(1.00)	15.0(1.00)	20.0(1.00)
0.85	10.0(1.00)	15.0(1.00)	20.0(1.00)
0.90	10.0(1.00)	15.0(1.00)	20.0(1.00)
0.95	10.0(1.00)	15.0(1.00)	20.0(1.00)

Figure 5.4.2.1
Mean of q for the five stopping criteria over the levels of the factor P

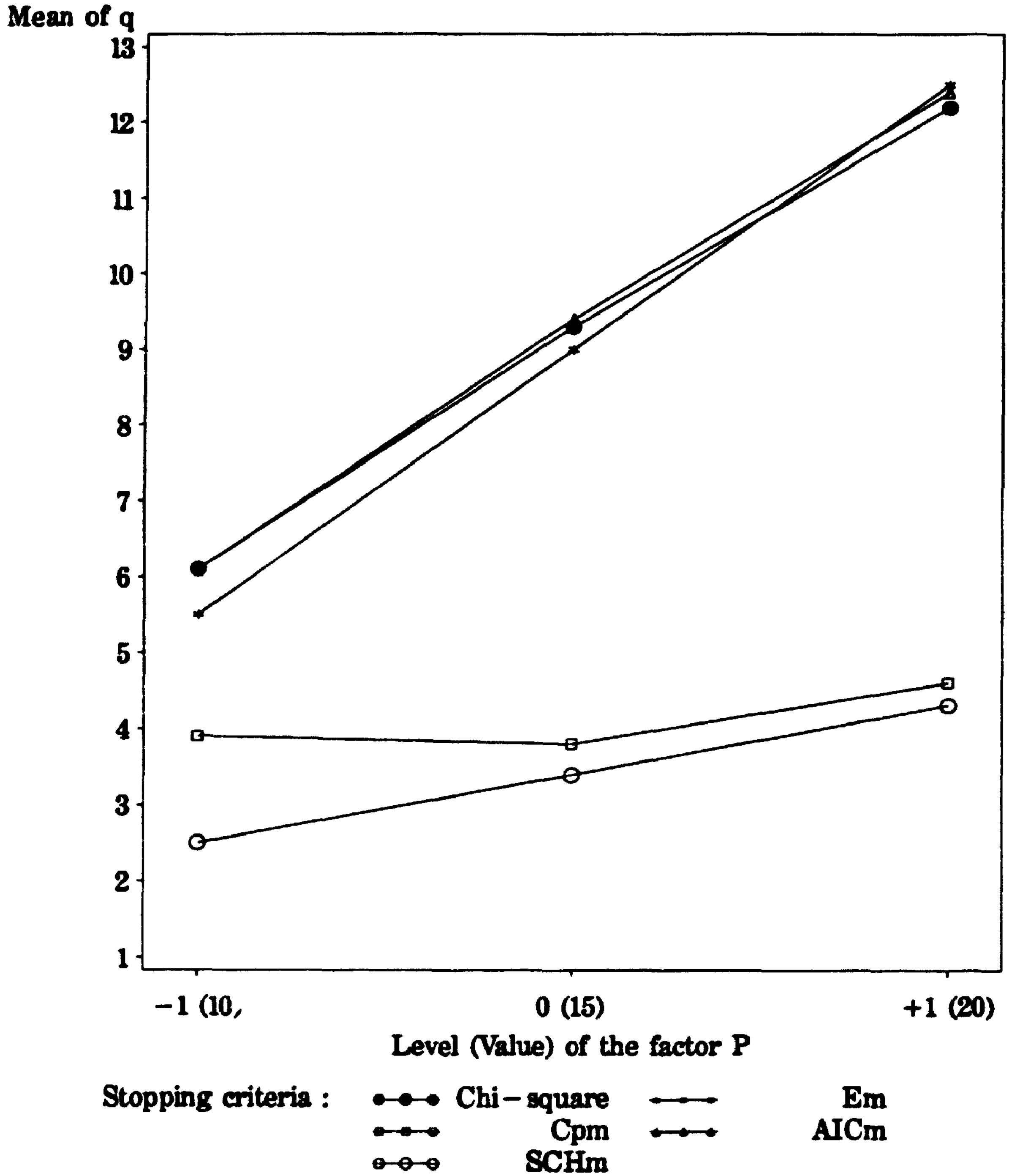


Figure 5.4.2.2

Mean of q for the five stopping criteria over the levels of the factor B

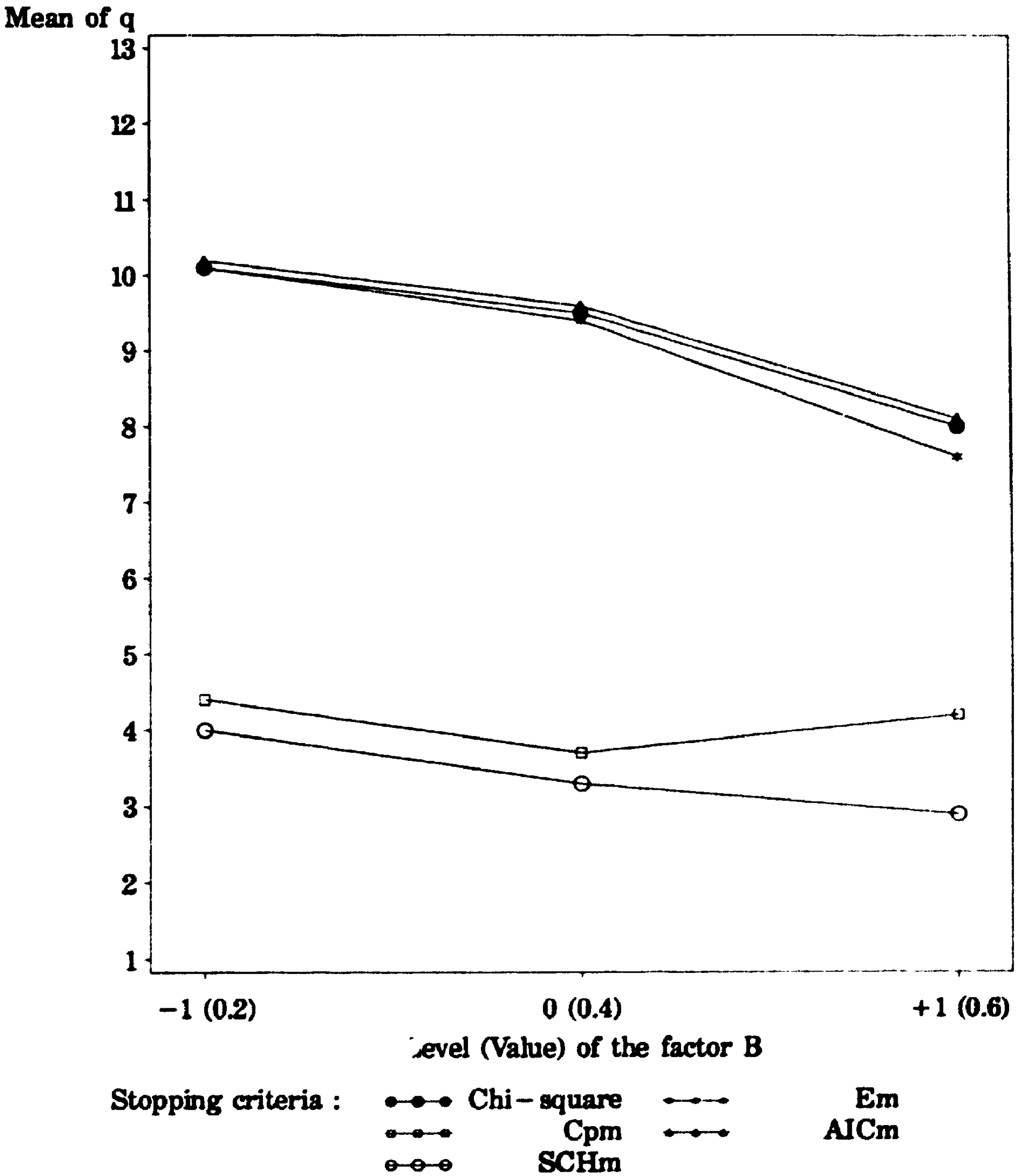


Figure 5.4.2.3

Mean of q for the five stopping criteria over the levels of the factor M

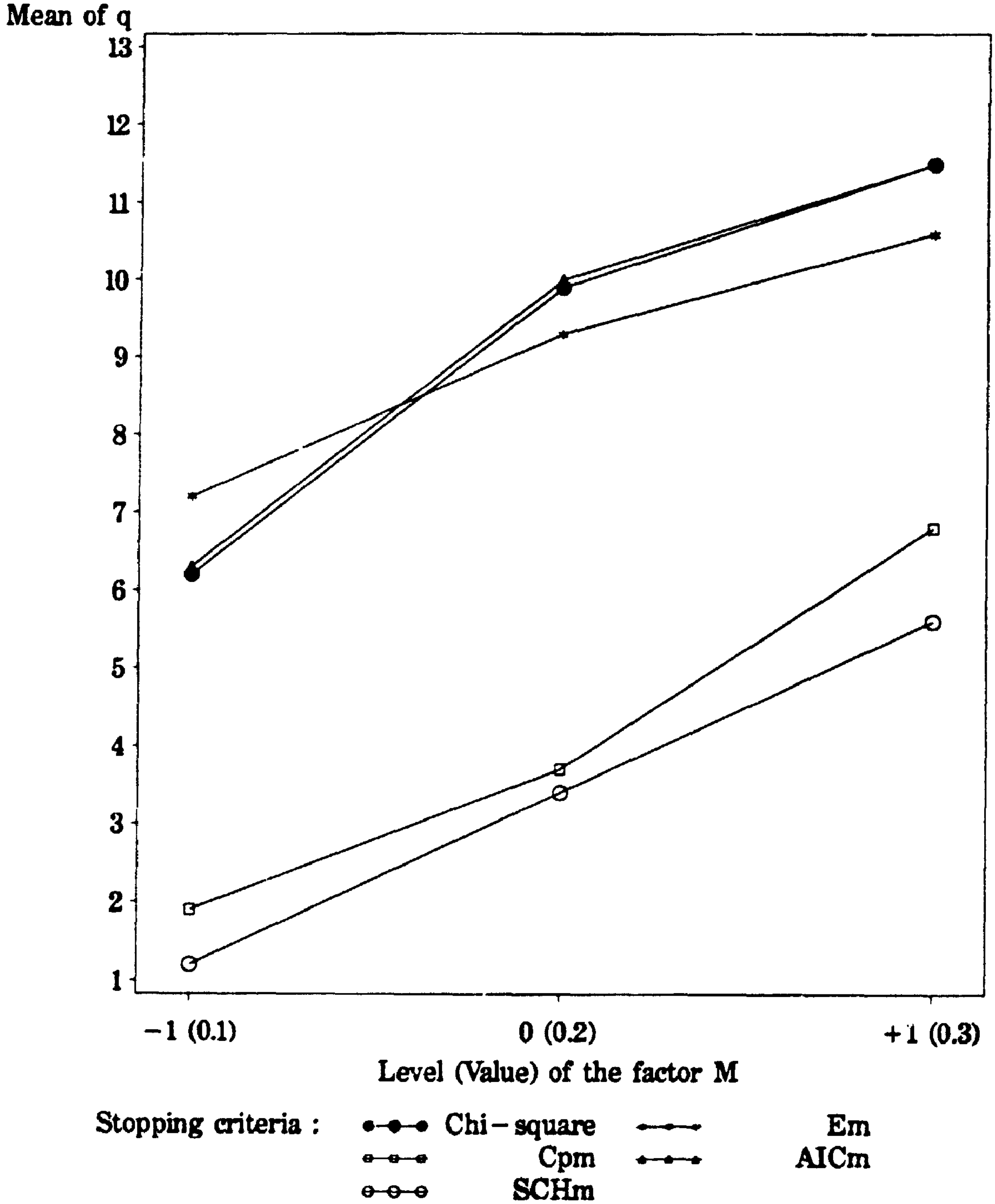


Figure 5.4.2.4

Mean of q for the five stopping criteria over the levels of the factor N

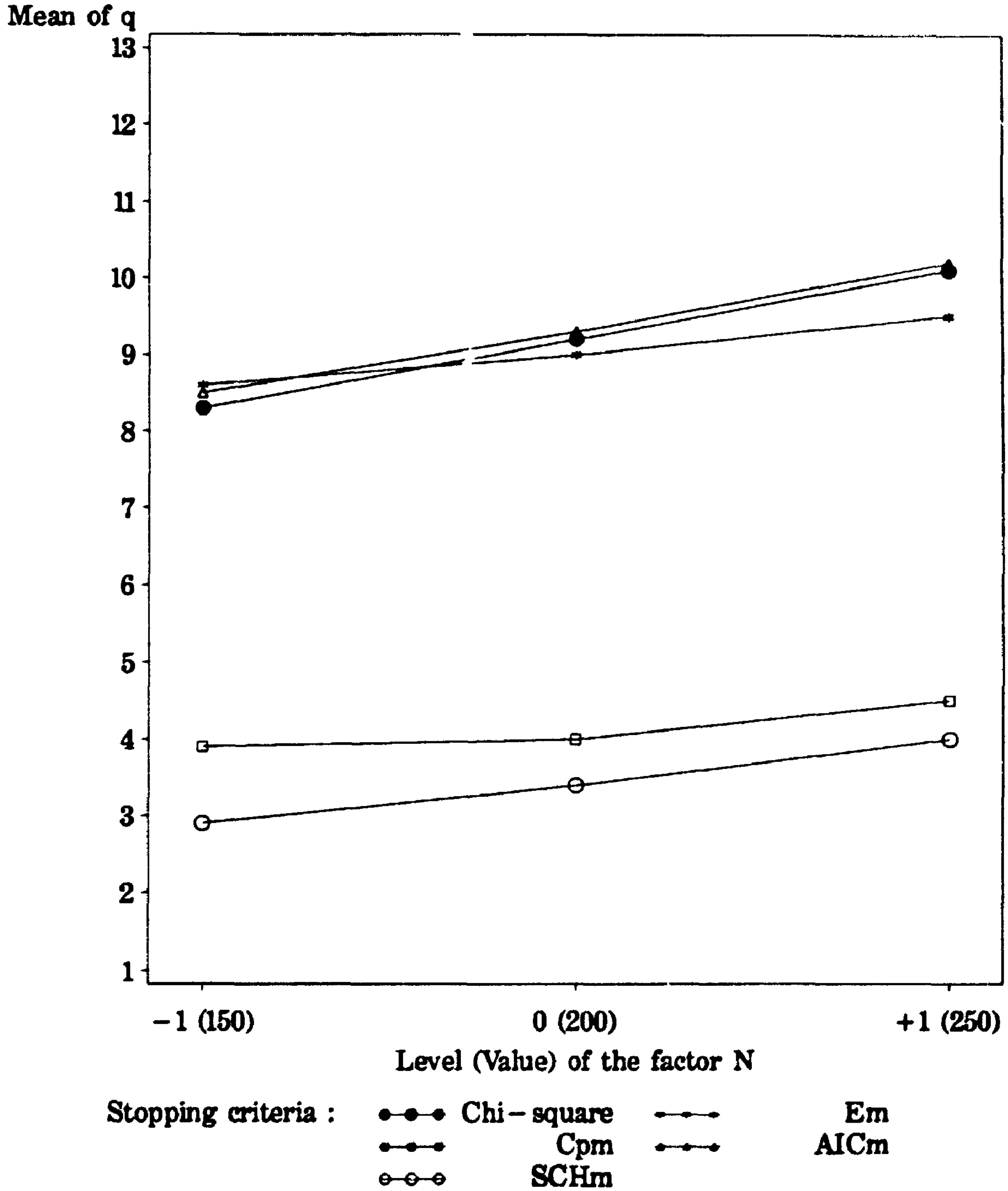


Figure 5.4.2.5

Effect of PM interaction on q for the Chi−square stopping criterion

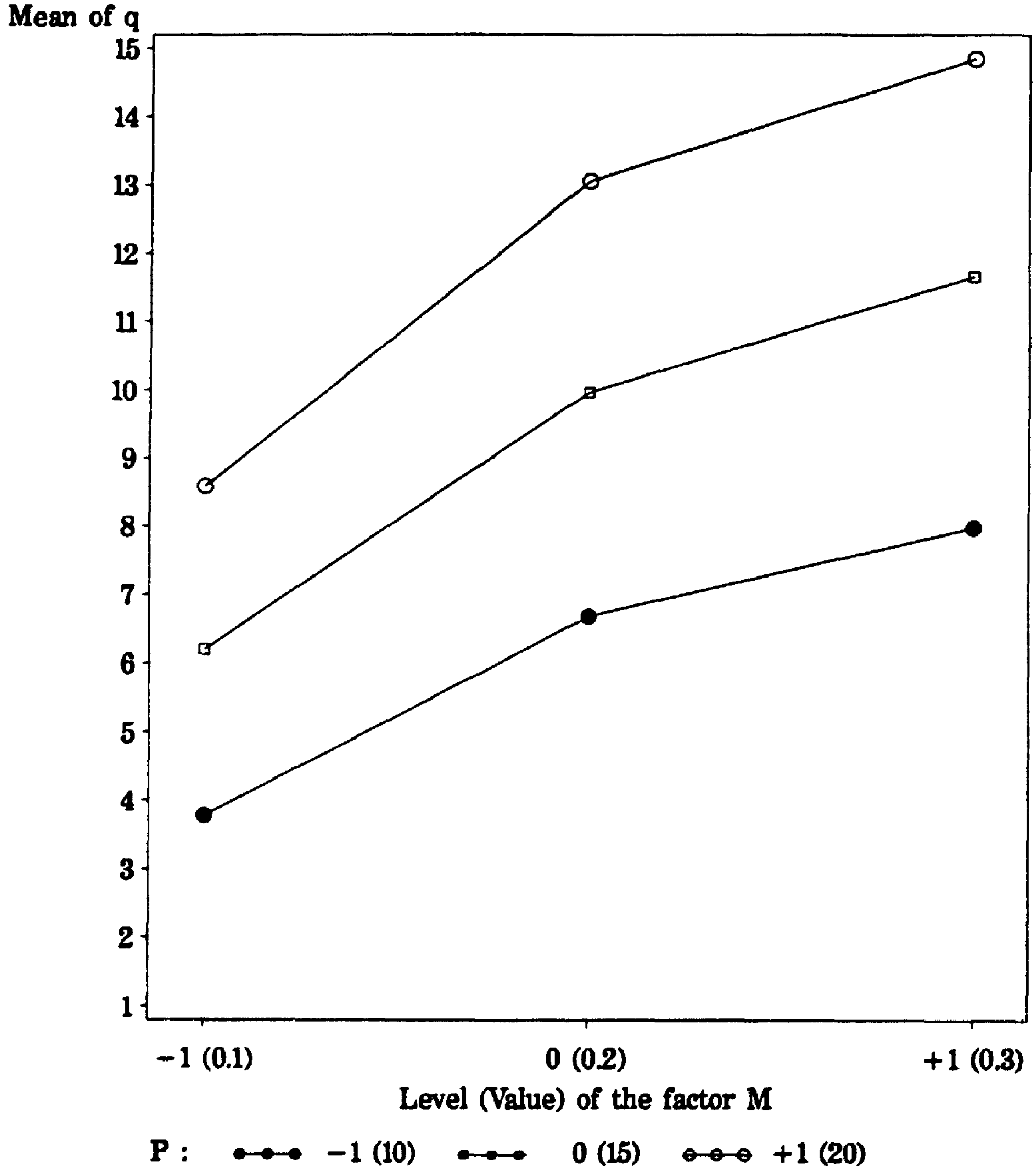


Figure 5.4.2.6

Effect of BM interaction on q
for the Chi – square stopping criterion

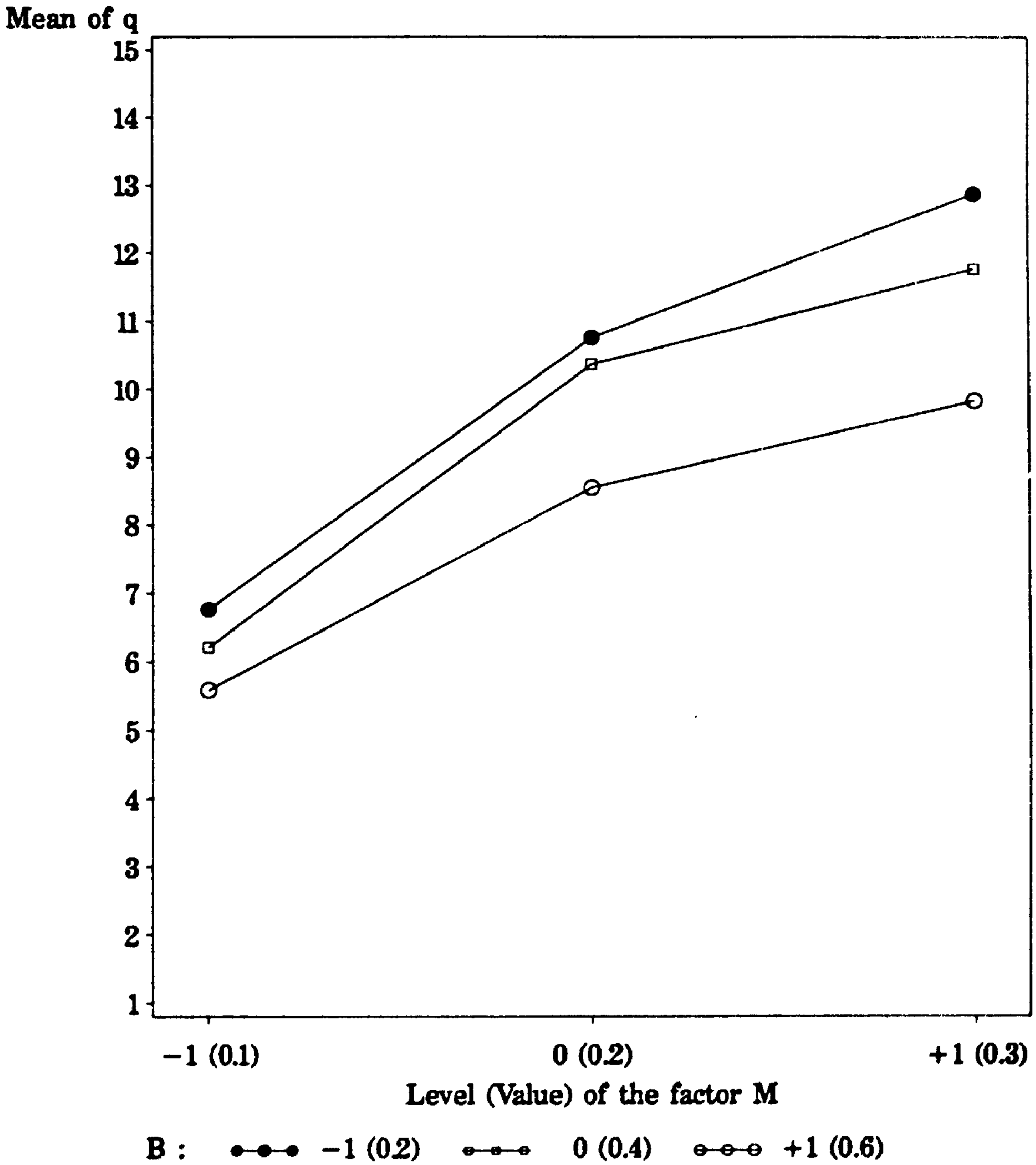


Figure 5.4.2.7

Effect of PM interaction on q for the Em stopping criterion

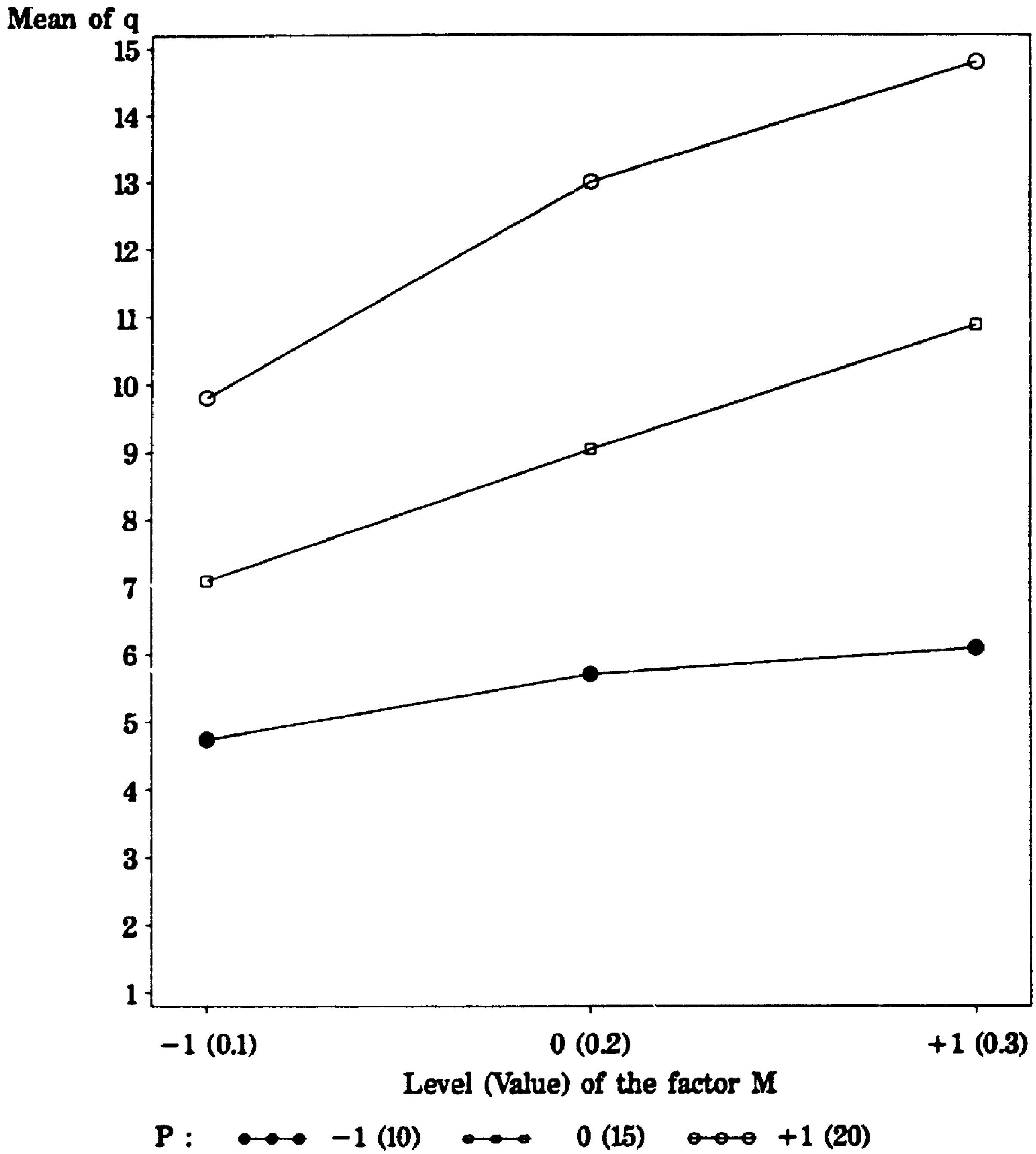


Figure 5.4.2.8

Effect of BM interaction on q for the Em stopping criterion

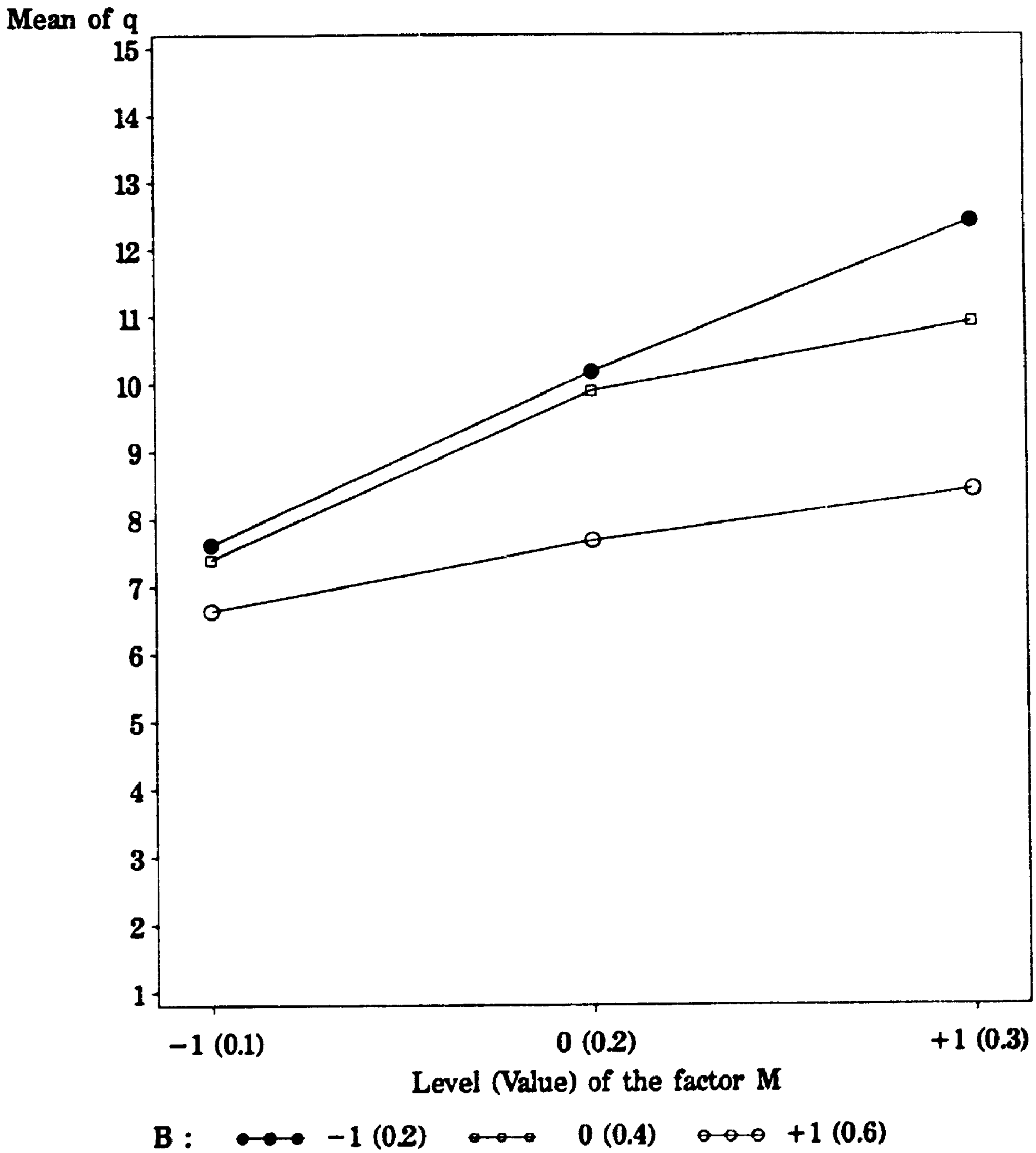


Figure 5.4.2.9
Effect of PM interaction on q
for the AICm stopping criterion

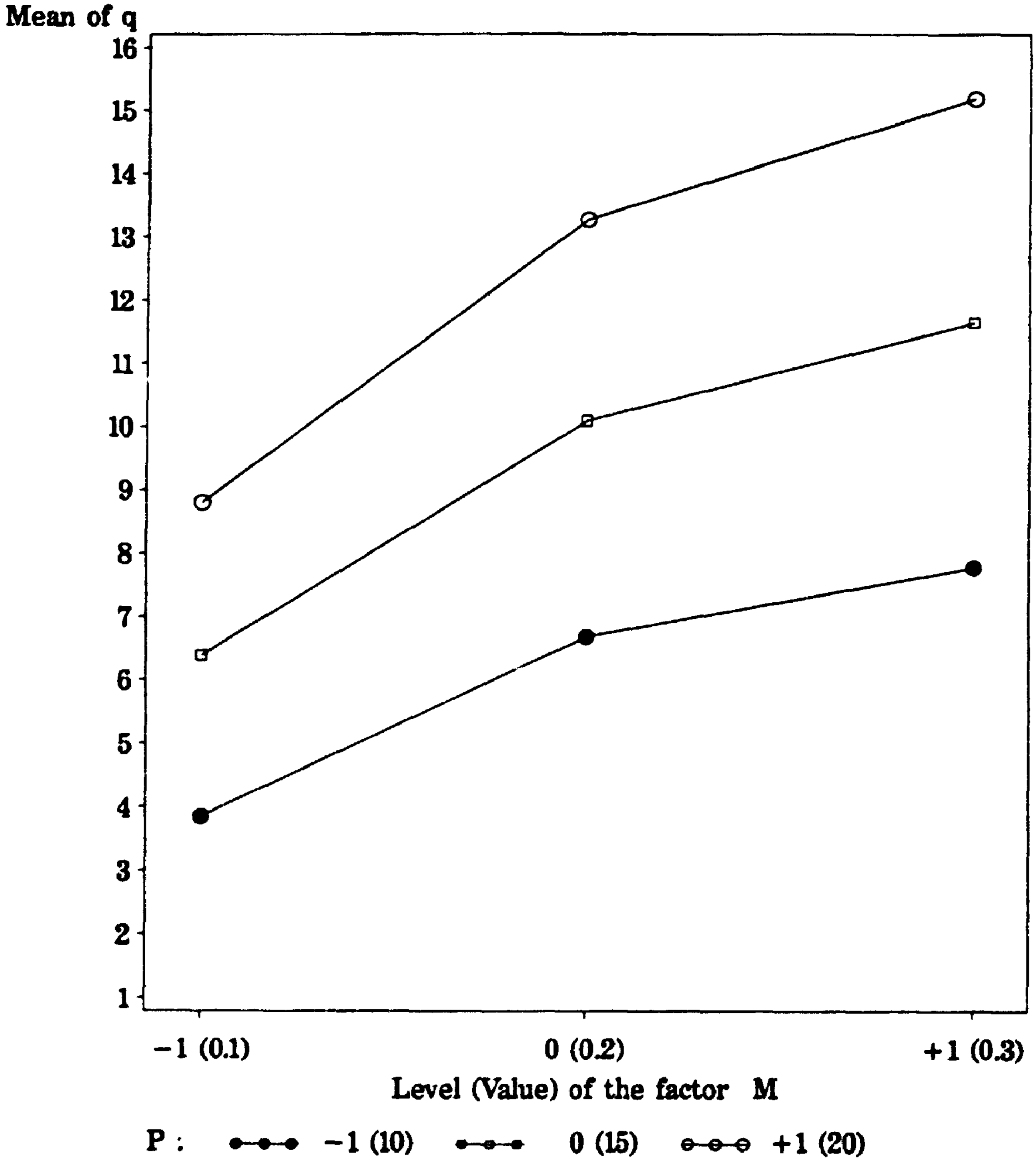


Figure 5.4.2.10

Effect of BM interaction on q for the AICm stopping criterion

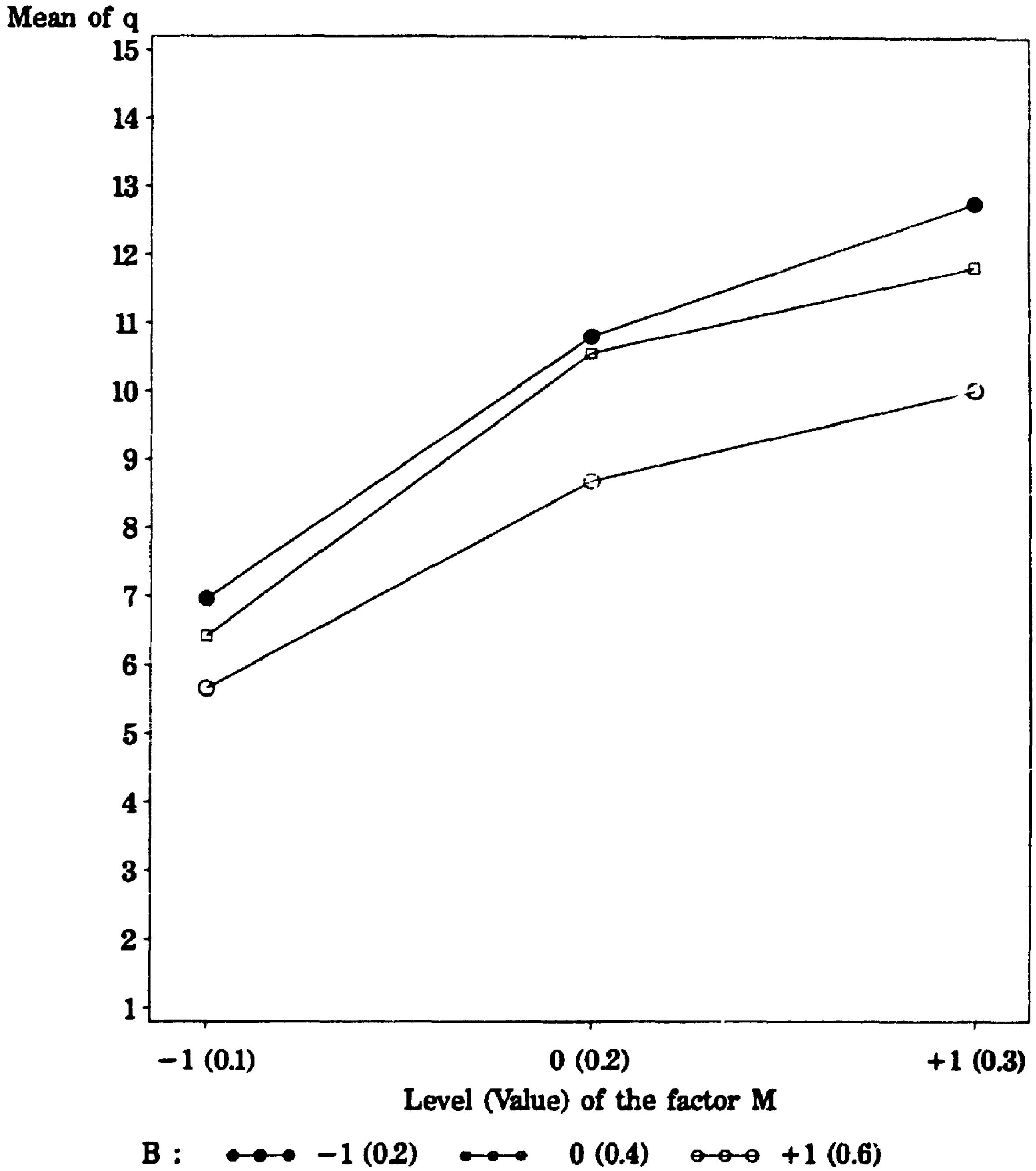


Figure 5.4.2.11
Effect of PM interaction on q for the Cpm stopping criterion

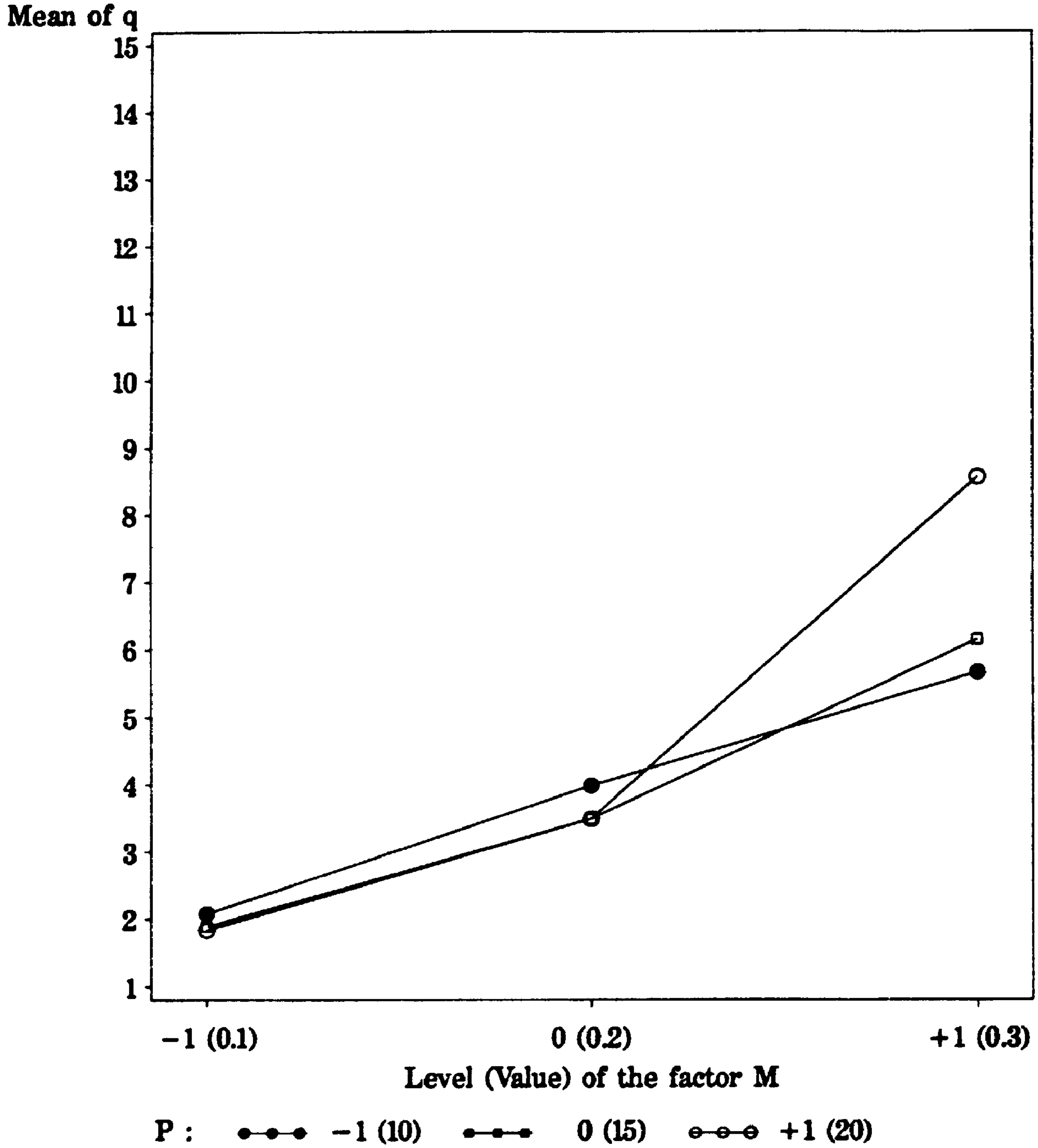


Figure 5.4.2.12

Effect of BM interaction on q for the Cpm stopping criterion

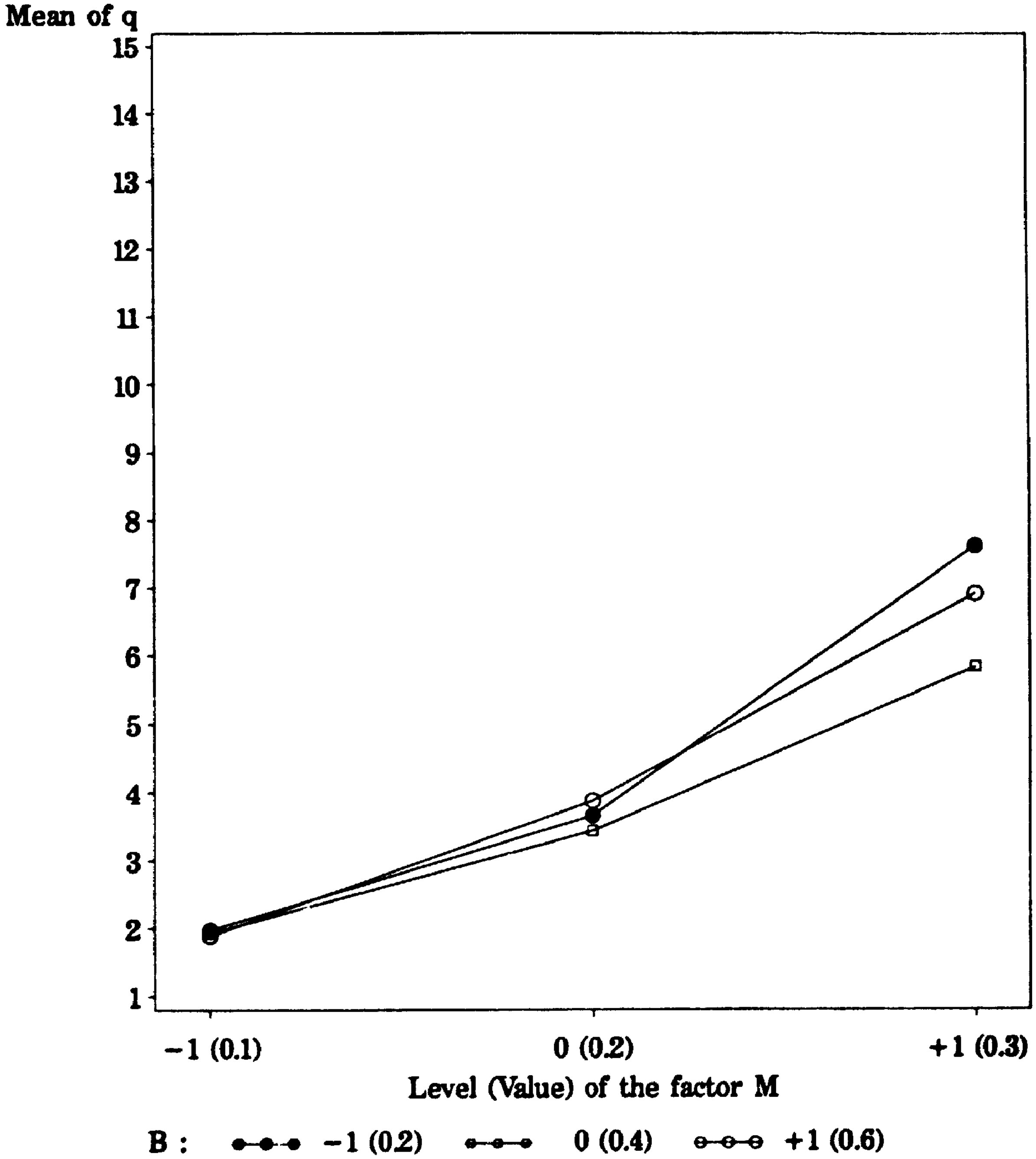


Figure 5.4.2.13
Effect of PM interaction on q
for the SCHm stopping criterion

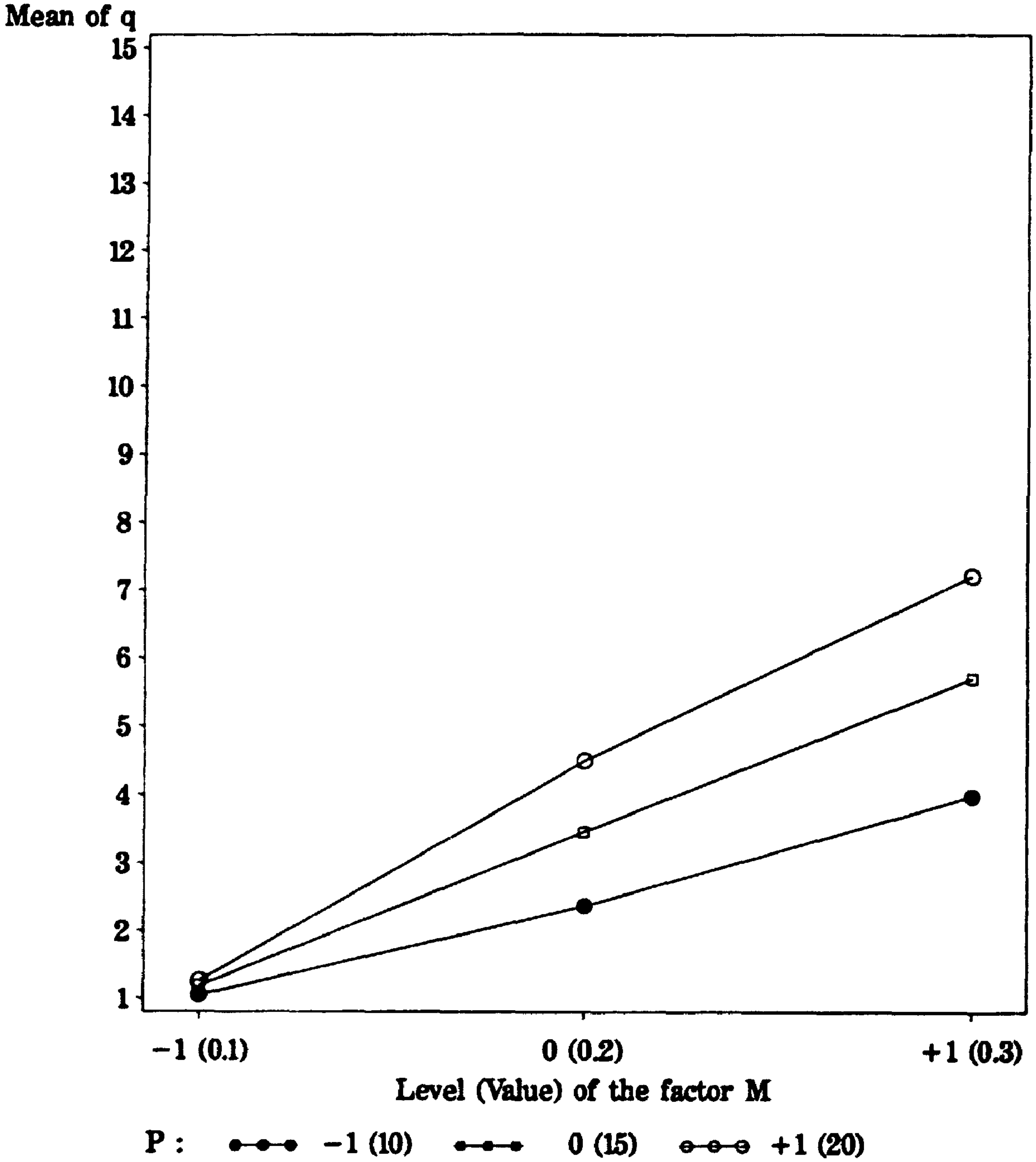


Figure 5.4.2.14

Effect of BM interaction on q for the SCHm stopping criterion

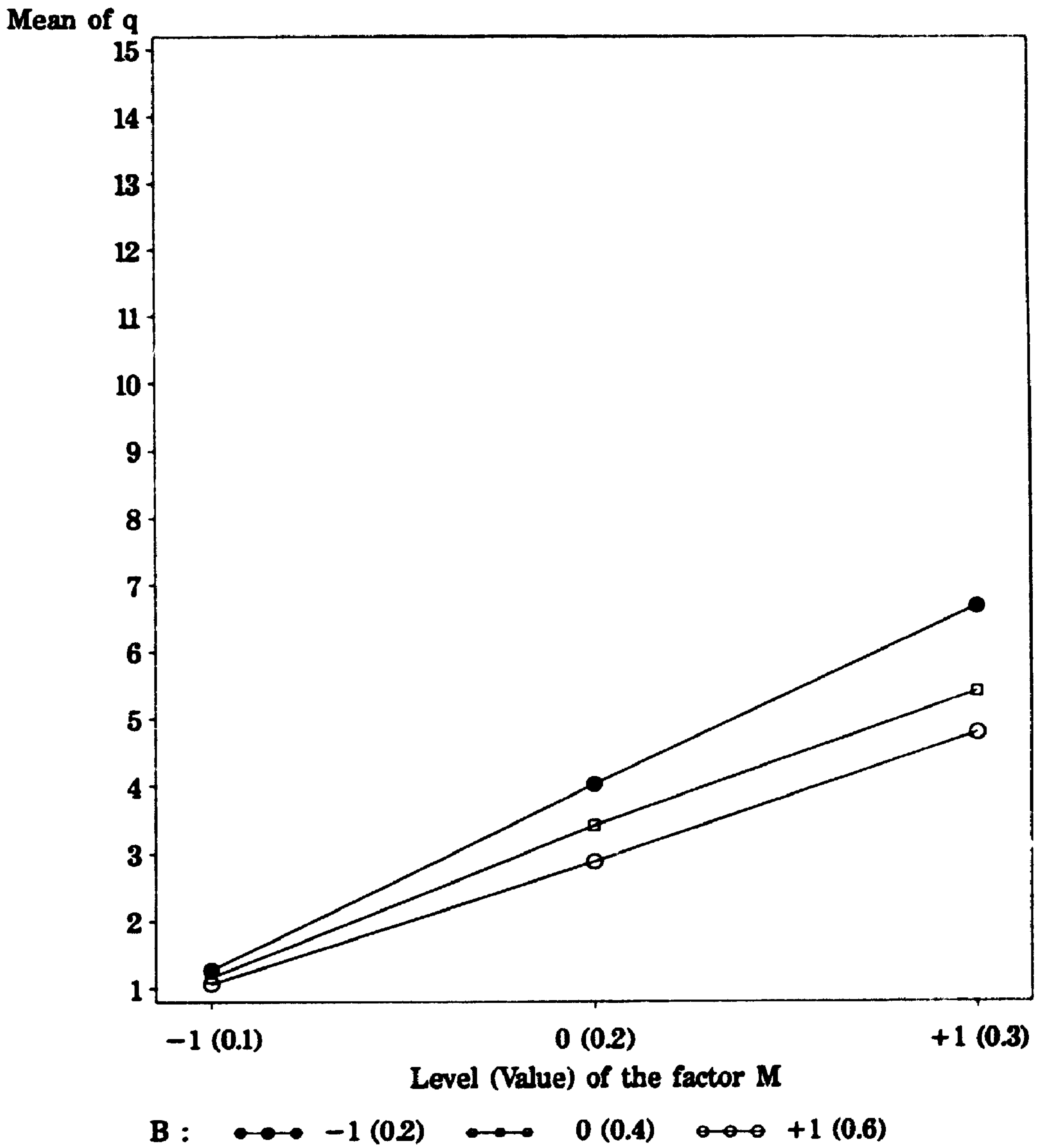
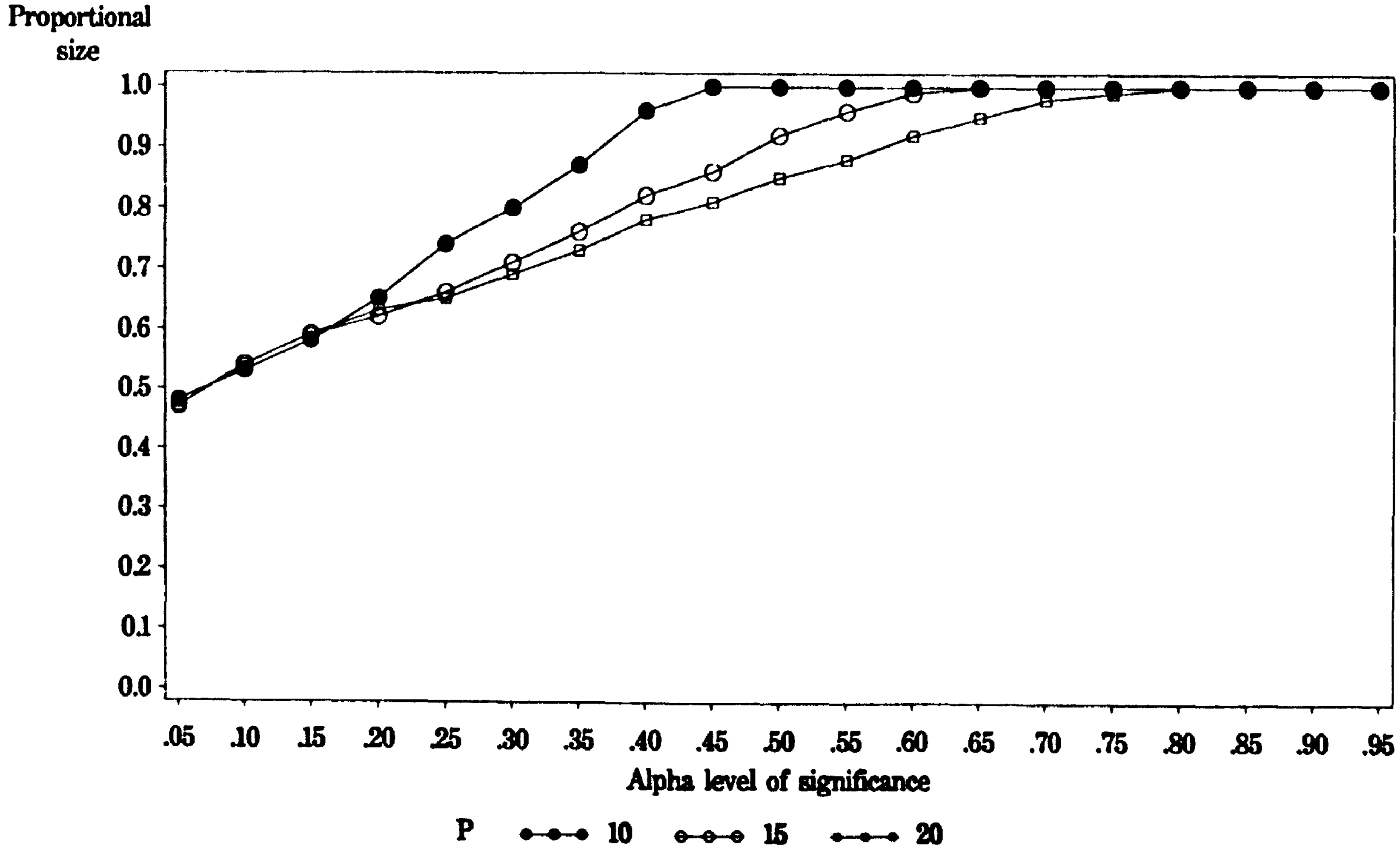


Figure 5.4.2.15

The proportional size of the logistic model over alpha level of significance for different values of P



5.5 Performance of selection and stopping criteria in terms of $\hat{ERR}$

The seven selection criteria LR, LS, WD, SC, PH, LK, and SW, and the five stopping criteria $\chi^2_{(\alpha)}$, E_m, C_{pm}, AIC_m, and SCH_m, have been defined in sections 3.2 and 3.3, respectively. For easy reference, let the symbols 'C' and 'S' denote the seven selection criteria and the five stopping criteria, respectively.

This section has three main purposes: 1) to examine the effects of the five between-subject factors P, V, Δ^2, D, and N and two within-subject factors C and S on $\hat{ERR}$ in the multivariate normal case, and of the four between-subject factors P, B, M, and N and two within-subject factors C and S on $\hat{ERR}$ in the multivariate binary case by using a repeated measures analysis of variance of $\sin^{-1}\hat{ERR}^{1/2}$; 2) to compare the performances of the seven selection criteria and the five stopping criteria in terms of $\hat{ERR}$ by using Bonferroni method of multiple-comparisons procedure; and 3) to examine the effects of the five factors P, V, Δ^2, D, and N on $\hat{ERR}$ in the multivariate normal case, and of the four factors P, B, M, and N on $\hat{ERR}$ in the multivariate binary case for the five stopping criteria in conjunction with the LR selection criterion (the reason for the selection of the LR criterion will be given in sections 5.5.1 and 5.5.2).

5.5.1 Multivariate normal case

Table 5.5.1.1 gives a summary for the results of overall mean of $\hat{ERR}$ for all 48 sampling situations with the seven selection criteria and the five stopping criteria. The overall mean of $\hat{ERR}$ ranges between 0.1520 to 0.1736 with standard error between 0.0019 to 0.0021 for all cases. The results of table 5.5.1.1 are graphically presented in figure 5.5.1.1. Table 5.5.1.1 will be analyzed in two different ways in the middle of this section by using Bonferroni method of multiple comparisons procedure.

In order to assess the effects of the five between-subject factors P, V, Δ^2, D, and N, and two within-subject factors C and S, a repeated measures analysis of variance of $\sin^{-1}\hat{ERR}^{1/2}$ was employed with the 32 factorial sampling situations. It is assumed in this study that all of the seven factors P, V, Δ^2, D, N, C, and S are fixed factors.

A repeated measures analysis of variance with five between-subject factors and two within-subject factors may be seen as a straightforward extension of the case of the five between-subject factors and one within-subject factor which has been used in section 5.4. However, when we consider a design with two within-subject factors, the problem becomes slightly more complicated due to the presence of additional error terms.

There are four error terms for this design. These four error terms will be labeled $Error_{between}$, $Error_{c-within}$, $Error_{s-within}$, and $Error_{cs-within}$. Each error term is separately used for a number of different tests. Specifically, $Error_{between}$ is used to test the between-subject effects. When it comes to the within-subject effects, C and its interaction with the five between-subject factors are tested by $Error_{c-within}$; S and its interaction with the five between-subject factors are tested by $Error_{s-within}$; and CS and its interaction with the five between-subject factors are tested by $Error_{cs-within}$.

Table 5.5.1.2 gives the results of the repeated measures of analysis of $\sin^{-1}\hat{ERR}^{1/2}$ for the five between-subject factors and two within-subject factors (note that table 5.5.1.2 contains four pages). The first page of table 5.5.1.2, shows that the main effect of the factor M is highly significant (F-ratio = 3142.50) and that the main effects of two factors P and D are moderately significant (F-ratio = 24.21 and 5.72, respectively). The results in this page are based on the 'overall' levels of the two within-subject factors C and S by combining all levels of these two factors.

The second page of table 5.5.1.2 gives the results for the effects which include the factor C. It is observed that the main effect of the factor C is not statistically significant. This result will be also seen in Bonferroni's multiple comparisons among the seven selection criteria. It is also observed that none of the interaction terms between the factor C and the five between-subject factors are statistically significant.

The third page of table 5.5.1.2 gives the results for the effects which include the factor S. It is noted that the main effect of the factor S is highly significant (F-ratio = 406.90). This result will be also seen in Bonferroni's multiple comparisons among the five stopping criteria. It is noted that the magnitudes of the F-ratio for the interaction terms are very small, although some are statistically significant, compared to the magnitude of the F-ratio for the main effect S. For instance, the magnitude of the F-ratio for the SN interaction which is most significant among all interaction terms was less than 5 percents of that for the main effect S (17.44 versus 406.90).

Lastly, the fourth page of table 5.5.1.2 gives the results for the effects which include the CS interaction. It is observed that the CS interaction is statistically significant (F-ratio = 3.69). This implies that we should examine the effects of the factor S at each level of the factor C. However, the magnitude of the F-ratio for the CS interaction is negligible, compared to that of the F-ratio for the factor S (3.69 versus 406.90). It is evident from figure 5.5.1.1 that the significant CS interaction is caused by the SW selection criterion.

Table 5.5.1.3 gives a summary of the results for the means of $\hat{ERR}$ over the levels of the five factors for the five stopping criteria in conjunction with the LR selection criterion. Figures 5.5.1.2 through 5.5.1.6 display the means of $\hat{ERR}$ over the levels of the five factors P, V, M, D, and N, respectively, for the five stopping criteria. These figures illustrate the previous findings in the repeated ANOVA in table 5.5.1.2.

Referring to table 5.5.1.1 again, it can be analyzed in two different ways by using Bonferroni method of multiple-comparisons procedure; one way is the comparisons of the seven selection criteria in which pairwise comparisons are made for the means of $E\hat{R}R$ for the seven selection criteria given a stopping criterion; and another way is the comparisons of the five stopping criteria in which pairwise comparisons are made for the means of $E\hat{R}R$ for the five stopping criteria given a selection criterion.

For the comparisons of the seven selection criteria, it is found that none of pairwise comparisons for the seven selection criteria for each of all five stopping criteria were statistically significant at $\alpha = 0.05$ level. The $\chi^2_{(0.20)}$ stopping criterion was chosen to illustrate the results of Bonferroni's multiple comparisons for the seven selection criteria. The first entry for each of the seven selection criteria in table 5.5.1.1 shows that the means of $E\hat{R}R$ for the seven selection criteria LR, LS, WD, SC, PH, LK, and SW for the $\chi^2_{(0.20)}$ stopping criterion are 0.1637, 0.1636, 0.1648, 0.1638, 0.1638, 0.1638, and 0.1642, respectively. In this case, the minimum significant difference, MSD, which any pair of means must differ by was 0.0088 with α=0.05, k=10, v=6713, $t_{(0.05,10,6713)}$=3.04, MSE=0.004055, and N_1 and N_2=960 (see section 5.4.1 for the definitions of these parameters). It is clear from these numbers that none of pairs of means are statistically significant at $\alpha = 0.05$ level. In figure 5.5.1.1 the horizontal line along with X-axis confirms this finding.

For the comparisons of the five stopping criteria, since the same results were found for all seven selection criteria, the LR selection criterion was chosen to illustrate the results of Bonferroni's multiple comparisons for the five stopping criteria.

Accordingly, we have, from the first five entries of the top in table 5.5.1.1, that the means of $E\hat{R}R$ for the five stopping criteria $\chi^2_{(0.20)}$, E_m, C_{pm}, AIC_m, and SCH_m for the LR selection criterion are 0.1637, 0.1520, 0.1669, 0.1633, and 0.1734, respectively. In this case, the minimum significant difference, MSD, which any pair of means must differ by was 0.0080 with α=0.05, k=10, v=4795, $t_{(0.05,10,4795)}$=2.81, MSE=0.003882, and N_1=N_2=960. The results of Bonferroni's multiple comparisons show that three groups of stopping criteria can be composed out of the five stopping criteria; $\{E_m\}$, $\{AIC_m, \chi^2_{(0.20)}, C_{pm}\}$, and $\{C_{pm}, SCH_m\}$. The interpretation of groups can be made such a way that any pairs within a group are not statistically significant, while any 'distinct' pairs between groups are statistically significant. We note that C_{pm} appears in two groups; that is the reason the word 'distinct' was used to differentiate pairs with C_{pm} from other pairs. Figure 5.5.1.1 illustrates these findings.

Based on the above results of Bonferroni's multiple comparisons and of the repeated measures ANOVA, it is not necessary to consider all selection criteria in order to examine the effects of the five factors P, V, Δ^2, D, and N on $E\hat{R}R$ for the five stopping criteria. Therefore, the LR selection criterion, which is the 'standard' criterion in forward stepwise logistic regression, was chosen to assess further the effects of the five factors P, V, Δ^2, D, and N on $E\hat{R}R$ for the five stopping criteria $\chi^2_{(0.20)}$, E_m, C_{pm}, AIC_m, and SCH_m.

In order to assess the effects of the five factors P, V, Δ^2, D, and N on $E\hat{R}R$,a response surface analysis was employed for each of the five stopping criteria. Tables 5.5.1.4 through 5.5.1.8 give the results of the response surface analyses of $\sin^{-1}E\hat{R}R^{1/2}$ for the five stopping criteria $\chi^2_{(0.20)}$, C_{pm}, AIC_m, E_m, and SCH_m, respectively. Due to many similarities of the results for the three stopping criteria, namely $\chi^2_{(0.20)}$, C_{pm}, and AIC_m, we discuss the results of these three stopping criteria together. The results of the E_m and SCH_m stopping criteria will be discussed separately in this order.

Before embarking on the major discussions of the results of the response surface analyses, one remark with regard to lack-of-fit test may be made. The lack-of-fit tests for the five stopping criteria have been found to be significant. However, we justify the use of the results of the response surface analyses with the reasons given in section 5.4.1.

Tables 5.5.1.4 through 5.5.1.6 for the $\chi^2_{(0.20)}$, C_{pm}, and AIC_m stopping criteria, respectively, show that variables P, M, and D are statistically significant, while variables V and N are not. Variables P and M have negative $\hat{\beta}$, while variable D has positive $\hat{\beta}$'s. The slopes for variables P, M, and D can be seen in figures 5.5.1.2, 5.5.1.4, and 5.5.1.5. Furthermore, the effects of variable D on $\sin^{-1}\hat{ERR}^{1/2}$ are found to be not only linear, but also quadratic. Equally striking are the results for variable N^2 (see convex curves in figure 5.5.1.6). Among all interaction terms only the PN interaction is statistically significant. Figures 5.5.1.7 through 5.5.1.9 for the $\chi^2_{(0.20)}$, C_{pm}, and AIC_m stopping criteria, respectively, show that there are crossover interactions. The interpretation of these PN interactions on $\hat{ERR}$ can made as follows: when P = 10 the smaller N (150) has a larger $\hat{ERR}$ than the larger N (250), while when P = 20 the larger N (250) has a larger $\hat{ERR}$ than the smaller N (150).

Table 5.5.1.5 for the E_m stopping criterion shows, in general, that the results are the same as for the $\chi^2_{(0.20)}$, C_{pm}, and AIC_m stopping criteria. There are two differences; one is the significant linear effect of variable N, and another is that the PN interaction is not crossed, but non-parallel (see figure 5.5.1.10).

Table 5.5.1.8 for the SCH_m stopping criterion shows, in general, the results the same as for the $\chi^2_{(0.20)}$, C_{pm}, and AIC_m stopping criteria except for three discrepancies: 1) variable V is significant (see figure 5.5.1.3); 2) the effect of variable P is reduced (see figure 5.5.1.2); and 3) the PN interaction is not significant.

5.5.2 Multivariate binary case

The tables and graphs for the multivariate binary case are analogous to those in the multivariate normal case.

In order to assess the effects of the four between-subject factors P, B, M, and N, and two within-subject factors C and S on $\hat{ERR}$, a repeated measures analysis of variance was employed with the 81 sampling situations.

Table 5.5.2.2 gives the results of the repeated measures analysis of $\sin^{-1}\hat{ERR}^{1/2}$ for the five between-subject factors and two within-subject factors. The first page of table 5.5.2.2 shows that the main effects of the factors P, B, and M are highly significant (F-ratio = 774.22, 297.57, and 6777.32, respectively) and that the main effect of the factor N is slightly significant (F-ratio = 5.87). However, the results of this part are of little practical importance because they are based on the 'overall' levels of the two within-subject factors C and S by collapsing all levels of these two factors.

The second page of table 5.5.2.2 shows the results for the effects which include the factor C. It is observed that the main effect of the factor C is moderately significant (F-ratio = 17.27). This result will be also seen in Bonferroni's multiple comparisons among the seven selection criteria; that is, there are differences in $\hat{ERR}$ between the SW selection criterion and other six selection criteria for the C_{pm} stopping criterion.

The third page of table 5.5.2.2 shows the results for the effects which include the factor S. It is observed that the main effect of the factor S is highly significant (F-ratio = 2431.61). This finding will be also seen in Bonferroni's multiple comparisons among the five stopping criteria. It is noted that the magnitudes of the F-ratio for the interaction terms are very small, although some are statistically significant, compared to the magnitude of the F-ratio for the main effect S. For instance, the F-ratio for the SP interaction which was the largest among all interaction terms was less than 10 percent of the F-ratio for the main effect S (232.30 versus 2431.61).

Lastly, the fourth page of table 5.5.2.2 shows the results for the effects which include the CS interaction. It is observed that the effect of the CS interaction is statistically significant (F-ratio = 61.18). The significant interaction CS implies that we should examine the effects of the factor S at each level of the factor C. It is evident in figure 5.5.2.1 that the significant interaction CS is caused by the SW selection criterion.

Table 5.5.2.3 gives a summary of the results for the means of $E\hat{R}R$ over the levels of the four factors for the five stopping criteria in conjunction with the LR selection criterion. Figures 5.5.2.2 through 5.5.2.4 display the means of $E\hat{R}R$ over the levels of the four factors P, B, M, and N, respectively, for the five stopping criteria

Given the results of table 5.5.2.1, one can analyze in two different ways by using Bonferroni method of multiple-comparisons procedure; one way is the comparisons of the seven selection criteria in which pairwise comparisons are made for the means of $E\hat{R}R$ for the seven selection criteria given a stopping criterion; and another way is the comparisons of the five stopping criteria in which pairwise comparisons are made for the means of $E\hat{R}R$ for the five stopping criteria given a selection criterion.

For the comparisons of the seven selection criteria given a stopping criterion, none of pairwise comparisons for the seven selection criteria for each of the five stopping criteria except for the C_{pm} stopping criterion were statistically significant at $\alpha = 0.05$ level. For Bonferroni's multiple comparisons among the seven selection criteria for the C_{pm} stopping criterion, the third entry of each of the seven selection criteria in table 5.5.2.1 gives the means of $E\hat{R}R$ for the seven selection criteria LR, LS, WD, SC, PH, LK, and SW; they are 0.2924, 0.2924, 0.2907, 0.2923, 0.2923, 0.2923, and 0.3049, respectively. In this case, the minimum significant difference, MSD, which any pair of means must differ by was 0.0105 with α=0.05, k=10, v=11333, $t_{(0.05,10,11333)}$=3.04, MSE=0.009635, and N_1=N_2=1620. It is found from these numbers that none of pairs of means are statistically significant at $\alpha = 0.05$ level. This finding is confirmed in figure 5.5.2.1 by the horizontal line along with X-axis.

For the comparisons of the five stopping criteria given a selection criterion, the same results were found for all six selection criteria except for the SW selection criterion. The LR selection criterion was chosen from the other selection criteria to illustrate the results of Bonferroni's multiple comparisons for the five stopping criteria. Accordingly, the first five entries of the top in table 5.5.2.1 give the means of $E\hat{R}R$ for the five stopping criteria $\chi^2_{(0.20)}$, E_m, C_{pm}, AIC_m, and SCH_m; they are 0.2558, 0.2405, 0.2924, 0.2556, and 0.2880, respectively. In this case, the minimum significant difference, MSD, which any pair of means must differ by was 0.0097 with α=0.05, k=10, v=8095, $t_{(0.05,10,8095)}$=2.81, MSE=0.009592, and N_1=N_2=1620. The results of Bonferroni's multiple comparisons show that three groups of stopping criteria can be composed out of the five stopping criteria; $\{E_m\}$, $\{AIC_m, \chi^2_{(0.15)}\}$, $\{SCH_m, C_{pm}\}$. The interpretation of the grouping was given in section 5.5.1. It was also examined for the SW selection criterion to see that there are four groups instead of three groups in such a way that C_{pm} itself is an additional group.

Finally, there is one interesting result; in terms of $\hat{ERR}$ among the five stopping criteria, SCH_m is the worst in the multivariate normal case (see figure 5.5.1.1), while C_{pm} is the worst the multivariate binary case (see figure 5.5.2.1).

Based on the above results of Bonferroni's multiple comparisons and of the repeated measures ANOVA, it is not necessary to consider all selection criteria in order to examine the effects of the four factors P, B, M, and N for the five stopping criteria. Therefore, the LR selection criterion, which is the 'standard' criterion in forward stepwise logistic regression, was chosen to assess further the effects of the four factors P, B, M, and N for the five stopping criteria $\chi^2_{(0.15)}$, E_m, C_{pm}, AIC_m, and SCH_m.

Tables 5.5.2.4 through 5.5.2.8 give the results of the analysis of variance of $\sin^{-1}\hat{ERR}^{1/2}$ for the five stopping criteria $\chi^2_{(0.20)}$, E_m, C_{pm}, AIC_m, and SCH_m, respectively. We note that the values of R^2 for tables 5.5.2.4 through 5.5.2.8 are 0.89, 0.90, 0.80, 0.89, and 0.88, respectively, which indicate that these ANOVAs fit the data well. The results are similar for all five stopping criteria.

First, although the main effects of the factors P, B, and M are statistically significant for all five stopping criteria, the magnitudes of the F-ratio are very different. The magnitudes of the F-ratio for the four main effects in descending order may be written as M > P > B > N.

Second, the PM and BM interactions are two most significant among all interaction terms in all five stopping criteria. Figures 5.5.2.5 through 5.5.2.15 show the effects of the PM and BM interactions on $\hat{ERR}$ for the five stopping criteria $\chi^2_{(0.15)}$, E_m, C_{pm}, AIC_m, and SCH_m, respectively. All these interactions were found to be slightly non-parallel.

5.5.3 Conclusions

1. There are no differences in $\hat{ERR}$ between the seven selection criteria for each of the five stopping criteria in the multivariate normal and multivariate binary cases.

2. There are differences in $\hat{ERR}$ between the five stopping criteria for each of the seven selection criteria in the multivariate normal and multivariate binary cases.

3. The ascending order of the overall means of $\hat{ERR}$ for the five stopping criteria is $E_m < AIC_m < \chi^2_{(0.20)} < C_{pm} < SCH_m$ in the multivariate normal case.

4. The ascending order of the overall means of $\hat{ERR}$ for the five stopping criteria is $E_m < AIC_m < \chi^2_{(0.15)} < SCH_m < C_m$ in the multivariate binary case.

5. $\{E_m\}$, $\{AIC_m, \chi^2_{(0.20)}, C_{pm}\}$, and $\{C_{pm}, SCH_m\}$ are three groups of stopping criteria which can be distinguished in terms of their effect on $\hat{ERR}$ in the multivariate normal case.

6. $\{E_m\}$, $\{AIC_m, \chi^2_{(0.15)}\}$, and $\{SCH_m, C_{pm}\}$ are three groups of stopping criteria which can be distinguished in terms of their effect on $\hat{ERR}$ in the multivariate binary case.

7. Δ^2 and M are the most influential factors on $\hat{ERR}$ in the multivariate normal and multivariate binary cases, respectively.

8. There is an interaction between P and N in their effect on $\hat{ERR}$ in the multivariate normal case.

9. There is an interaction between P and M in their effect on $\hat{ERR}$ in the multivariate binary case.

Table 5.5.1.1 Mean of $\hat{ERR}$ for all 48 sampling situations with the seven selection criteria and five stopping criteria		
Selection criteria	Stopping criteria	$\overline{\hat{ERR}}$ (s.e.)
LR	$\chi_{(0.20)}$	0.1637 (0.0020)
	E_m	0.1520 (0.0019)
	C_{pm}	0.1669 (0.0020)
	AIC_m	0.1633 (0.0020)
	SCH_m	0.1734 (0.0021)
LS	$\chi_{(0.20)}$	0.1636 (0.0020)
	E_m	0.1520 (0.0019)
	C_{pm}	0.1669 (0.0020)
	AIC_m	0.1633 (0.0020)
	SCH_m	0.1734 (0.0021)
WD	$\chi_{(0.20)}$	0.1648 (0.0020)
	E_m	0.1530 (0.0020)
	C_{pm}	0.1679 (0.0020)
	AIC_m	0.1643 (0.0021)
	SCH_m	0.17[illegible] (0.0021)
SC	$\chi_{(0.20)}$	0.1638 (0.0020)
	E_m	0.1521 (0.0019)
	C_{pm}	0.1669 (0.0020)
	AIC_m	0.1633 (0.0020)
	SCH_m	0.1736 (0.0020)
PH	$\chi_{(0.20)}$	0.1638 (0.0020)
	E_m	0.1521 (0.0019)
	C_{pm}	0.1669 (0.0020)
	AIC_m	0.1633 (0.0020)
	SCH_m	0.1736 (0.0020)
LK	$\chi_{(0.20)}$	0.1638 (0.0020)
	E_m	0.1521 (0.0019)
	C_{pm}	0.1669 (0.0020)
	AIC_m	0.1633 (0.0020)
	SCH_m	0.1736 (0.0020)
SW	$\chi_{(0.20)}$	0.1642 (0.0020)
	E_m	0.1522 (0.0019)
	C_{pm}	0.1678 (0.0020)
	AIC_m	0.1632 (0.0020)
	SCH_m	0.1734 (0.0020)

Table 5.5.1.2
Repeated measures analysis of variance of $\sin^{-1}\widehat{ERR}^{1/2}$
Test of Between Subject Effect

Effect	d.f.	Mean Square	F-ratio	p-value
[a]P	1	1.19940	24.21	0.00
V	1	0.03400	0.69	0.41
[a]M	1	155.66440	3142.50	0.00
[b]D	1	0.28312	5.72	0.02
N	1	0.10710	2.16	0.14
PV	1	0.00266	0.05	0.82
PM	1	0.00939	0.19	0.66
PD	1	0.00083	0.02	0.90
[b]PN	1	0.24373	4.92	0.03
VM	1	0.00035	0.01	0.93
VD	1	0.03510	0.71	0.40
VN	1	0.03573	0.72	0.40
MD	1	0.01088	0.22	0.64
MN	1	0.00016	0.00	0.95
DN	1	0.00011	0.00	0.96
PVM	1	0.00280	0.06	0.81
PVD	1	0.09722	1.96	0.16
[b]PVN	1	0.19548	3.95	0.05
PMD	1	0.04822	0.97	0.32
PMN	1	0.03904	0.79	0.38
PDN	1	0.00232	0.05	0.83
VMD	1	0.01698	0.34	0.56
VMN	1	0.05954	1.20	0.27
VDN	1	0.10027	2.02	0.16
MDN	1	0.00150	0.03	0.86
[b]PVMD	1	0.37973	7.67	0.01
PVMN	1	0.03645	0.74	0.39
PVDN	1	0.09792	1.98	0.16
PMDN	1	0.06713	1.36	0.24
VMDN	1	0.00177	0.04	0.85
PVMDN	1	0.15505	3.13	0.08
Error(between)	608	0.04954		

[a]significant with $p < 0.001$
[b]significant with $p < 0.05$

Table 5.5.1.2 (continued) Repeated measures analysis of variance of $\sin^{-1}\hat{ERR}^{1/2}$ Test of C Within Subject Effect				
Effect	d.f.	Mean Square	F-ratio	p–value[c]
C	6	0.00020	0.27	0.62
CP	6	0.00118	1.61	0.21
CV	6	0.00006	0.09	0.79
CM	6	0.00038	0.52	0.48
CD	6	0.00154	2.10	0.15
CN	6	0.00023	0.31	0.59
CPV	6	0.00053	0.63	0.40
CPM	6	0.00180	2.47	0.11
CPD	6	0.00001	0.01	0.93
CPN	6	0.00045	0.61	0.45
CVM	6	0.00003	0.04	0.86
CVD	6	0.00066	0.90	0.35
CVN	6	0.00005	0.07	0.81
CMD	6	0.00060	0.81	0.38
CMN	6	0.00013	0.18	0.69
CDN	6	0.00243	3.32	0.07
CPVM	6	0.00040	0.55	0.47
CPVD	6	0.00020	0.28	0.61
CPVN	6	0.00051	0.70	0.41
CPMD	6	0.00021	0.29	0.61
CPMN	6	0.00107	1.46	0.23
CPDN	6	0.00003	0.05	0.85
CVMD	6	0.00059	0.80	0.38
CVMN	6	0.00016	0.22	0.66
CVDN	6	0.00140	1.92	0.17
CMDN	6	0.00006	0.08	0.79
CPVMD	6	0.00115	1.58	0.21
CPVMN	6	0.00045	0.61	0.45
CPVDN	6	0.00004	0.05	0.84
CPMDN	6	0.00005	0.06	0.82
CVMDN	6	0.00023	0.32	0.59
CPVMDN	6	0.00121	1.66	0.20
Error(C within)	3648	0.00073		

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

[c]Greenhouse-Geisser's method

Table 5.5.1.2 (continued)
Repeated measures analysis of variance of $\sin^{-1}\hat{ERR}^{1/2}$
Test of S Within Subject Effect

Effect	d.f.	Mean Square	F-ratio	p-value[c]
[a]S	4	0.57546	406.90	0.00
[a]SP	4	0.02070	14.64	0.00
[a]SV	4	0.01574	11.13	0.00
SM	4	0.00363	2.57	0.06
SD	4	0.00271	1.92	0.13
[a]SN	4	0.02467	17.44	0.00
SPV	4	0.00110	0.77	0.49
[b]SPM	4	0.00451	3.19	0.03
SPD	4	0.00241	1.70	0.17
SPN	4	0.00263	1.86	0.14
SVM	4	0.00191	1.35	0.26
[b]SVD	4	0.00495	3.50	0.02
SVN	4	0.00060	0.42	0.71
SMD	4	0.00009	0.06	0.97
SMN	4	0.00193	1.36	0.25
SDN	4	0.00133	0.94	0.41
SPVM	4	0.00014	0.10	0.94
SPVD	4	0.00065	0.46	0.68
SPVN	4	0.00116	0.82	0.47
SPMD	4	0.00112	0.79	0.48
SPMN	4	0.00300	2.12	0.10
[b]SPDN	4	0.00471	3.33	0.02
SVMD	4	0.00075	0.53	0.64
SVMN	4	0.00164	1.16	0.32
SVDN	4	0.00112	0.79	0.48
SMDN	4	0.00022	0.15	0.91
SPVMD	4	0.00148	1.05	0.36
SPVMN	4	0.00070	0.50	0.66
SPVDN	4	0.00068	0.48	0.67
SPMDN	4	0.00008	0.06	0.97
SVMDN	4	0.00369	2.61	0.06
SPVMDN	4	0.00183	1.29	0.28
Error(S within)	2432	0.00141		

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

[c]Greenhouse-Geisser's method

Table 5.5.1.2 (continued) Repeated measures analysis of variance of $\sin^{-1}\hat{ERR}^{1/2}$ Test of CS Within Subject Effect				
Effect	d.f.	Mean Square	F-ratio	p–value[c]
[a]CS	24	0.00015	3.69	0.00
[a]CSP	24	0.00014	3.23	0.00
CSV	24	0.00006	1.40	0.21
[a]CSM	24	0.00016	3.71	0.00
[b]CSD	24	0.00010	2.30	0.03
[b]CSN	24	0.00012	2.93	0.01
CSPV	24	0.00005	1.18	0.31
[b]CSPM	24	0.00012	2.83	0.01
CSPD	24	0.00005	1.15	0.33
CSPN	24	0.00005	1.26	0.27
CSVM	24	0.00008	1.97	0.06
CSVD	24	0.00002	0.60	0.74
CSVN	24	0.00005	1.09	0.37
CSMD	24	0.00003	0.72	0.64
[b]CSMN	24	0.00012	2.97	0.01
CSDN	24	0.00005	1.08	0.37
CSPVM	24	0.00002	0.40	0.88
CSPVD	24	0.00004	0.96	0.45
CSPVN	24	0.00002	0.43	0.86
CSPMD	24	0.00004	1.01	0.42
CSPMN	24	0.00005	1.27	0.27
CSPDN	24	0.00009	2.15	0.04
CSVMD	24	0.00003	0.70	0.65
CSVMN	24	0.00004	0.87	0.52
CSVDN	24	0.00005	1.13	0.34
CSMDN	24	0.00002	0.55	0.78
CSPVMD	24	0.00002	0.44	0.85
CSPVMN	24	0.00004	0.97	0.44
CSPVDN	24	0.00003	0.83	0.55
CSPMDN	24	0.00004	1.03	0.40
CSVMDN	24	0.00003	0.78	0.59
[b]CSPVMDN	24	0.00010	2.39	0.03
Error(CS within)	14592	0.00004		

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

[c]Greenhouse-Geisser's method

Table 5.5.1.3
Mean of $\hat{ERR}$ for the five stopping criteria with the LR selection criterion over the levels of the five factors P, V, M, D, and N

Factor	[a]Level	Value	Stopping criteria				
			$\chi^2_{(0.20)}$	E_m	C_{pm}	AIC_m	SCH_m
P	-2	5	0.165	0.159	0.165	0.163	0.172
	-1	10	0.169	0.158	0.170	0.168	0.177
	0	15	0.164	0.153	0.167	0.164	0.173
	+1	20	0.158	0.145	0.163	0.158	0.170
	+2	25	0.168	0.152	0.174	0.166	0.175
V	-2	0.2	0.159	0.148	0.157	0.159	0.167
	-1	0.4	0.163	0.151	0.166	0.163	0.170
	0	0.6	0.165	0.153	0.168	0.164	0.173
	+1	0.8	0.163	0.152	0.167	0.163	0.177
	+2	1.0	0.169	0.158	0.175	0.170	0.182
M	-2	1.0	0.298	0.285	0.308	0.298	0.315
	-1	1.5	0.224	0.210	0.227	0.223	0.234
	0	2.0	0.162	0.151	0.165	0.162	0.171
	+1	2.5	0.103	0.093	0.107	0.103	0.112
	+2	3.0	0.065	0.057	0.065	0.064	0.069
D	-2	0.2	0.148	0.138	0.152	0.148	0.156
	-1	0.4	0.161	0.149	0.164	0.161	0.170
	0	0.6	0.161	0.150	0.164	0.160	0.170
	+1	0.8	0.166	0.154	0.169	0.165	0.177
	+2	1.0	0.229	0.216	0.228	0.230	0.239
N	-2	100	0.152	0.131	0.156	0.149	0.162
	-1	150	0.162	0.148	0.166	0.162	0.174
	0	200	0.166	0.155	0.169	0.166	0.175
	+1	250	0.164	0.155	0.167	0.164	0.173
	+2	300	0.156	0.147	0.158	0.156	0.162
Overall			0.164	0.152	0.167	0.163	0.173

[a]The number of sampling situations for each level is 1 for levels ±2, 16 for levels ±1, and 14 for level 0.

Table 5.5.1.4
Response surface analysis of $\sin^{-1}\hat{ERR}^{1/2}$
for the $\chi^2_{(0.20)}$ stopping criterion

Variable	d.f.	$\hat{\beta}$	s.e.	t-value	p-value
[a]P	1	-0.006224	0.001459	-4.27	0.000
V	1	0.000366	0.001459	0.25	0.802
[a]M	1	-0.083353	0.001459	-57.14	0.000
[a]D	1	0.008060	0.001459	5.53	0.000
N	1	0.001869	0.001459	1.28	0.201
P^2	1	0.000676	0.001673	0.40	0.687
V^2	1	0.000054	0.001673	0.03	0.975
M^2	1	0.000113	0.001673	0.07	0.946
[a]D^2	1	0.007519	0.001673	4.49	0.000
[b]N^2	1	-0.003598	0.001673	-2.15	0.032
PV	1	-0.000247	0.001631	-0.15	0.880
PM	1	-0.001012	0.001631	-0.62	0.535
PD	1	0.000464	0.001631	0.28	0.776
[b]PN	1	0.003894	0.001631	2.39	0.017
VM	1	-0.001186	0.001631	-0.73	0.467
VD	1	-0.001633	0.001631	-1.00	0.317
VN	1	0.001670	0.001631	1.02	0.306
MD	1	0.000660	0.006131	0.41	0.686
MN	1	0.000030	0.006131	0.02	0.985
DN	1	0.000297	0.006131	0.18	0.855

[a]significant with $p < 0.001$.

[b]significant with $p < 0.05$.

Lack-of-fit test: $F_{22,917} = 4.043$ ($p < 0.001$)

Table 5.5.1.5
Response surface analysis of $\sin^{-1}\hat{ERR}^{1/2}$
for the C_{pm} stopping criterion

Variable	d.f.	$\hat{\beta}$	s.e.	t-value	p-value
[b]P	1	-0.003521	0.001443	-2.44	0.015
V	1	0.002015	0.001443	1.40	0.163
[a]M	1	-0.082826	0.001443	-57.39	0.000
[a]D	1	0.007835	0.001443	5.43	0.000
N	1	0.001045	0.001443	0.72	0.469
P^2	1	0.000924	0.001655	0.56	0.577
V^2	1	-0.000089	0.001655	-0.05	0.957
M^2	1	0.000576	0.001655	0.34	0.732
[a]D^2	1	0.007071	0.001655	4.27	0.000
[b]N^2	1	-0.003441	0.001655	-2.08	0.040
PV	1	-0.000156	0.001613	-0.10	0.923
PM	1	-0.000480	0.001613	-0.30	0.766
PD	1	-0.000611	0.001613	-0.38	0.705
[b]PN	1	0.003326	0.001613	2.06	0.040
VM	1	0.000037	0.001613	0.02	0.982
VD	1	-0.001589	0.001613	-0.99	0.325
VN	1	0.001310	0.001613	0.81	0.417
MD	1	0.000249	0.001613	0.16	0.877
MN	1	0.000180	0.001613	0.11	0.911
DN	1	0.000513	0.001613	0.32	0.751

[a]significant with $p < 0.001$.

[b]significant with $p < 0.05$.

Lack-of-fit test: $F_{22,917} = 4.111$ ($p < 0.001$)

Table 5.5.1.6
Response surface analysis of $\sin^{-1}\hat{ERR}^{1/2}$
for the AIC_m stopping criterion

Variable	d.f.	$\hat{\beta}$	s.e.	t-value	p-value
[a]P	1	-0.006493	0.001504	-4.32	0.000
V	1	0.001244	0.001504	0.83	0.408
[a]M	1	-0.083678	0.001504	-55.66	0.000
[a]D	1	0.007913	0.001504	5.26	0.000
N	1	0.002466	0.001504	1.64	0.101
P^2	1	0.000395	0.001725	0.23	0.819
V^2	1	0.000369	0.001725	0.21	0.830
M^2	1	-0.000114	0.001725	-0.07	0.947
[a]D^2	1	0.007792	0.001725	4.52	0.000
[b]N^2	1	-0.004024	0.001725	-2.33	0.020
PV	1	-0.000491	0.001681	-0.29	0.770
PM	1	-0.001524	0.001681	-0.91	0.365
PD	1	0.000013	0.001681	0.01	0.994
[b]PN	1	0.004425	0.001681	2.63	0.009
VM	1	0.000153	0.001681	0.09	0.927
VD	1	-0.002820	0.001681	-1.68	0.094
VN	1	0.001555	0.001681	0.93	0.355
MD	1	0.000553	0.001681	0.33	0.742
MN	1	0.000560	0.001681	0.33	0.739
DN	1	0.000618	0.001681	0.37	0.713

[a]significant with $p < 0.001$.

[b]significant with $p < 0.05$.

Lack-of-fit test: $F_{22,917} = 4.100$ ($p < 0.001$)

Table 5.5.1.7
Response surface analysis of $\sin^{-1}\hat{ERR}^{1/2}$
for the E_m stopping criterion

Variable	d.f.	$\hat{\beta}$	s.e.	t-value	p-value
[a]P	1	-0.008460	0.001450	-5.83	0.000
V	1	0.001352	0.001450	0.93	0.352
[a]M	1	-0.083445	0.001450	-57.54	0.000
[a]D	1	0.007944	0.001450	5.48	0.000
[a]N	1	0.006002	0.004150	4.14	0.000
P^2	1	0.000818	0.001664	0.49	0.623
V^2	1	-0.000108	0.001664	-0.07	0.948
M^2	1	0.000258	0.001664	-0.16	0.877
[a]D^2	1	0.007538	0.001664	4.53	0.000
[b]N^2	1	-0.004942	0.001664	-2.97	0.003
PV	1	-0.000861	0.001621	-0.53	0.596
PM	1	-0.000776	0.001621	-0.48	0.632
PD	1	0.000481	0.001621	0.30	0.767
[b]PN	1	0.003639	0.001621	2.24	0.025
VM	1	0.000206	0.001621	0.13	0.899
VD	1	-0.001434	0.001621	-0.89	0.377
VN	1	0.000967	0.001621	0.60	0.551
MD	1	0.000537	0.001621	0.33	0.741
MN	1	0.000341	0.001621	0.21	0.833
DN	1	0.000446	0.006121	0.28	0.783

[a]significant with $p < 0.001$.

[b]significant with $p < 0.05$.

Lack-of-fit test: $F_{22,917} = 4.163$ ($p < 0.001$)

Table 5.5.1.8
Response surface analysis of $\sin^{-1}\hat{ERR}^{1/2}$
for the SCH_m stopping criterion

Variable	d.f.	$\hat{\beta}$	s.e.	t-value	p-value
[b]P	1	-0.003502	0.001364	-2.57	0.010
[b]V	1	0.004581	0.001364	3.36	0.001
[a]M	1	-0.082385	0.001364	-60.41	0.000
[a]D	1	0.009207	0.001364	6.75	0.000
N	1	-0.000180	0.001364	-0.13	0.895
P^2	1	0.000712	0.001564	0.46	0.649
V^2	1	0.000837	0.001564	0.54	0.593
M^2	1	0.000825	0.001564	0.53	0.598
[a]D^2	1	0.007824	0.001564	5.00	0.000
[b]N^2	1	-0.003240	0.001564	-2.07	0.039
PV	1	0.000751	0.001525	0.49	0.622
PM	1	0.001228	0.001525	0.81	0.421
PD	1	-0.001301	0.001525	-0.85	0.394
PN	1	0.002224	0.001525	1.46	0.145
VM	1	0.000906	0.001525	0.59	0.553
VD	1	0.000041	0.001525	0.03	0.979
VN	1	0.000666	0.001525	0.44	0.662
MD	1	0.000449	0.001525	0.29	0.769
MN	1	-0.000766	0.001525	-0.50	0.615
DN	1	-0.000803	0.001525	-0.53	0.599

[a]significant with $p < 0.001$.

[b]significant with $p < 0.05$.

Lack-of-fit test: $F_{22,917} = 3.812$ ($p < 0.001$)

Figure 5.5.1.1

Mean of estimated ERR for all 48 sampling situations with the seven selection criteria and five stopping criteria

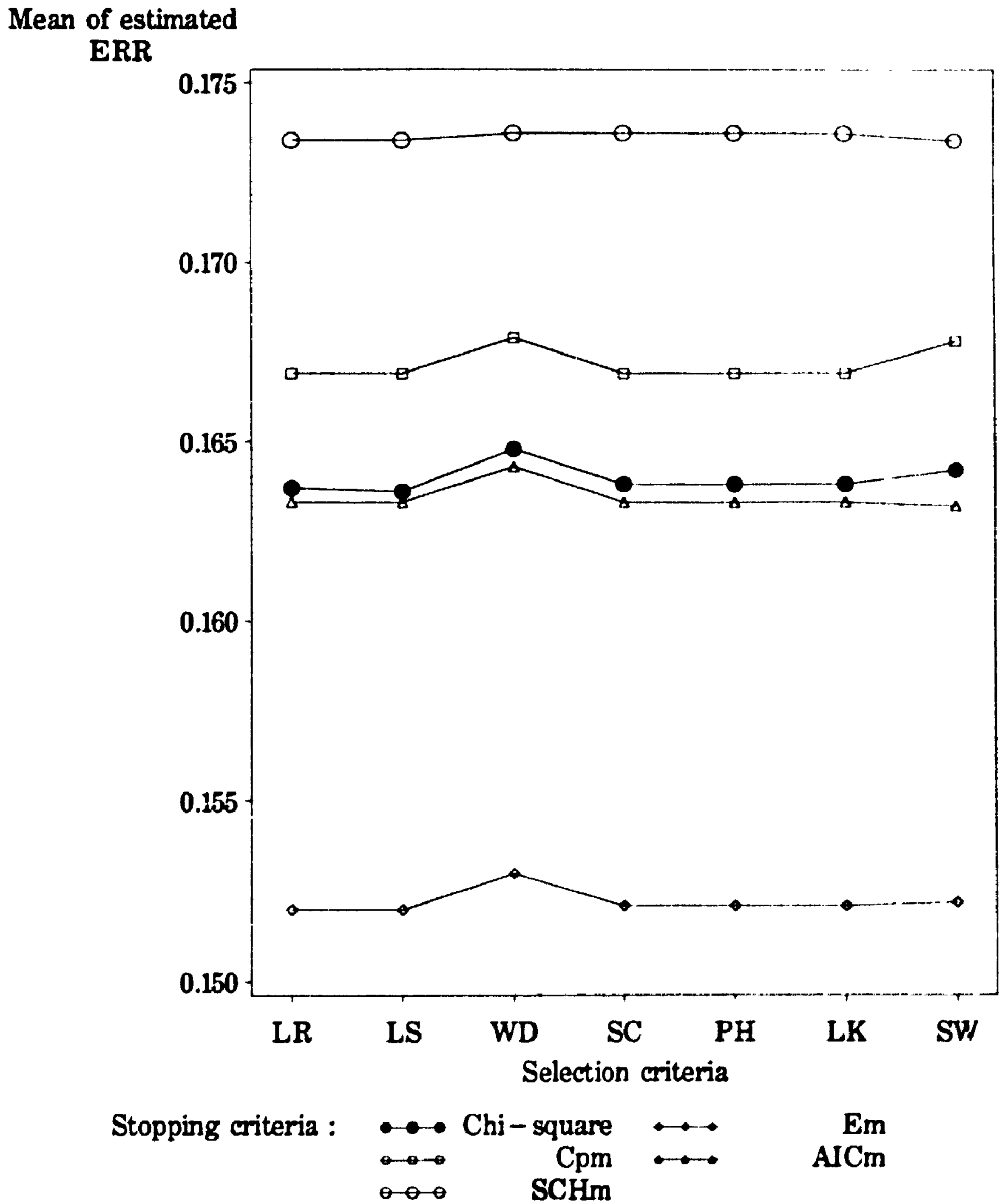


Figure 5.5.1.2

Mean of estimated ERR for the five stopping criteria over the levels of the factor P

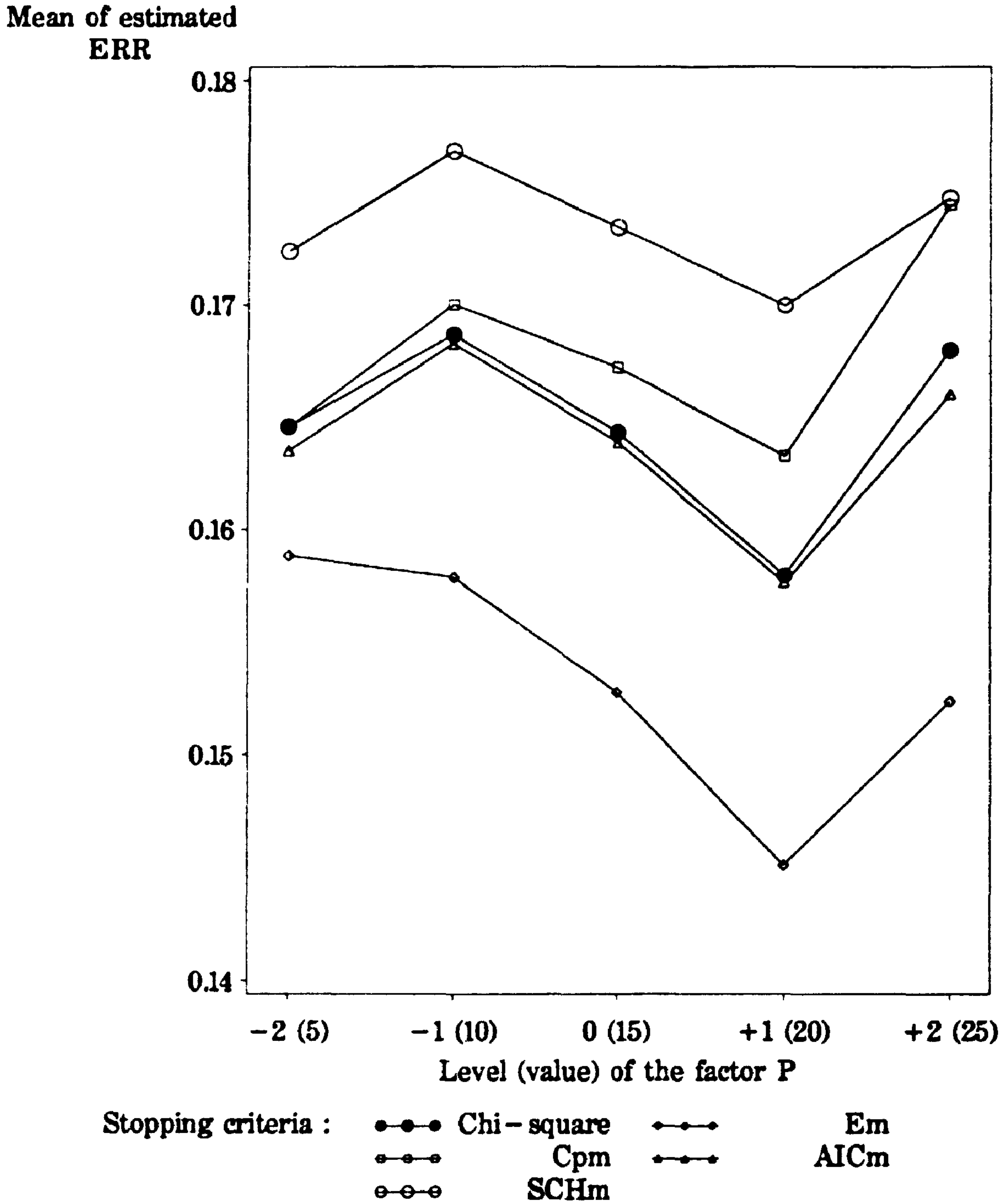


Figure 5.5.1.3

Mean of estimated ERR for the five stopping criteria over the levels of the factor V

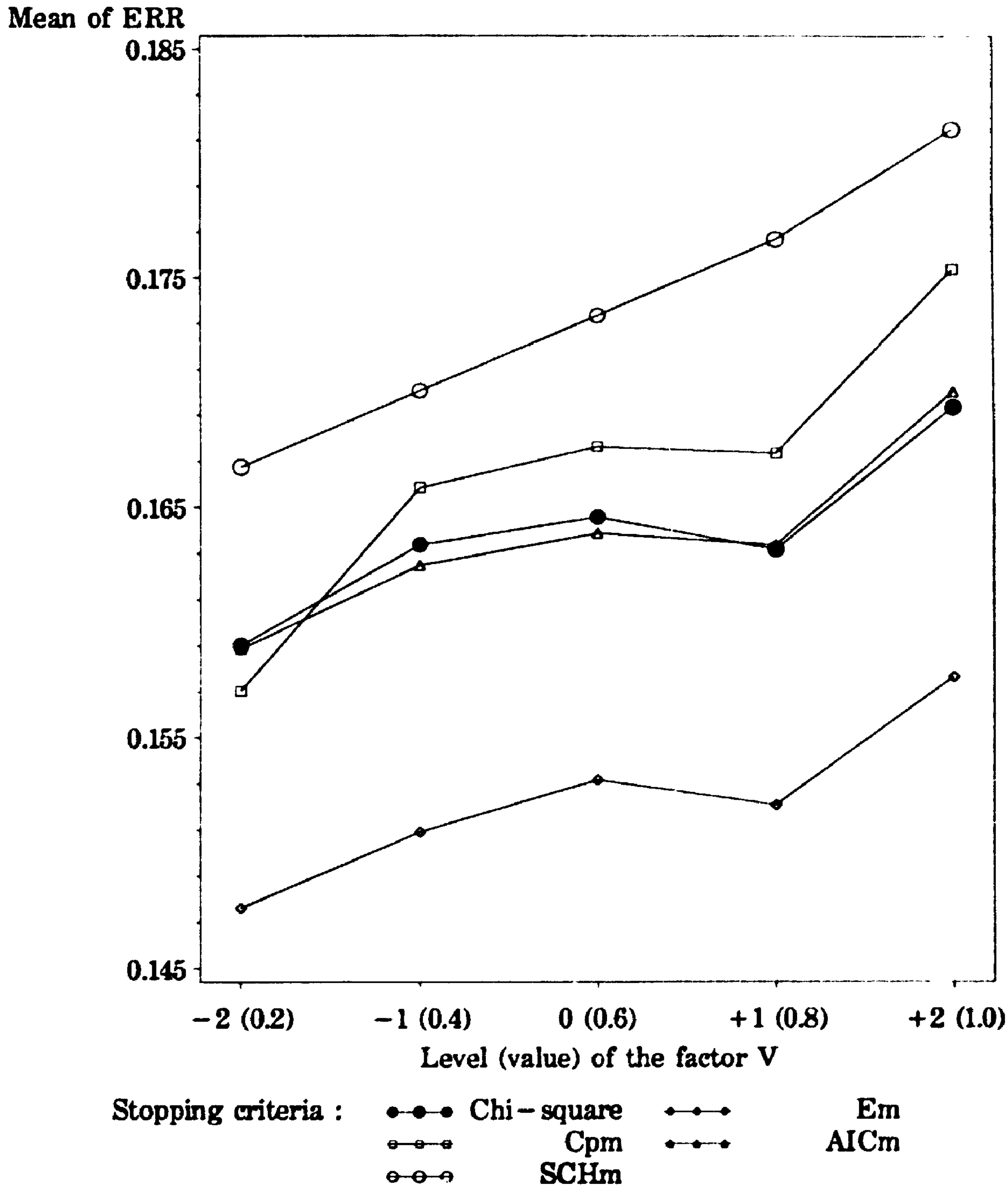


Figure 5.5.1.4

Mean of estimated ERR for the five stopping criteria over the levels of the factor M

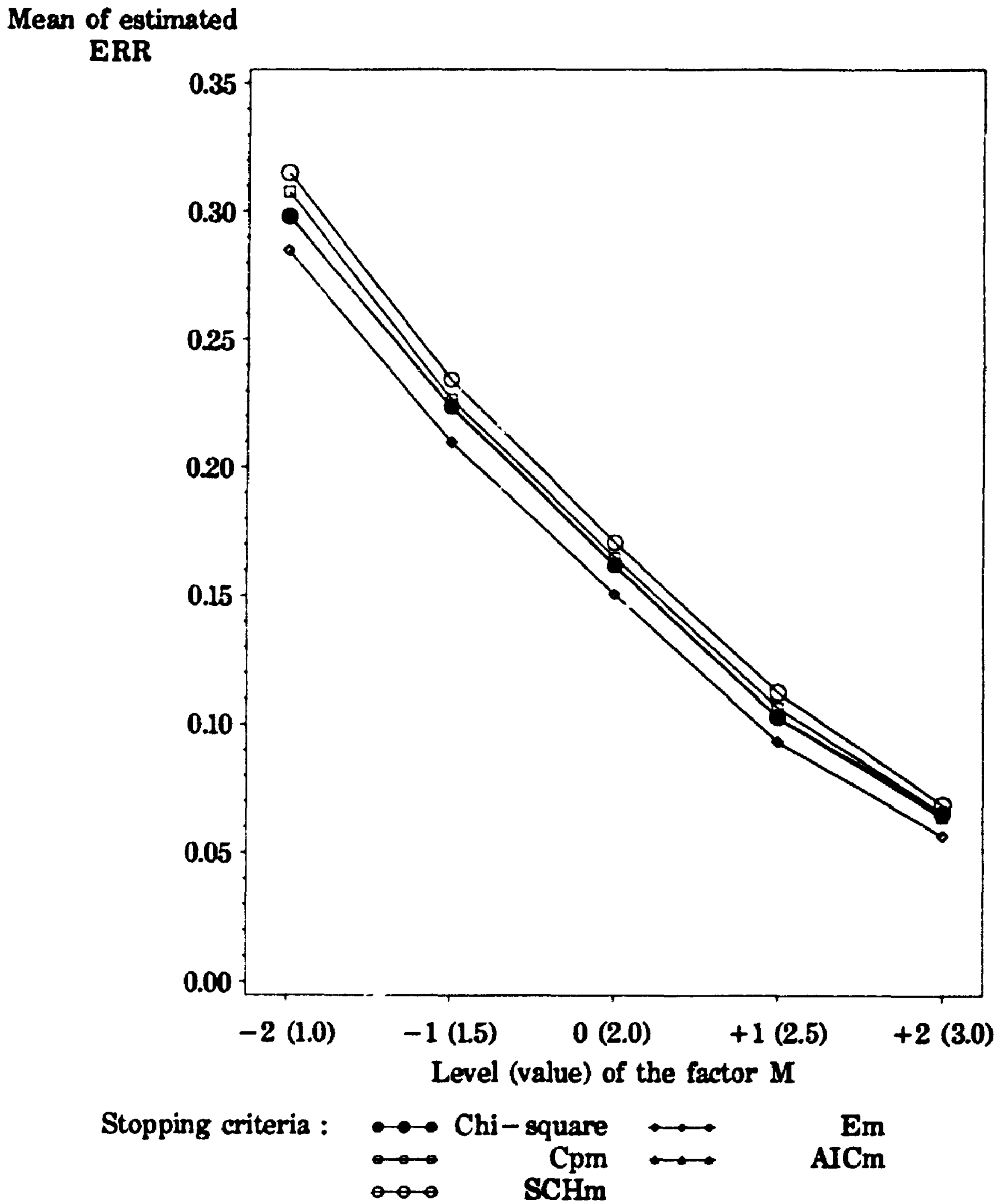


Figure 5.5.1.5

Mean of estimated ERR for the five stopping criteria over the levels of the factor D

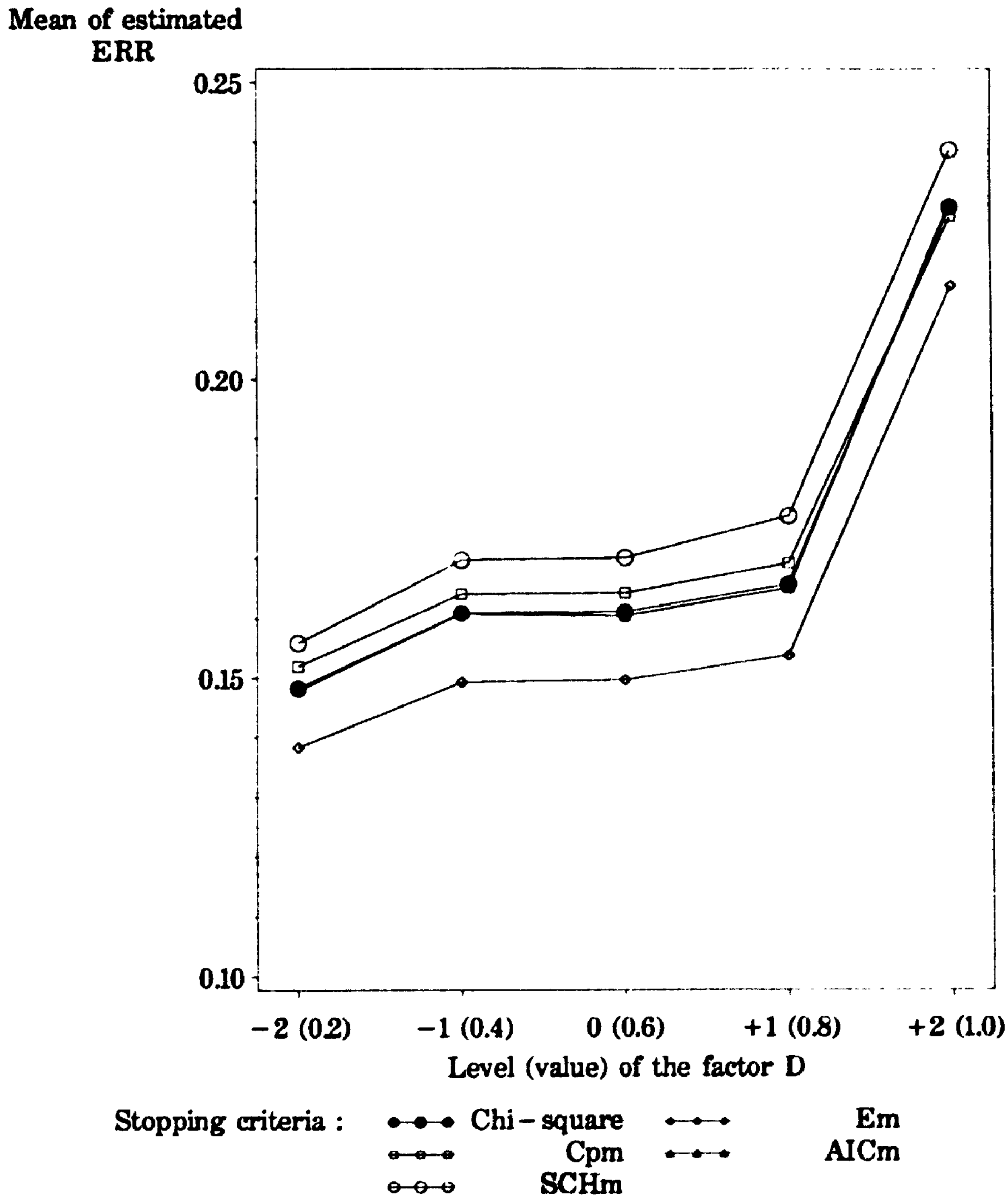


Figure 5.5.1.6

Mean of estimated ERR for the five stopping criteria over the levels of the factor N

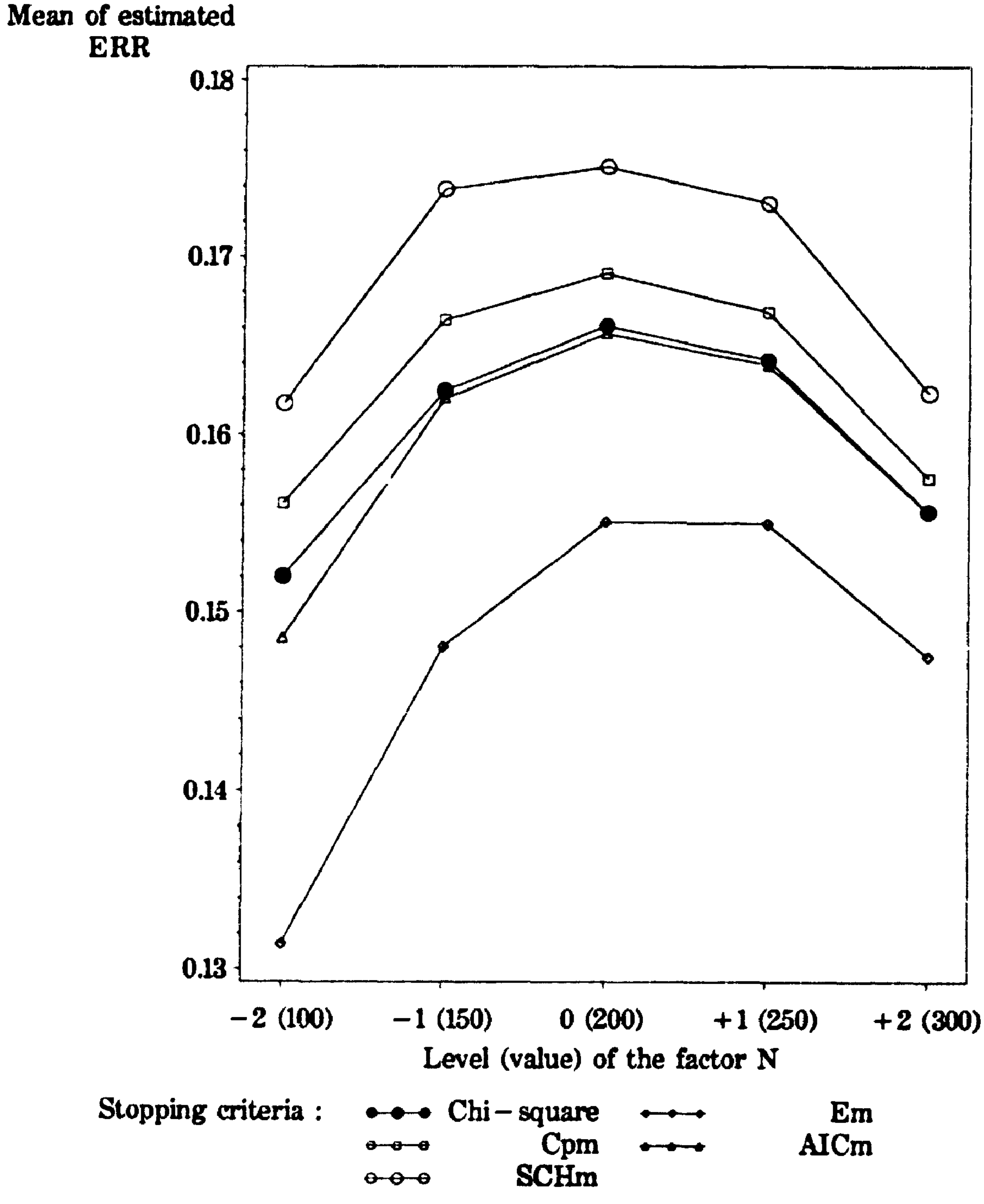


Figure 5.5.1.7

Effect of PN interaction on estimated ERR for the Chi-square stopping criterion

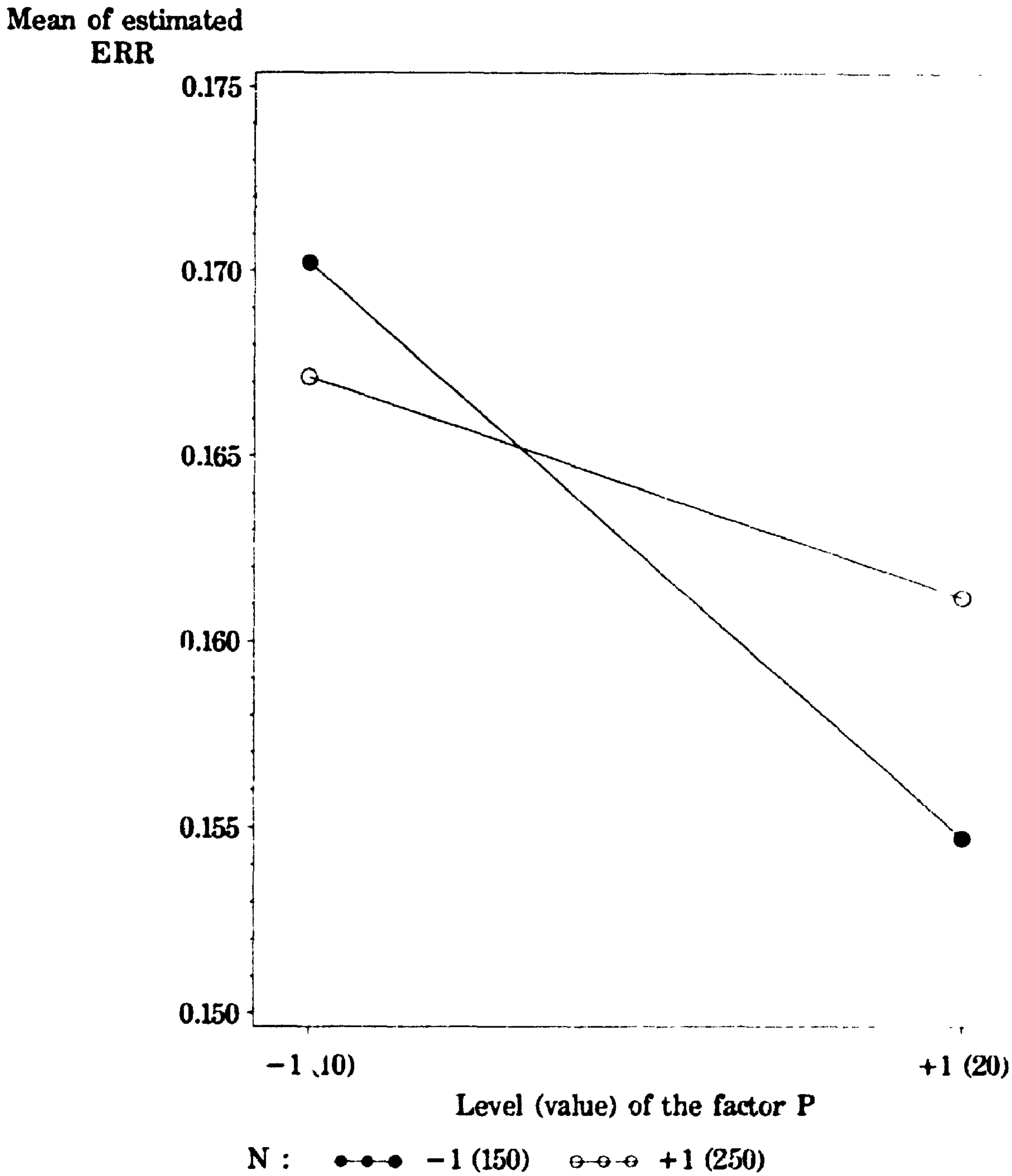


Figure 5.5.1.8

Effect of PN interaction on estimated ERR for the Cpm stopping criterion

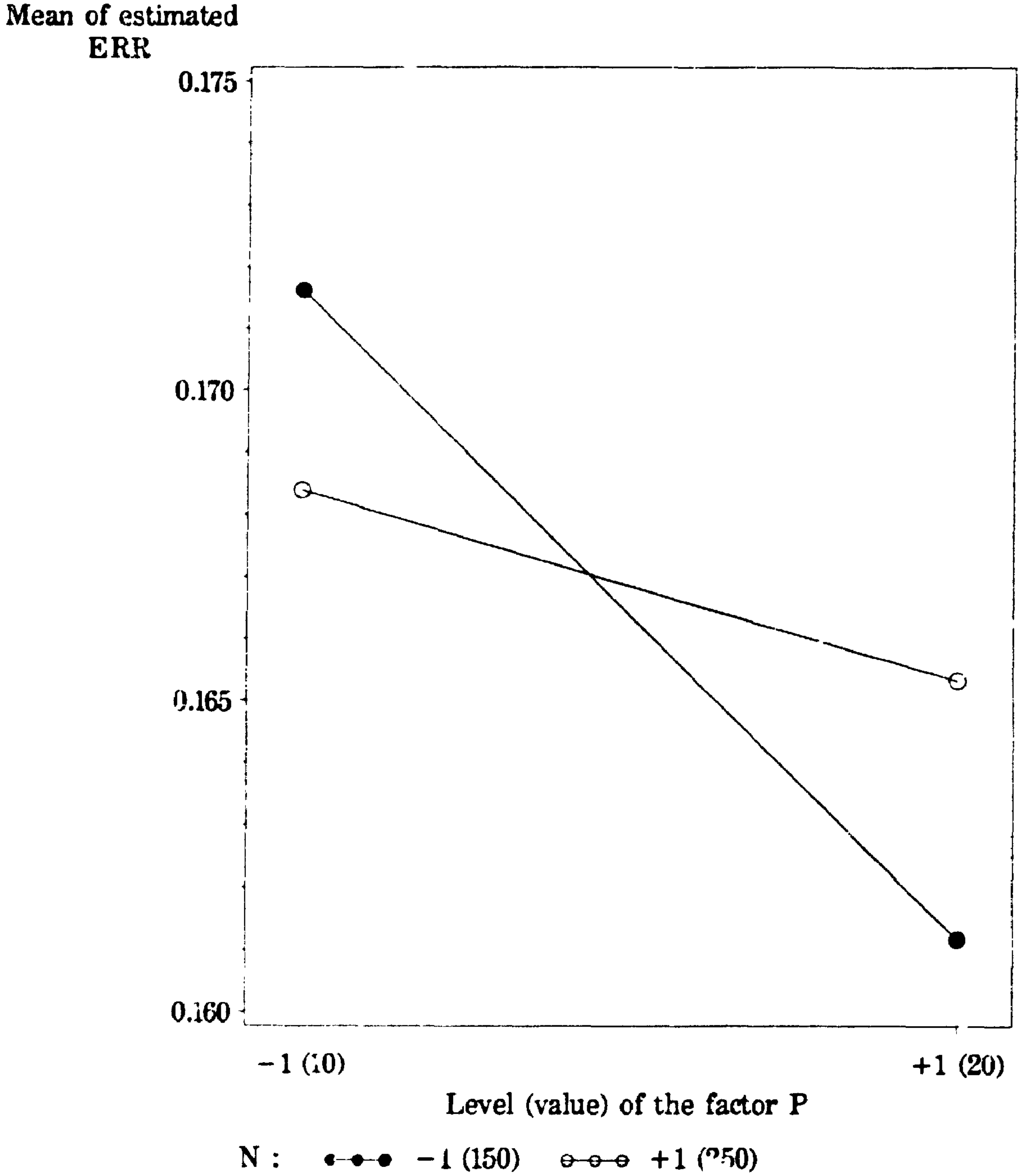


Figure 5.5.1.9

Effect of PN interaction on estimated ERR for the AICm stopping criterion

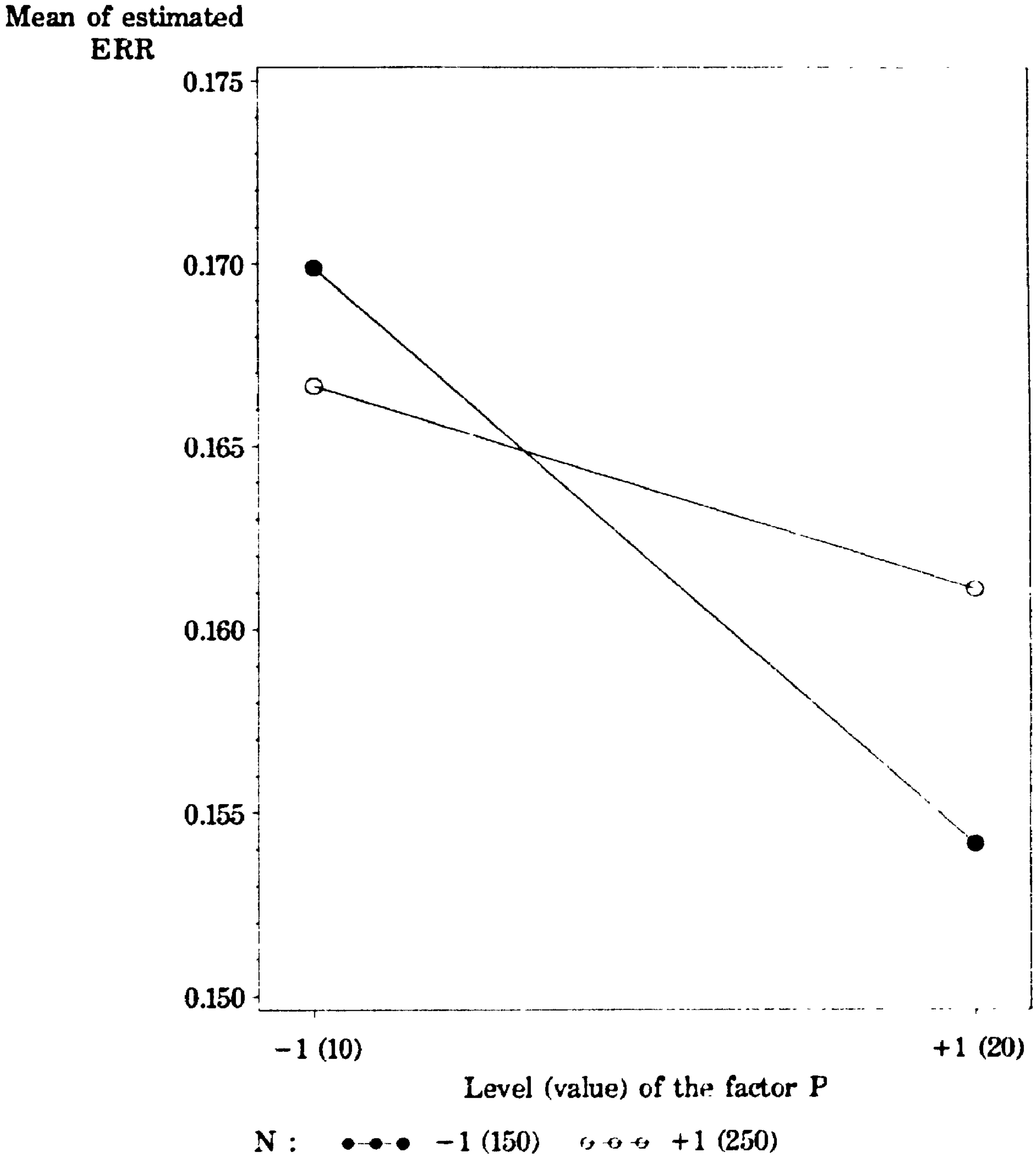


Figure 5.5.1.10

Effect of PN interaction on estimated ERR for the Em stopping criterion

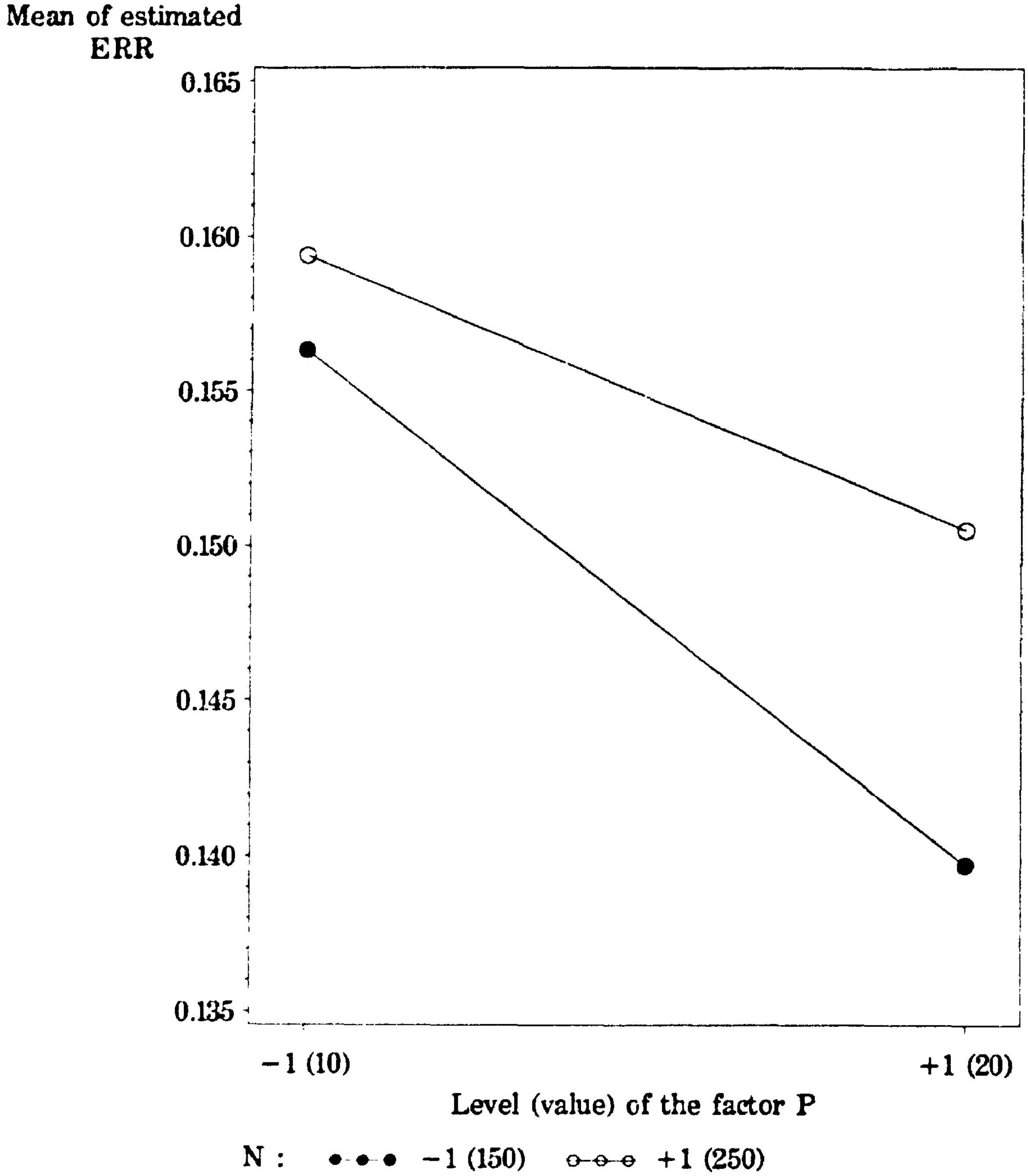


Table 5.5.2.1
Mean of $\hat{ERR}$ for all 81 sampling situations
with the seven selection criteria and five stopping criteria

Selection criteria	Stopping criteria	$\overline{\hat{ERR}}$ (s.e.)
LR	$\chi_{(0.15)}$	0.2558 (0.0024)
	E_m	0.2405 (0.0024)
	C_{pm}	0.2924 (0.0024)
	AIC_m	0.2556 (0.0025)
	SCH_m	0.2880 (0.0025)
LS	$\chi_{(0.15)}$	0.2558 (0.0025)
	E_m	0.2405 (0.0024)
	C_{pm}	0.2924 (0.0024)
	AIC_m	0.2556 (0.0025)
	SCH_m	0.2880 (0.0025)
WD	$\chi_{(0.15)}$	0.2544 (0.0024)
	E_m	0.2396 (0.0023)
	C_{pm}	0.2907 (0.0024)
	AIC_m	0.2544 (0.0024)
	SCH_m	0.2881 (0.0024)
SC	$\chi_{(0.15)}$	0.2557 (0.0025)
	E_m	0.2406 (0.0024)
	C_{pm}	0.2923 (0.0024)
	AIC_m	0.2556 (0.0025)
	SCH_m	0.2881 (0.0025)
PH	$\chi_{(0.15)}$	0.2557 (0.0025)
	E_m	0.2406 (0.0024)
	C_{pm}	0.2923 (0.0024)
	AIC_m	0.2556 (0.0025)
	SCH_m	0.2881 (0.0025)
LK	$\chi_{(0.15)}$	0.2557 (0.0025)
	E_m	0.2406 (0.0024)
	C_{pm}	0.2923 (0.0024)
	AIC_m	0.2556 (0.0025)
	SCH_m	0.2881 (0.0025)
SW	$\chi_{(0.15)}$	0.2571 (0.0024)
	E_m	0.2407 (0.0024)
	C_{pm}	0.3049 (0.0022)
	AIC_m	0.2556 (0.0024)
	SCH_m	0.2887 (0.0025)

Table 5.5.2.2
Repeated measures analysis of variance of $\sin^{-1}\hat{ERR}^{1/2}$
Test of Between Subject Effect

Effect	d.f.	Mean Square	F-ratio	p-value
[a]P	2	31.99652	774.22	0.0001
[a]B	2	12.29789	297.57	0.0001
[a]M	2	280.08922	6777.32	0.0001
[b]N	2	0.24267	5.87	0.0029
[a]PB	4	0.54034	13.07	0.0001
[a]PM	4	2.20732	53.41	0.0001
[b]PN	4	0.15139	3.66	0.0056
[a]BM	4	1.96944	47.65	0.0001
BN	4	0.02742	0.66	0.6174
MN	4	0.08575	2.07	0.0817
[a]PBM	8	0.21415	5.18	0.0001
PBN	8	0.06734	1.63	0.1116
PMN	8	0.07009	1.70	0.0947
BMN	8	0.03341	0.81	0.5952
PBMN	16	0.05941	1.44	0.1155
Error(between)	1539	0.04133		

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

Table 5.5.2.2 (continued)
Repeated measures analysis of variance of $\sin^{-1}\mathrm{ERR}^{1/2}$
Test of C Within Subject Effect

Effect	d.f.	Mean Square	F-ratio	[c]p–value
[a]C	6	0.01993	17.27	0.0001
[a]CP	12	0.00893	7.73	0.0002
[b]CB	12	0.00357	3.10	0.0369
[a]CM	12	0.01000	8.66	0.0001
CN	12	0.00190	1.65	0.1865
CPB	24	0.00199	1.72	0.1296
[a]CPM	24	0.00534	4.63	0.0004
CPN	24	0.00199	1.72	0.1306
CBM	24	0.00202	1.75	0.1242
CBN	24	0.00092	0.79	0.5485
CMN	24	0.00201	1.74	0.1253
CPBM	48	0.00133	1.15	0.3217
CPBN	48	0.00115	1.00	0.4416
CPMN	48	0.00114	0.99	0.4469
CBMN	48	0.00108	0.93	0.4977
CPBMN	96	0.00050	0.43	0.9848
Error(C within)	9234	0.00115		

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

[c]Greenhouse-Geisser's method

Table 5.5.2.2 (continued)
Repeated measures analysis of variance of $\sin^{-1}\hat{ERR}^{1/2}$
Test of S Within Subject Effect

Effect	d.f.	Mean Square	F-ratio	[c]p–value
[a]S	4	8.49238	2431.61	0.0001
[a]SP	8	0.81129	232.30	0.0001
[a]SB	8	0.38033	108.90	0.0001
[a]SM	8	0.09590	27.46	0.0001
[a]SN	8	0.04214	12.07	0.0001
[a]SPB	16	0.02809	8.04	0.0001
[a]SPM	16	0.02907	8.32	0.0001
[b]SPN	16	0.00850	2.43	0.0145
[a]SBM	16	0.04835	13.85	0.0001
[b]SBN	16	0.01159	3.32	0.0011
[a]SMN	16	0.01990	5.70	0.0001
[a]SPBM	32	0.01206	3.45	0.0001
[b]SPBN	32	0.00712	2.04	0.0099
[a]SPMN	32	0.01117	3.20	0.0001
[a]SBMN	32	0.00903	2.59	0.0007
[b]SPBMN	64	0.00551	1.58	0.0233
Error(S within)	6156	0.00349		

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

[c]Greenhouse-Geisser's method

Table 5.5.2.2 (continued)
Repeated measures analysis of variance of $\sin^{-1}\hat{ERR}^{1/2}$
Test of CS Within Subject Effect

Effect	d.f.	Mean Square	F-ratio	[c]p–value
[a]CS	24	0.01043	61.18	0.0001
[a]CSP	48	0.00476	27.91	0.0001
[a]CSB	48	0.00149	8.76	0.0001
[a]CSM	48	0.00400	23.45	0.0001
CSN	48	0.00033	1.92	0.0553
[a]CSPB	96	0.00068	4.01	0.0001
[a]CSPM	96	0.00214	12.57	0.0001
[b]CSPN	96	0.00037	2.17	0.0049
[a]CSBM	96	0.00067	3.91	0.0001
[b]CSBN	96	0.00035	2.04	0.0092
[b]CSMN	96	0.00028	1.66	0.0489
[a]CSPBM	192	0.00038	2.21	0.0001
[b]CSPBN	192	0.00025	1.49	0.0380
[b]CSPMN	912	0.00032	1.91	0.0018
CSBMN	192	0.00020	1.19	0.2138
CSPBMN	384	0.00021	1.22	0.1186
Error(CS within)	36936	0.00017		

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

[c]Greenhouse-Geisser's method

Table 5.5.2.3
Mean of $\hat{ERR}$ for the five stopping criteria with the LR selection criterion over the levels of the four factors P, B, M, and N

Factor	[a]Level	Value	Stopping criteria				
			$\chi^2_{(0.15)}$	E_m	C_{pm}	AIC_m	SCH_m
P	-1	10	0.297	0.281	0.309	0.297	0.314
	0	15	0.255	0.240	0.294	0.254	0.288
	+1	20	0.215	0.200	0.274	0.215	0.262
B	-1	0.2	0.229	0.216	0.279	0.228	0.270
	0	0.4	0.262	0.246	0.304	0.262	0.297
	+1	0.6	0.277	0.260	0.294	0.277	0.297
M	-1	0.1	0.359	0.341	0.391	0.359	0.397
	0	0.2	0.254	0.238	0.299	0.254	0.291
	+1	0.3	0.154	0.142	0.187	0.154	0.176
N	-1	150	0.252	0.235	0.291	0.252	0.289
	0	200	0.258	0.242	0.295	0.257	0.289
	+1	250	0.258	0.244	0.291	0.258	0.286
Overall			0.256	0.241	0.292	0.256	0.288

[a]The number of sampling situations for each level is 27.

Table 5.5.2.4
Analysis of variance of $\sin^{-1}\hat{ERR}^{1/2}$
for the $\chi^2_{(0.15)}$ stopping criterion

Effect	d.f.	Mean Square	F-ratio	p-value
[a]P	2	1.42674	858.41	0.0001
[a]B	2	0.56637	340.76	0.0001
[a]M	2	8.15630	4907.34	0.0001
[a]N	2	0.01674	10.07	0.0001
[a]PB	4	0.02375	14.29	0.0001
[a]PM	4	0.07633	45.93	0.0001
[b]PN	4	0.00589	3.54	0.0070
[a]BM	4	0.07430	44.70	0.0001
BN	4	0.00070	0.42	0.7950
MN	4	0.00138	0.83	0.5066
[a]PBM	8	0.00725	4.36	0.0001
PBN	8	0.00231	1.39	0.1968
PMN	8	0.00228	1.37	0.2051
BMN	8	0.00090	0.54	0.8261
PBMN	16	0.00199	1.20	0.2624
Error	1539	0.00166		

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

$R^2 = 0.89$

Table 5.5.2.5
Analysis of variance of $\sin^{-1}\hat{ERR}^{1/2}$
for the E_{in} stopping criterion

Effect	d.f.	Mean Square	F-ratio	p-value
[a]P	2	1.46023	974.42	0.0001
[a]B	2	0.48424	323.14	0.0001
[a]M	2	8.09955	5404.86	0.0001
[a]N	2	0.02524	16.84	0.0001
[a]PB	4	0.02532	16.90	0.0001
[a]PM	4	0.06420	42.84	0.0001
[a]PN	4	0.00939	6.27	0.0001
[a]BM	4	0.07184	47.94	0.0001
BN	4	0.00031	0.21	0.9332
MN	4	0.00177	1.18	0.3186
[a]PBM	8	0.00794	5.30	0.0001
PBN	8	0.00184	1.23	0.2767
[b]PMN	8	0.00324	2.16	0.0278
BMN	8	0.00123	0.82	0.5830
PBMN	16	0.00115	0.77	0.7232
Error	1539	0.00150		

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

$R^2 = 0.90$

Table 5.5.2.6
Analysis of variance of $\sin^{-1}\hat{ERR}^{1/2}$
for the C_{pm} stopping criterion

Effect	d.f.	Mean Square	F-ratio	p-value
[a]P	2	0.29158	103.37	0.0001
[a]B	2	0.14352	50.88	0.0001
[a]M	2	7.56462	2681.86	0.0001
N	2	0.00741	2.63	0.0727
PB	4	0.00226	0.80	0.5251
[a]PM	4	0.06669	23.64	0.0001
[b]PN	4	0.00709	2.51	0.0399
[a]BM	4	0.04007	14.21	0.0001
[b]BN	4	0.00950	3.37	0.0094
[a]MN	4	0.01846	6.54	0.0001
[a]PBM	8	0.01064	3.77	0.0002
[b]PBN	8	0.00791	2.80	0.0044
[a]PMN	8	0.01016	3.60	0.0004
[b]BMN	8	0.00771	2.73	0.0054
[a]PBMN	16	0.00736	2.61	0.0005
Error	1539	0.00282		

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

$R^2 = 0.80$

Table 5.5.2.7
Analysis of variance of $\sin^{-1}\hat{ERR}^{1/2}$
for the AIC_m stopping criterion

Effect	d.f.	Mean Square	F-ratio	p-value
[a]P	2	1.44208	855.30	0.0001
[a]B	2	0.57586	341.54	0.0001
[a]M	2	8.19469	4860.29	0.0001
[a]N	2	0.01547	9.17	0.0001
[a]PB	4	0.02581	15.31	0.0001
[a]PM	4	0.07739	45.90	0.0001
[b]PN	4	0.00653	3.88	0.0039
[a]BM	4	0.07634	45.28	0.0001
BN	4	0.00060	0.36	0.8405
MN	4	0.00144	0.86	0.4901
[a]PBM	8	0.00704	4.17	0.0001
PBN	8	0.00230	1.37	0.2071
PMN	8	0.00297	1.76	0.0806
BMN	8	0.00087	0.51	0.8472
PBMN	16	0.00183	1.08	0.3648
Error	1539	0.00169		

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

$R^2 = 0.89$

Table 5.5.2.8
Analysis of variance of $\sin^{-1}\hat{ERR}^{1/2}$
for the $SCH_{(a)}$ stopping criterion

Effect	d.f.	Mean Square	F-ratio	p-value
[a]P	2	0.54530	324.00	0.0001
[a]B	2	0.22072	131.15	0.0001
[a]M	2	8.67600	5155.04	0.0001
N	2	0.00160	0.95	0.3868
[a]PB	4	0.02287	13.59	0.0001
[a]PM	4	0.06479	38.50	0.0001
PN	4	0.00086	0.51	0.7288
[a]BM	4	0.06254	37.16	0.0001
BN	4	0.00281	1.67	0.1543
[b]MN	4	0.00469	2.78	0.0254
[a]PBM	8	0.00750	4.46	0.0001
PBN	8	0.00123	0.73	0.6650
PMN	8	0.00109	0.65	0.7385
BMN	8	0.00088	0.52	0.8420
PBMN	16	0.00143	0.85	0.6264
Error	1539	0.00168		

[a]significant with $p < 0.001$

[b]significant with $p < 0.05$

$R^2 = 0.88$

Figure 5.5.2.1

Mean of estimated ERR for all 81 sampling situations with the seven selection criteria and five stopping criteria

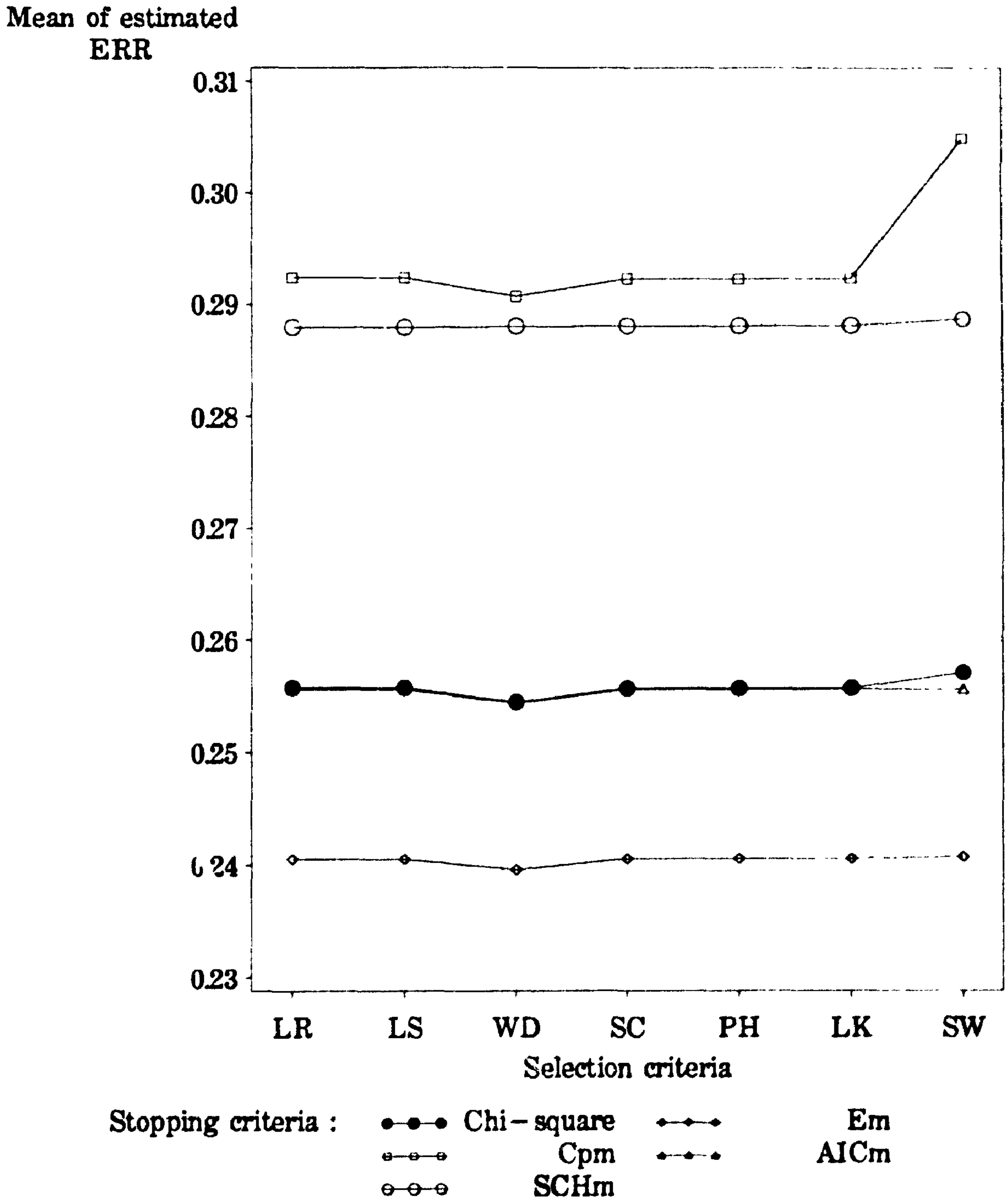


Figure 5.5.2.2

Mean of estimated ERR for the five stopping criteria over the levels of the factor P

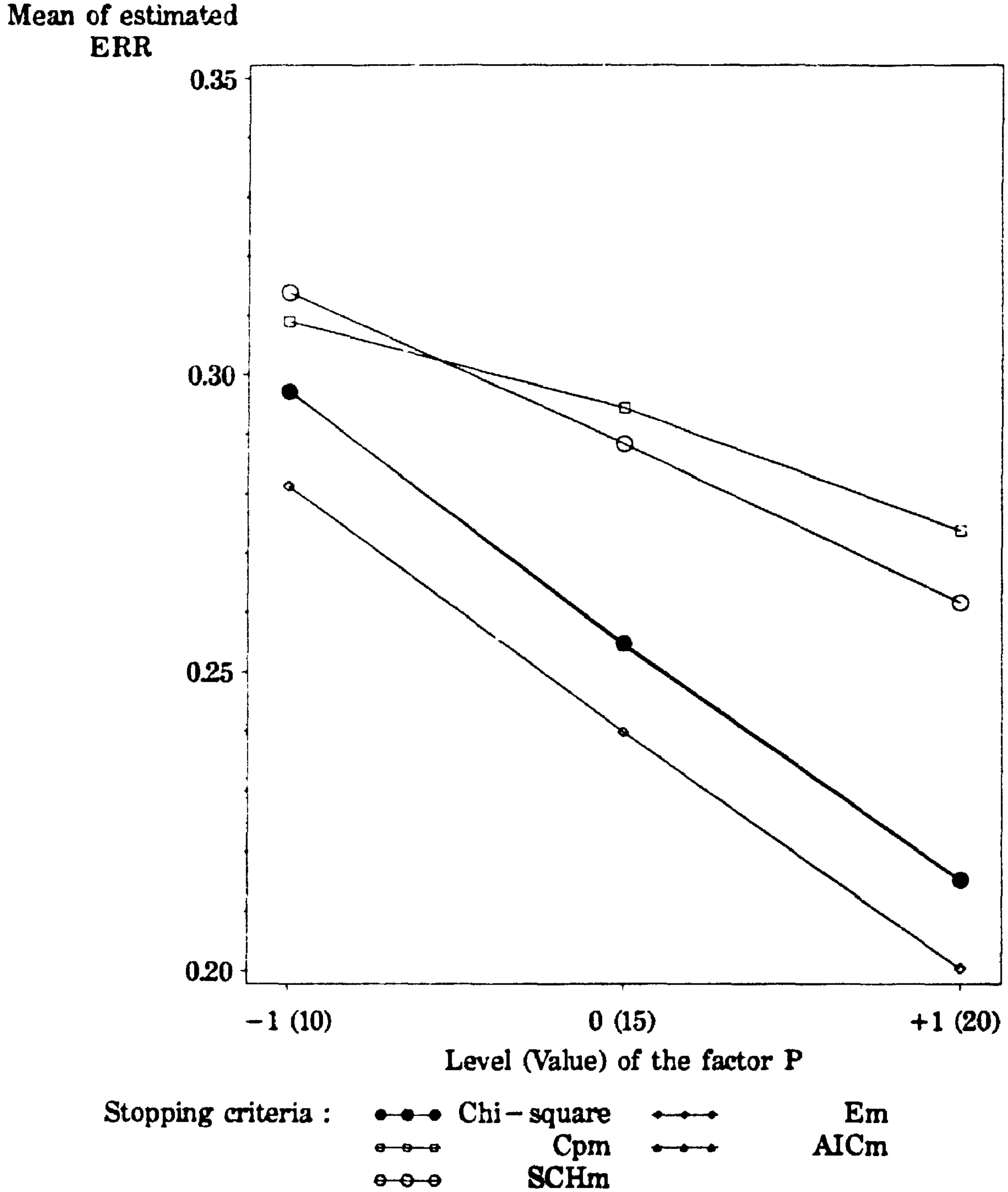


Figure 5.5.2.3

Mean of estimated ERR for the five stopping criteria over the levels of the factor B

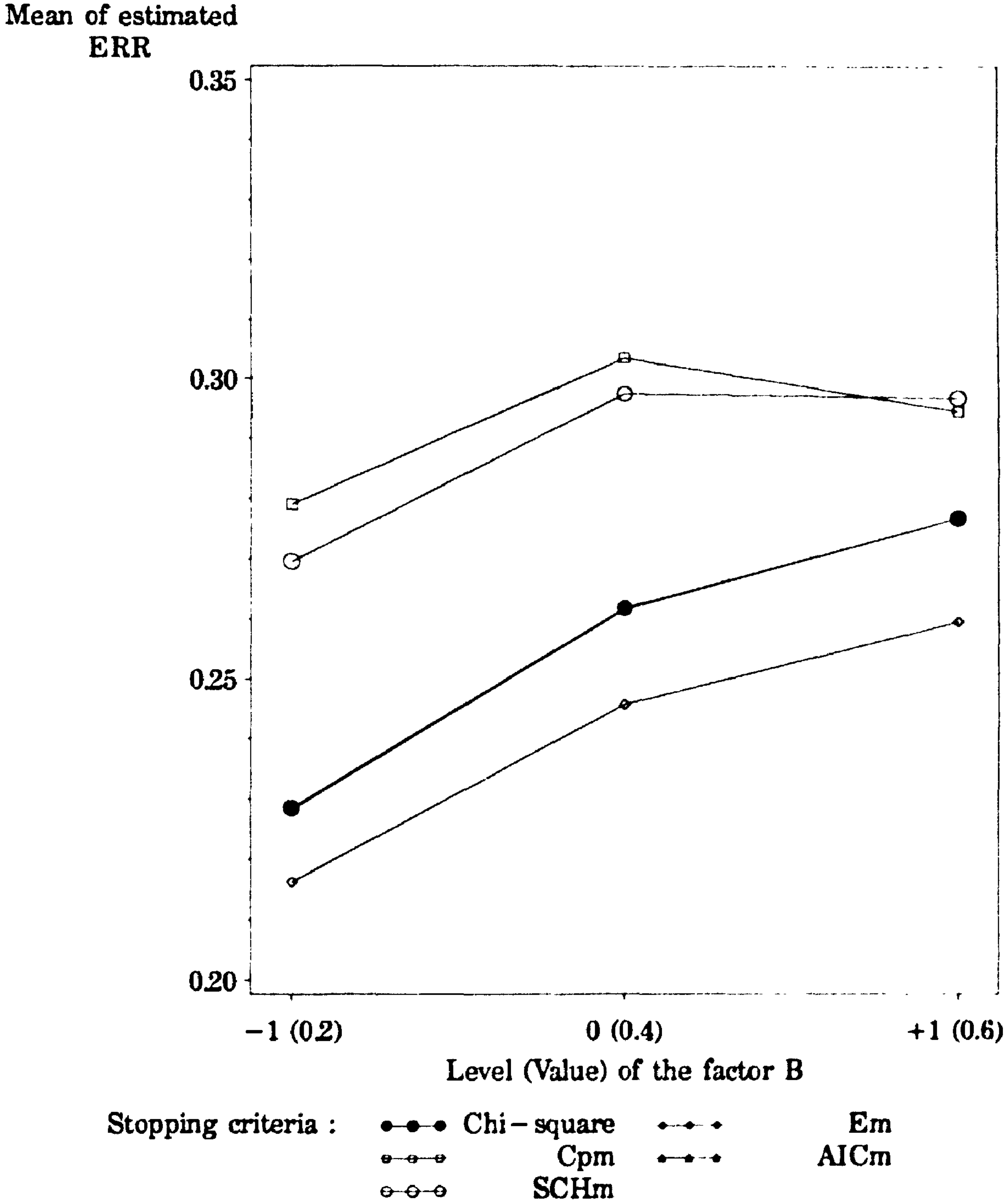


Figure 5.5.2.4

Mean of estimated ERR for the five stopping criteria over the levels of the factor M

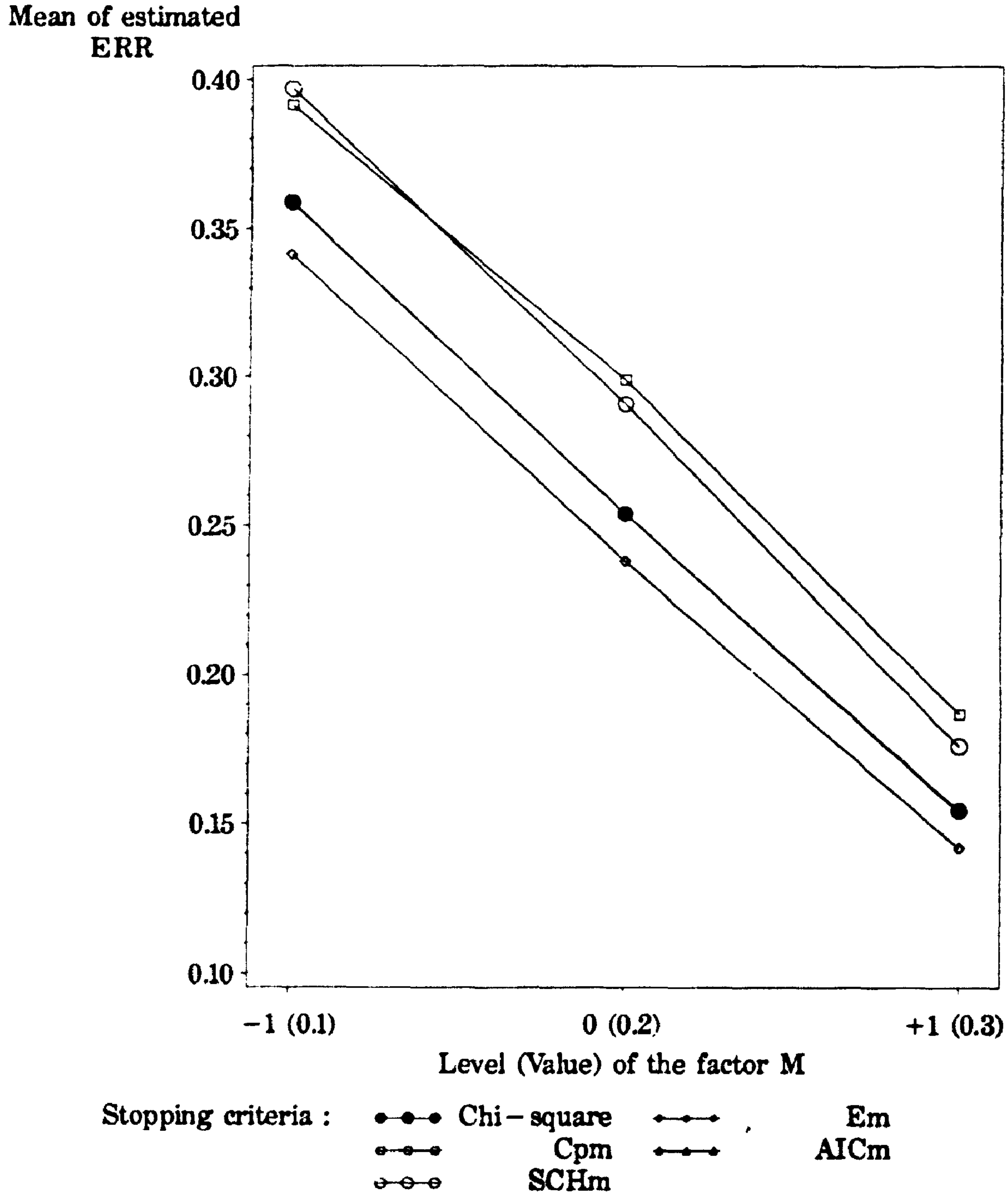


Figure 5.5.2.5

Mean of estimated ERR for the five stopping criteria over the levels of the factor N

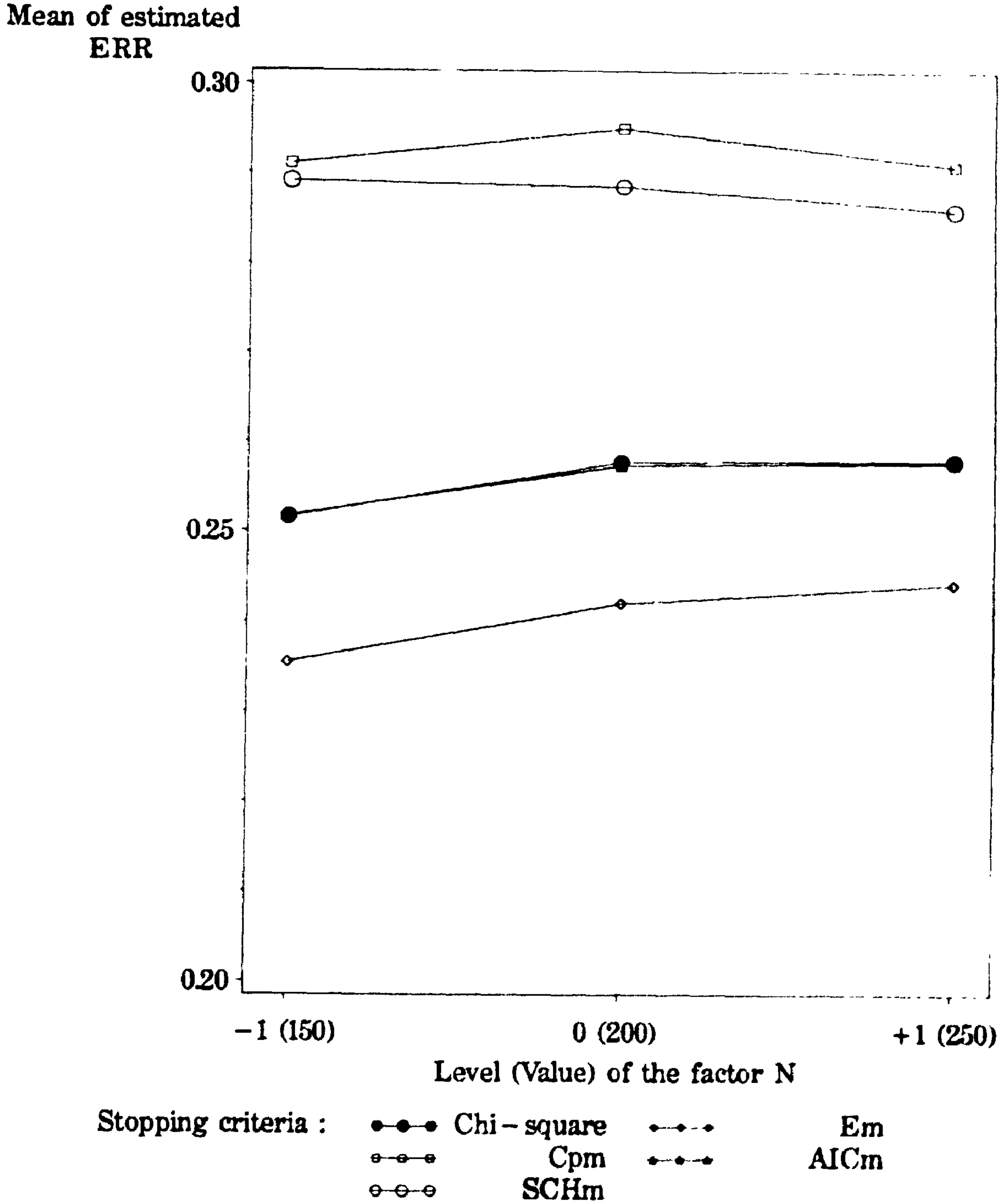


Figure 5.5.2.6

Effect of PM interaction on estimated ERR for the Chi-square stopping criterion

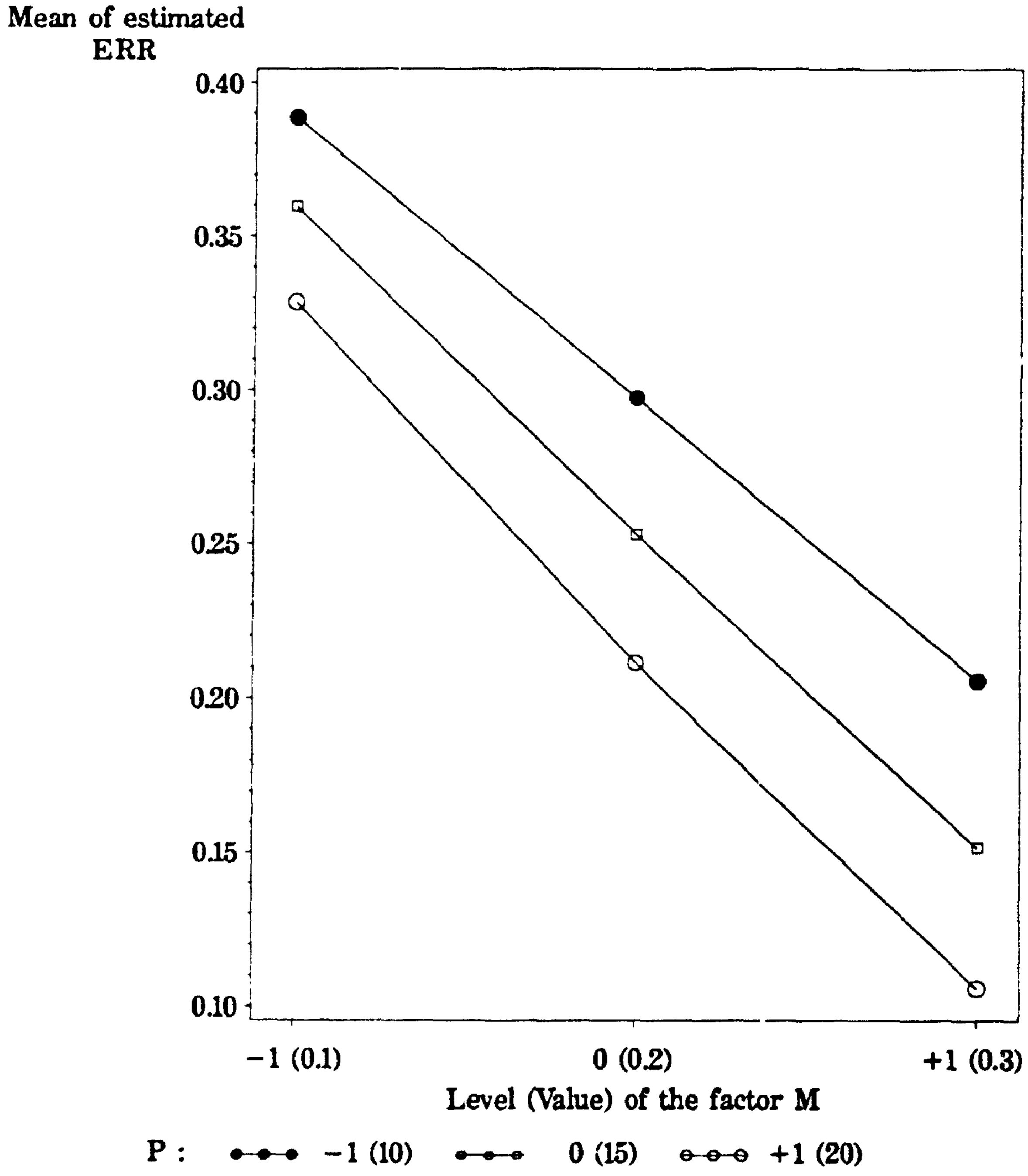


Figure 5.5.2.7

Effect of BM interaction on estimated ERR for the Chi−square stopping criterion

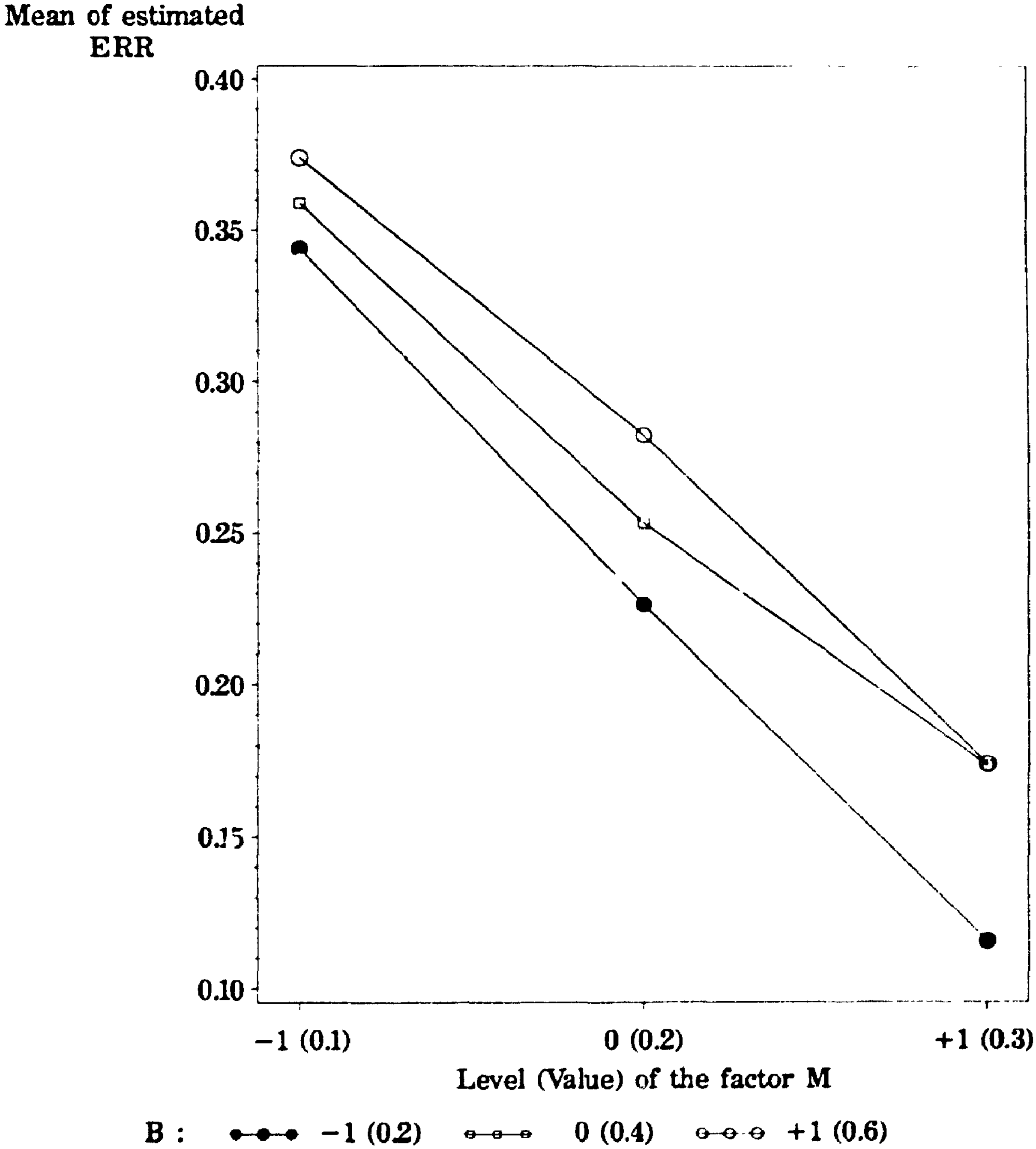


Figure 5.5.2.8
Effect of PM interaction on estimated ERR for the Em stopping criterion

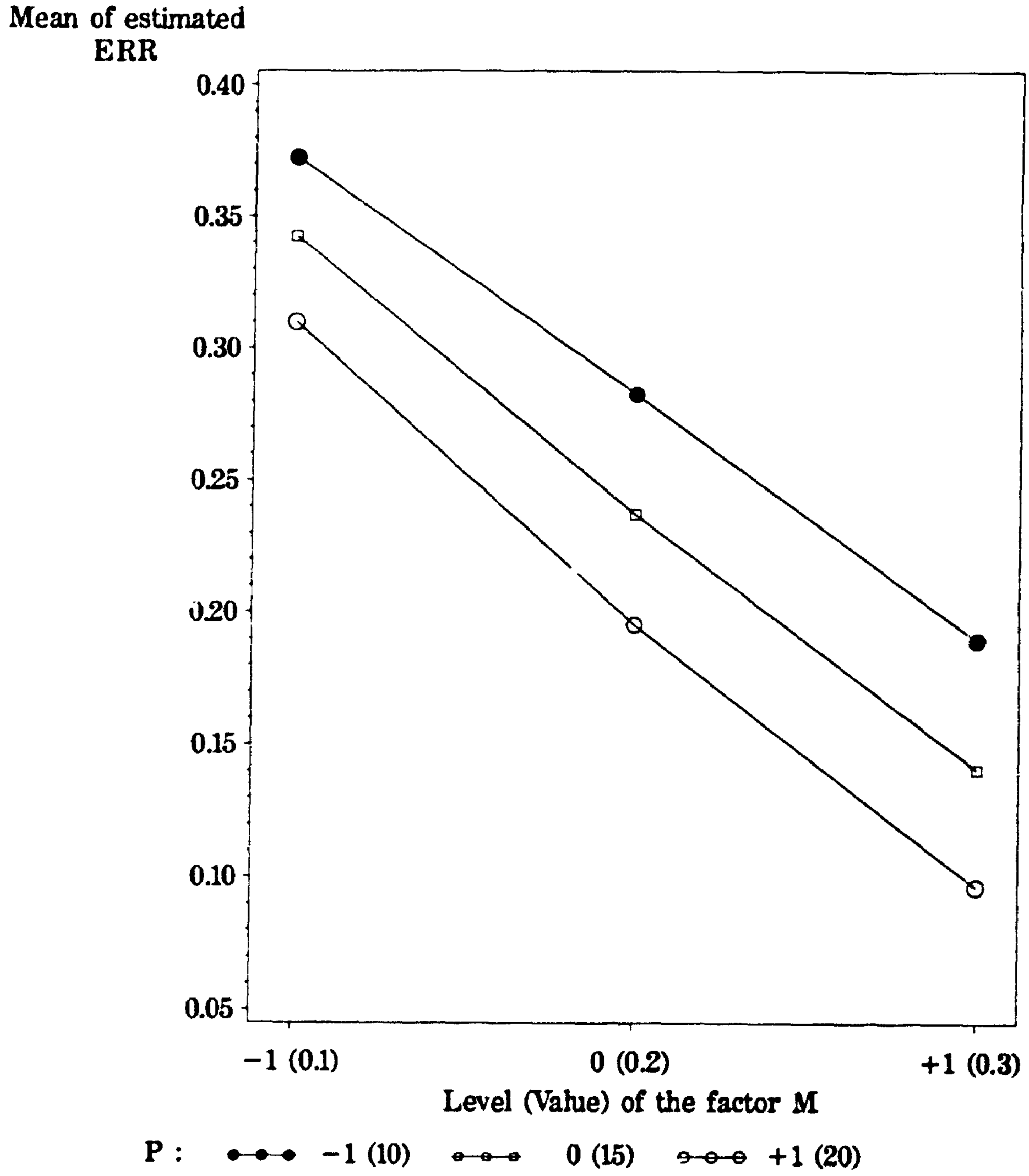


Figure 5.5.2.9

Effect of BM interaction on estimated ERR for the Em stopping criterion

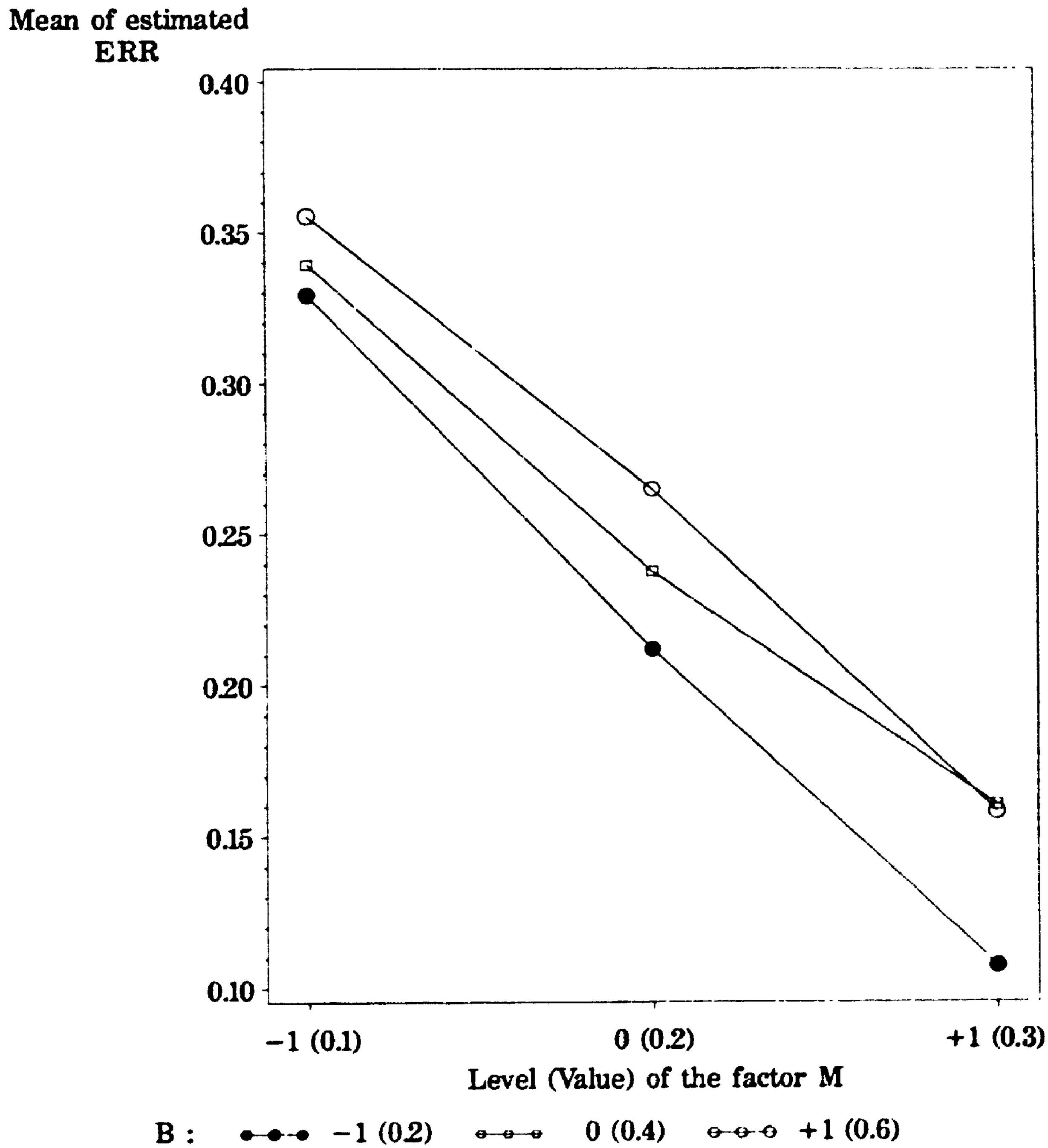


Figure 5.5.2.10

Effect of PM interaction on estimated ERR for the Cpm stopping criterion

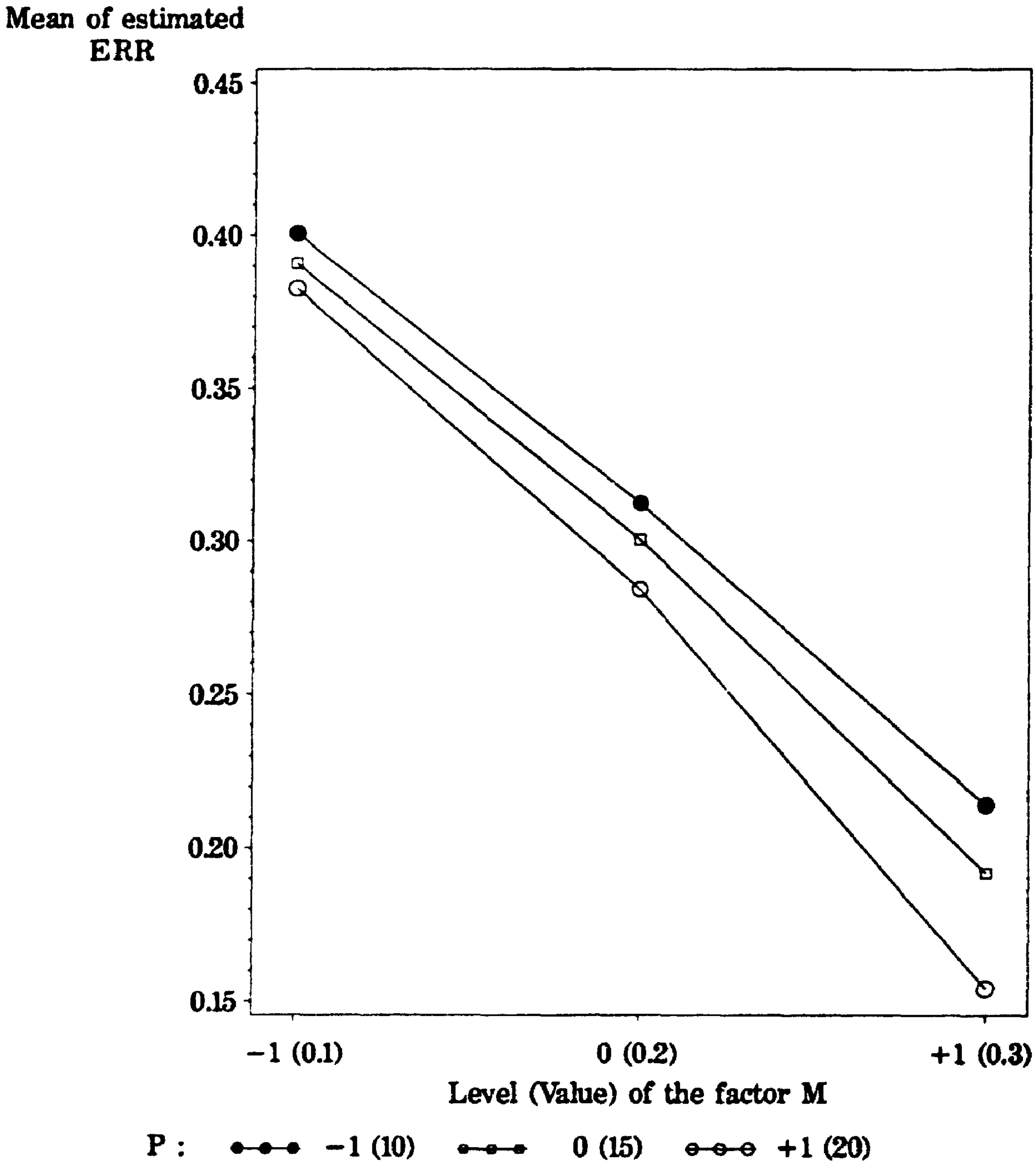


Figure 5.5.2.11

Effect of BM interaction on estimated ERR for the Cpm stopping criterion

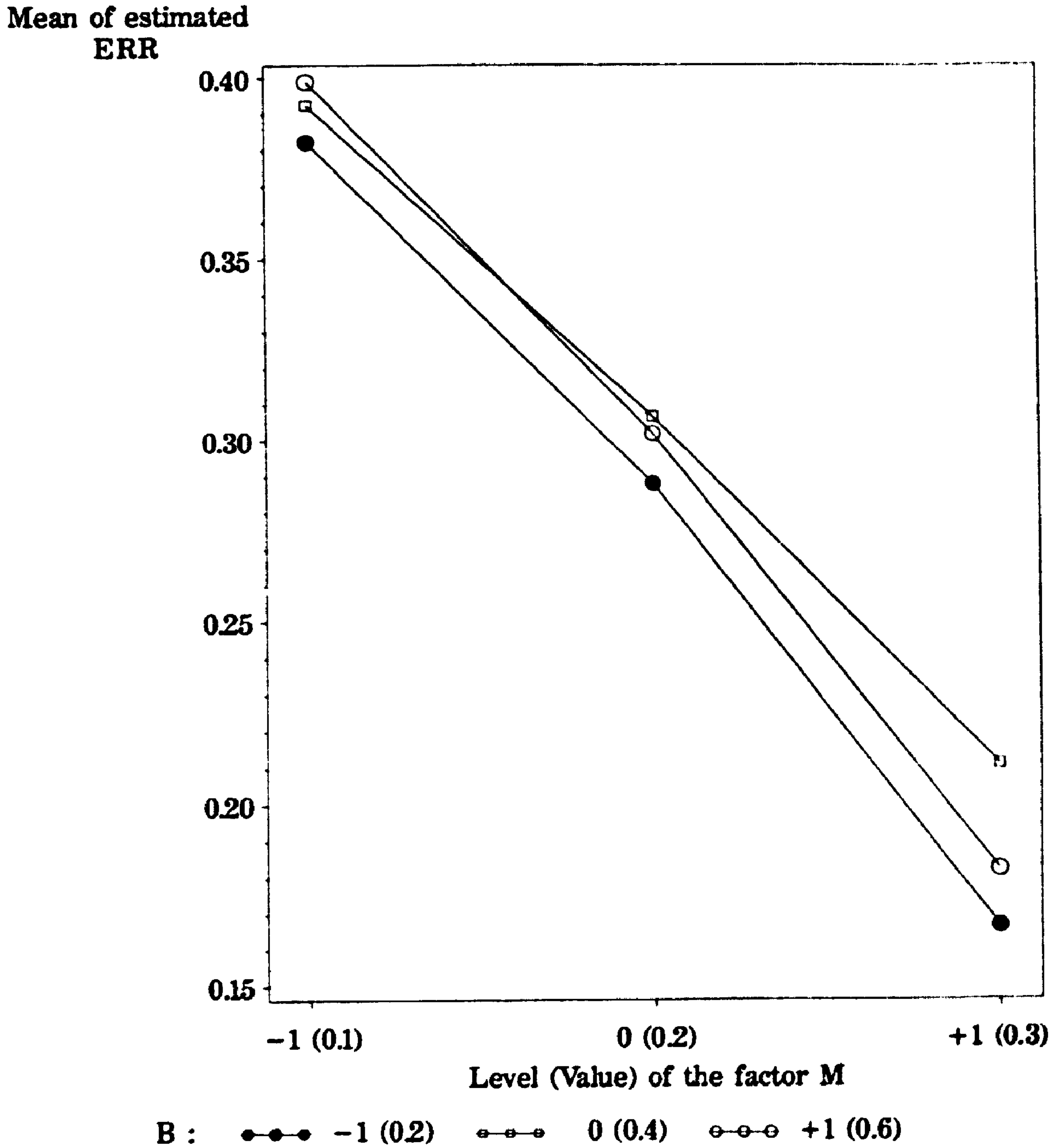


Figure 5.5.2.12

Effect of PM interaction on estimated ERR for the AICm stopping criterion

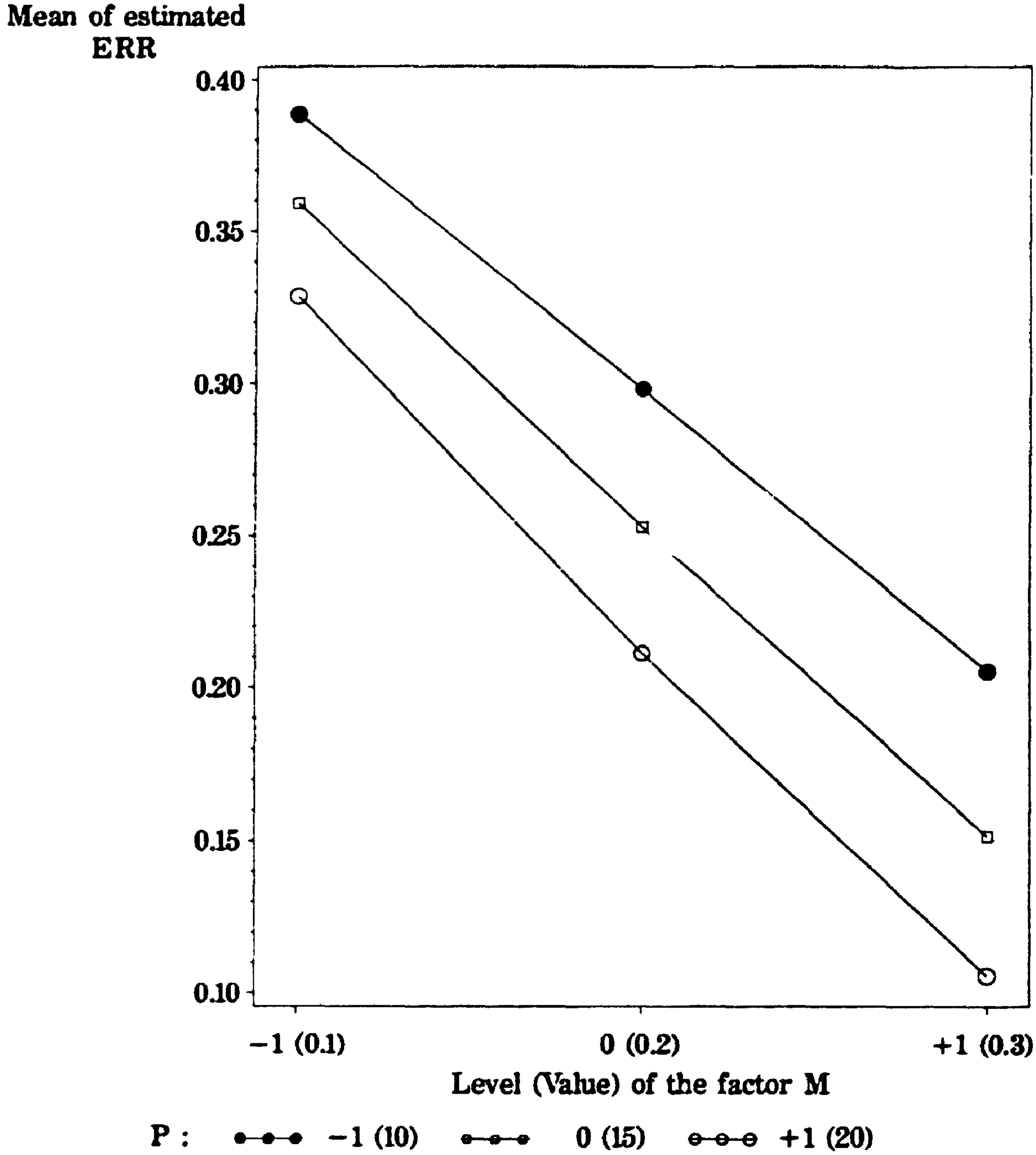


Figure 5.5.2.13

Effect of BM interaction on estimated ERR for the AICm stopping criterion

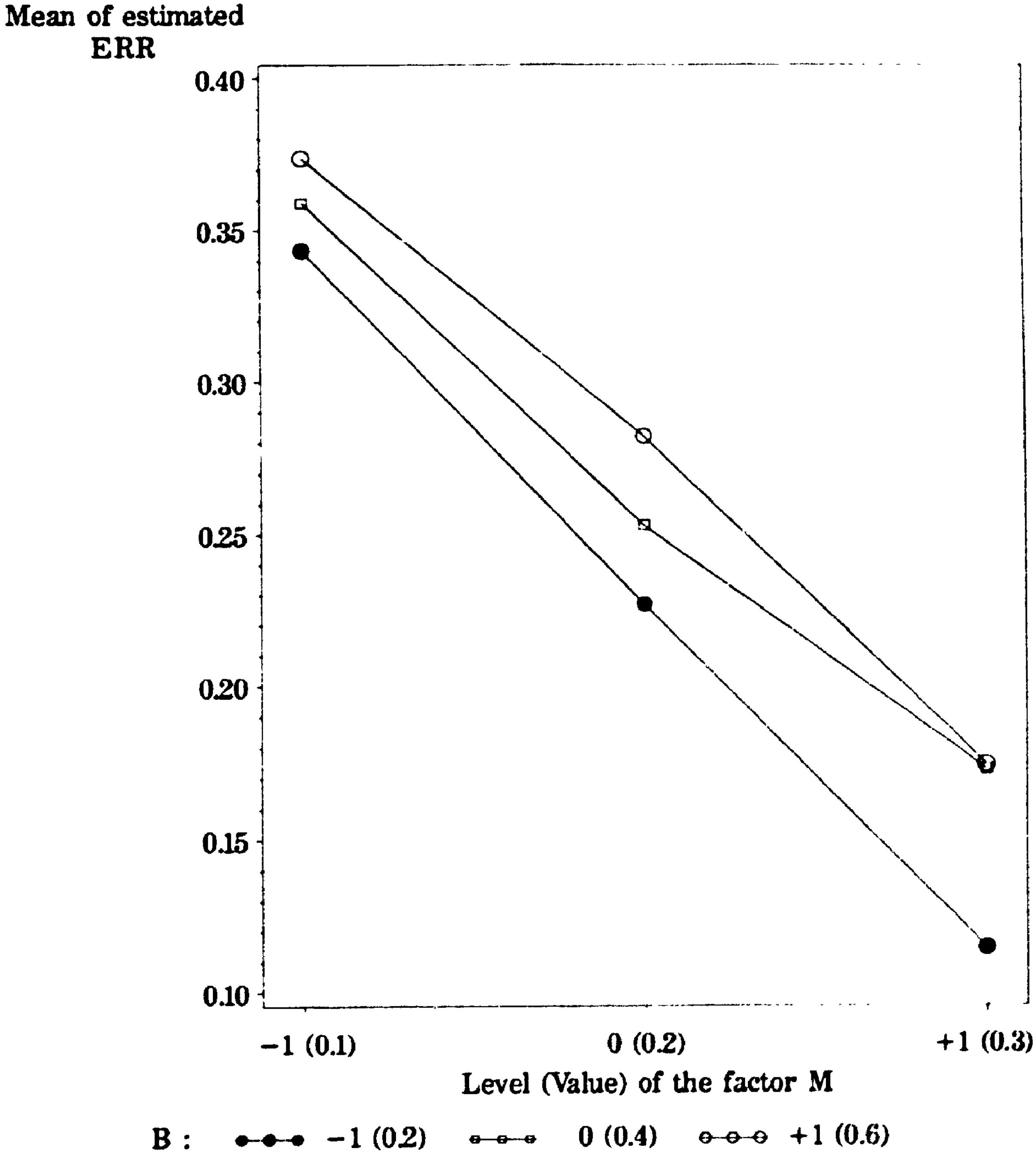


Figure 5.5.2.14

Effect of PM interaction on estimated ERR for the SCHm stopping criterion

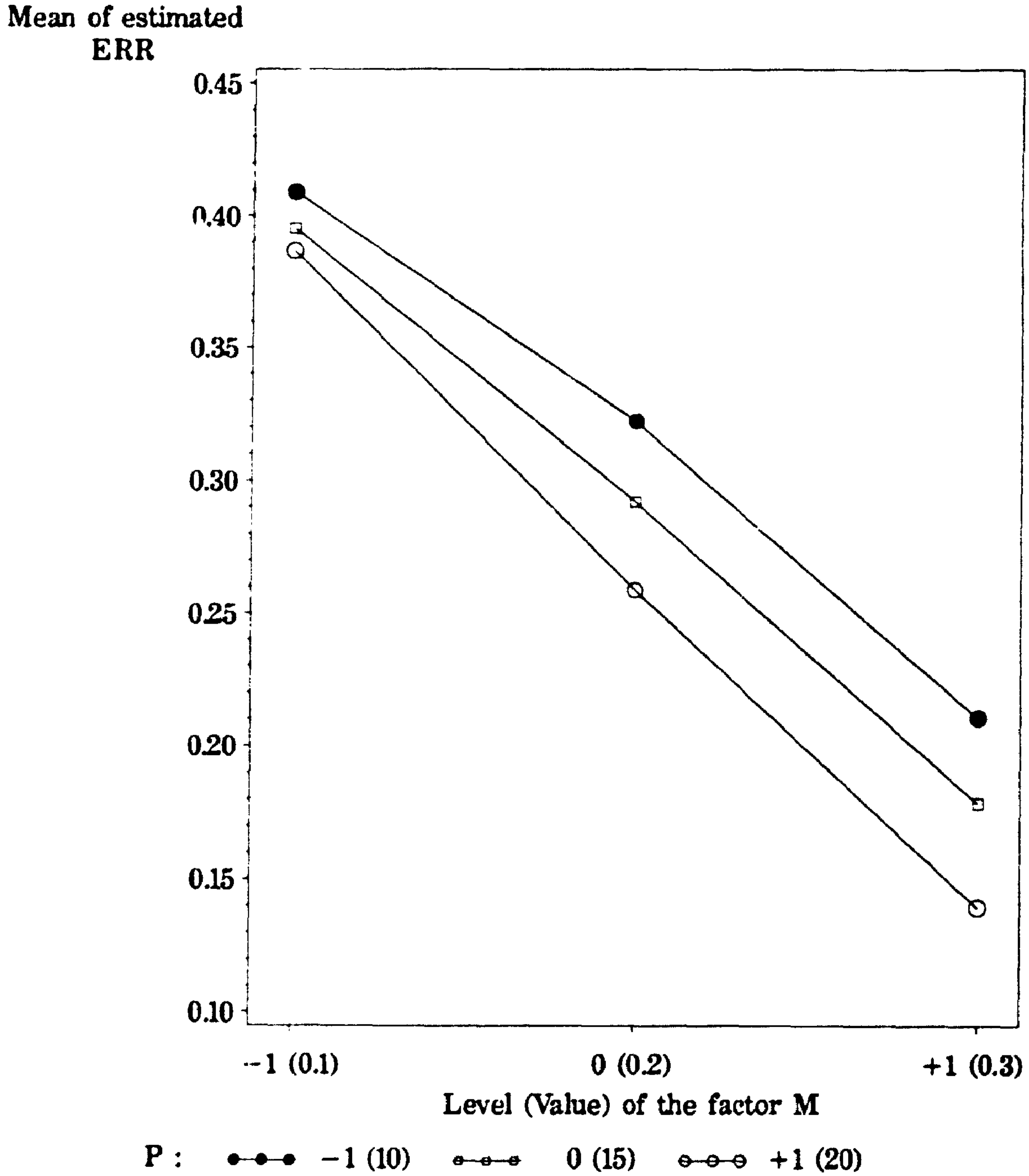
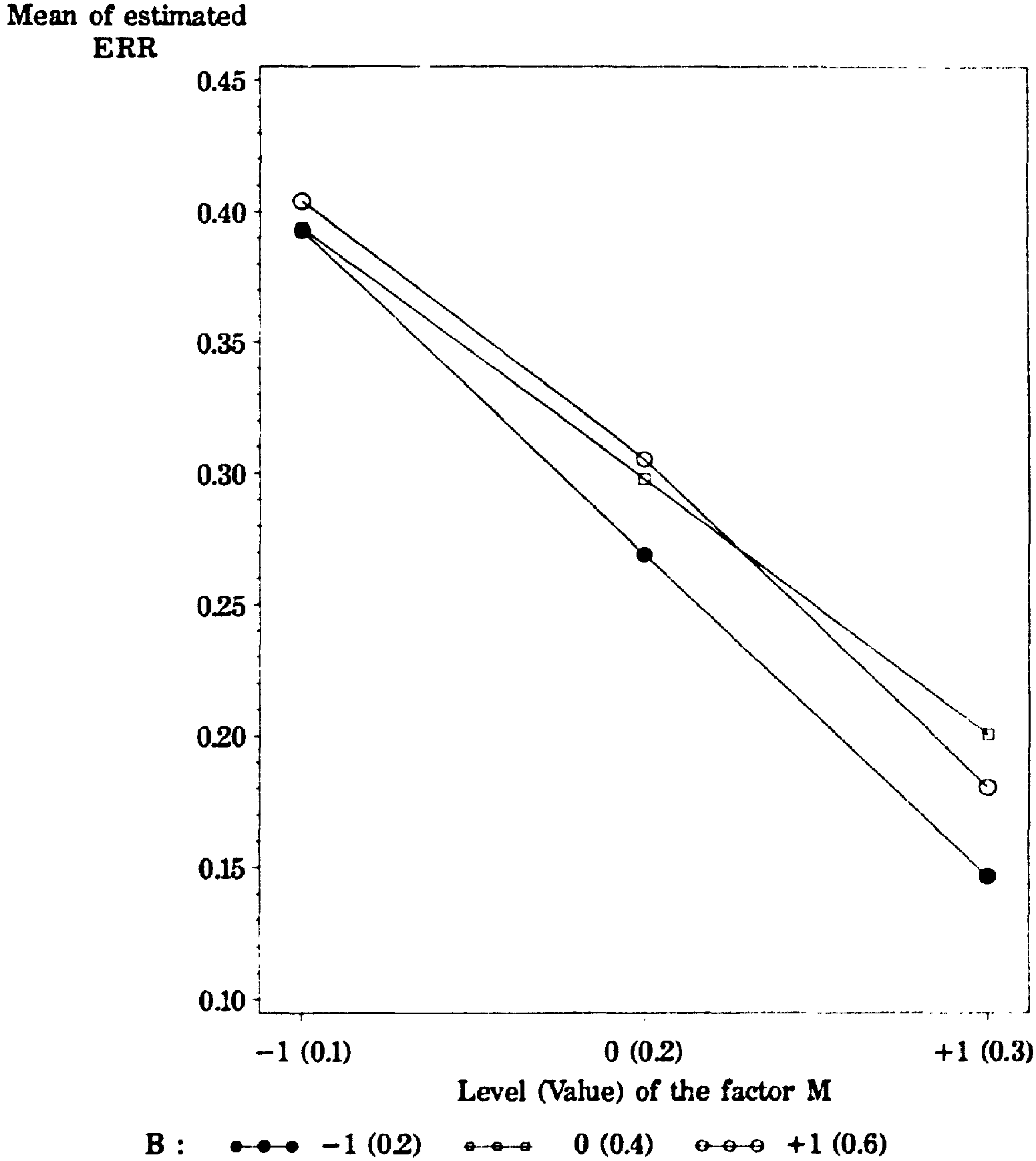


Figure 5.5.2.15

Effect of BM interaction on estimated ERR for the SCHm stopping criterion



CHAPTER 6

EXAMPLES, DISCUSSION, AND AREAS OF FURTHER STUDY

In this chapter three data sets are used to examine how well the results found in the simulation data hold in the practical data. These three data sets will be solely used for illustrative purposes; the topics described in the previous chapter will be investigated with these three data sets.

6.1 Example one

The first data set, the Low Birth Weight Data, consists of N = 100 ($n_0 = n_1 = 50$) and P = 8 ($P_c = P_b = 4$), which is the part of the data in appendix 1 in Hosmer (1989). Note that P_c and P_b denote the number of continuous and binary variables, respectively. The original data were collected to study risk factors associated with low infant birth weight at Baystate Medical Center in Springfield, Massachusetts, during 1986. More details of the data may be found in Hosmer (1989). The descriptions of the variables in the first data set are presented below.

The variables in the Low Birth Weight Data

Description	Variable
1. Subject number	SUB
2. Low birth weight (0 = No, 1 = Yes)	LOW
3. Age of the mother in years	AGE
4. Weight in pounds at the last menstrual period	LWT
5. Race (1 = White, 0 = Non-white)	RACE
6. Smoking status during pregnancy (0 = No, 1 = Yes)	SMOKE
7. History of premature labor (0 = None, 1 = One, etc.)	PTL
8. History of hypertension (0 = No, 1 = Yes)	HT

9. Presence of uterine irritability (0 = No, 1 = Yes) UI

10. Number of physician visits during the first trimester FTV

LOW is the response variable. AGE, LWT, PTL, and FTV are continuous variables, and RACE, SMOKE, HT, and UI are binary variables.

Table 6.1.1 gives ARR, Bias, and $\hat{ERR}$ at each step of forward stepwise logistic regression with the Low Birth Weight Data. The corresponding graph of ARR, Bias, and $\hat{ERR}$ is presented in figure 6.1.1. The effect of P, where P is the number of variables in the model, on ARR, Bias, and $\hat{ERR}$ can be seen in these results. Figure 6.1.1 shows that as step number increases, there is a decrease in ARR and an increase in Bias. Since $\hat{ERR}$ is a function of ARR and Bias, $\hat{ERR}$ may have a minimum anywhere from 1 to P, depending on the 'trade-offs' between these two opposite rates. It is observed that in this particular example $\hat{ERR}$ has a minimum at step 6. This means that the minimum of $\hat{ERR}$ is obtained with a subset of 6 variables rather than with the full model of 8 variables. This result supports the proposition explained in chapter 2; smaller $\hat{ERR}$ may be obtained with a subset model than with the full model.

Table 6.1.2 gives ARR, Bias, and $\hat{ERR}$ for different sample sizes in the Low Birth Weight Data and the corresponding graph is presented in figure 6.1.2. These results show the effect of N, where N is the sample size, on ARR, Bias, and $\hat{ERR}$. There is a decrease in Bias as N increases which agrees with the findings in the simulation data. However, there are no clear directional trends in both ARR and $\hat{ERR}$ as N increases.

The order of the variables selected by the seven selection criteria in forward stepwise logistic regression with the Low Birth Weight Data is given in table 6.1.3. It is found that only the WD selection criterion was different from the other six selection criteria in selecting the variables. The order of the variables selected by the LR and the other five selection criteria was UI, RACE, SMOKE, LWT, HT, FTV, PTL, and AGE, whereas RACE, SMOKE, UI, LWT, HT, FTV, PTL, and AGE was the order for the WD selection criterion. Only the first three variables were in different orders.

Table 6.1.4 shows the size of subset for the six stopping criteria with the LR and WD selection criteria in the Low Birth Weight Data. Note that only the two selection criteria, LR and WD, were included in the table because the results of the other five selection criteria are the same as those of the LR selection criterion. For both LR and WD selection criteria, the size of subset for the six stopping criteria may be written in descending order as $E_m > C_{pm} > \chi^2_{(0.15)} = \chi^2_{(0.20)} = AIC_m > SCH_m$. Given that the corresponding result in the multivariate normal case was $E_m > \{AIC_m > \chi^2_{(0.20)}\} > C_{pm} > SCH_m$, and that in the multivariate binary case $\{AIC_m > \chi^2_{(0.15)} > E_m\} > C_{pm} > SCH_m$, where { } indicates the same group, the following comparisons can be made: 1) C_{pm} was the fourth in size in the simulation data, whereas it is the second in size in the example data, 2) AIC_m is the same as $\chi^2_{(0.15)}$ and $\chi^2_{(0.20)}$ in size in the example data which agrees with the findings in the simulation data; 3) E_m is the first in size in the example data which agrees with the findings in the simulation data; and 4) SCH_m is the last in size in both simulation and example data.

Referring to table 6.1.4 again, for both LR and WD selection criteria, $\hat{ERR}$ for the six stopping criteria may be written in ascending order as $E_m < \chi^2_{(0.15)} = \chi^2_{(0.20)} = AIC_m < C_{pm} < SCH_m$. Given that the corresponding result in the multivariate normal case was $E_m < \{AIC_m < \chi^2_{(0.20)} < C_{pm}\}$, $\{C_{pm} < SCH_m\}$, and that in the multivariate binary case $E_m < \{AIC_m < \chi^2_{(0.15)}\} < \{SCH_m < C_{pm}\}$, there is no major discrepancy in the results between the example data and the simulation data.

Table 6.1.1
ARR, Bias, and E$\hat{R}$R at each step of forward stepwise logistic regression with the Low Birth Weight Data

[a]Step number	Variable entered	ARR	Bias	E$\hat{R}$R
1	UI	0.38000	0.00792	0.38792
2	RACE	0.32000	0.00653	0.32653
3	SMOKE	0.29000	0.01975	0.30975
4	LWT	0.25000	0.01570	0.26570
5	HT	0.25000	0.01762	0.26762
6	FTV	0.22000	0.02002	0.24002
7	PTL	0.22000	0.02306	0.24306
8	AGE	0.22000	0.02557	0.24557

[a]The LR selection criterion was used for the entry of the variables.

Table 6.1.2
ARR, Bias, and E$\hat{R}$R for different sample sizes of N in the Low Birth Weight Data

Subject number	N	ARR	Bias	E$\hat{R}$R
1-50, 51-100	100	0.22000	0.02557	0.24557
1-45, 51-95	90	0.20000	0.02757	0.22757
1-40, 51-90	80	0.21250	0.03086	0.24336
1-35, 51-85	70	0.27143	0.03436	0.30579
1-30, 51-80	60	0.22333	0.03262	0.26595
1-25, 51-75	50	0.20000	0.03726	0.23726

Table 6.1.3
[a]Order of variables selected by the seven selection criteria in forward stepwise logistic regression with the Low Birth Weight Data

Variable	Selection criteria						
	LR	LS	WD	SC	PH	LK	SW
AGE	8	8	8	8	8	8	8
LWT	4	4	4	4	4	4	4
RACE	2	2	1	2	2	2	2
SMOKE	3	3	2	3	3	3	3
PTL	7	7	7	7	7	7	7
HT	5	5	5	5	5	5	5
UI	1	1	3	1	1	1	1
FTV	6	6	6	6	6	6	6

[a]The number indicates the order of entry for the variable.

Table 6.1.4
Size of subset and $\hat{ERR}$ for the six stopping criteria
with the LR and WD selection criteria in the Low Birth Weight Data

Selection criteria	Stopping criteria	Size	$\hat{ERR}$
LR	$\chi^2_{(0.15)}$	4	0.26498
	$\chi^2_{(0.20)}$	4	0.26498
	E_m	6	0.24036
	C_{pm}	5	0.26733
	AIC_m	4	0.26498
	SCH_m	3	0.30185
WD	$\chi^2_{(0.15)}$	4	0.26498
	$\chi^2_{(0.20)}$	4	0.26498
	E_m	6	0.24036
	C_{pm}	5	0.26733
	AIC_m	4	0.26498
	SCH_m	2	0.29940

Figure 6.1.1

ARR, Bias, and estimated ERR at each step of forward stepwise logistic regression with the Low Birth Weight Data

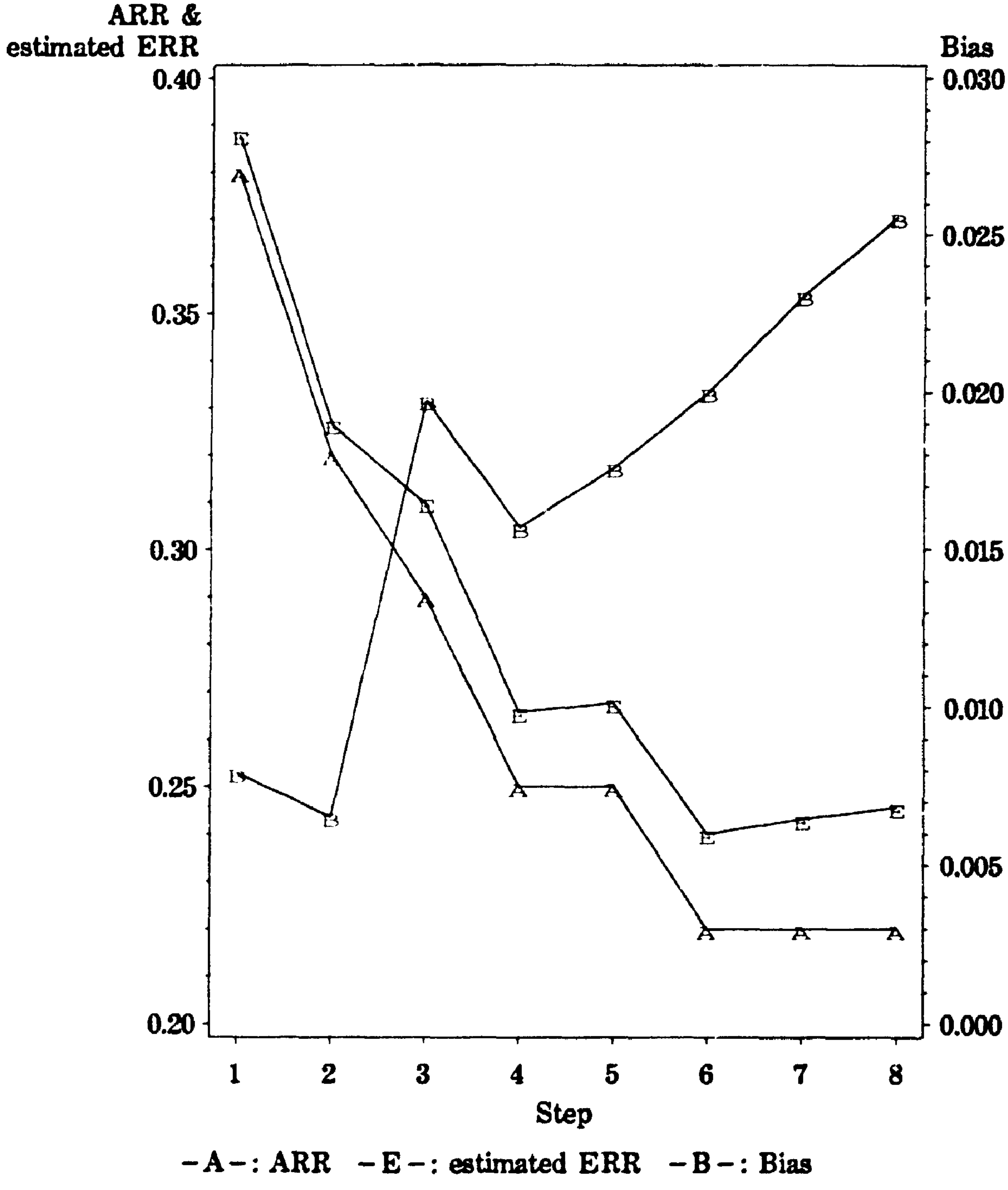
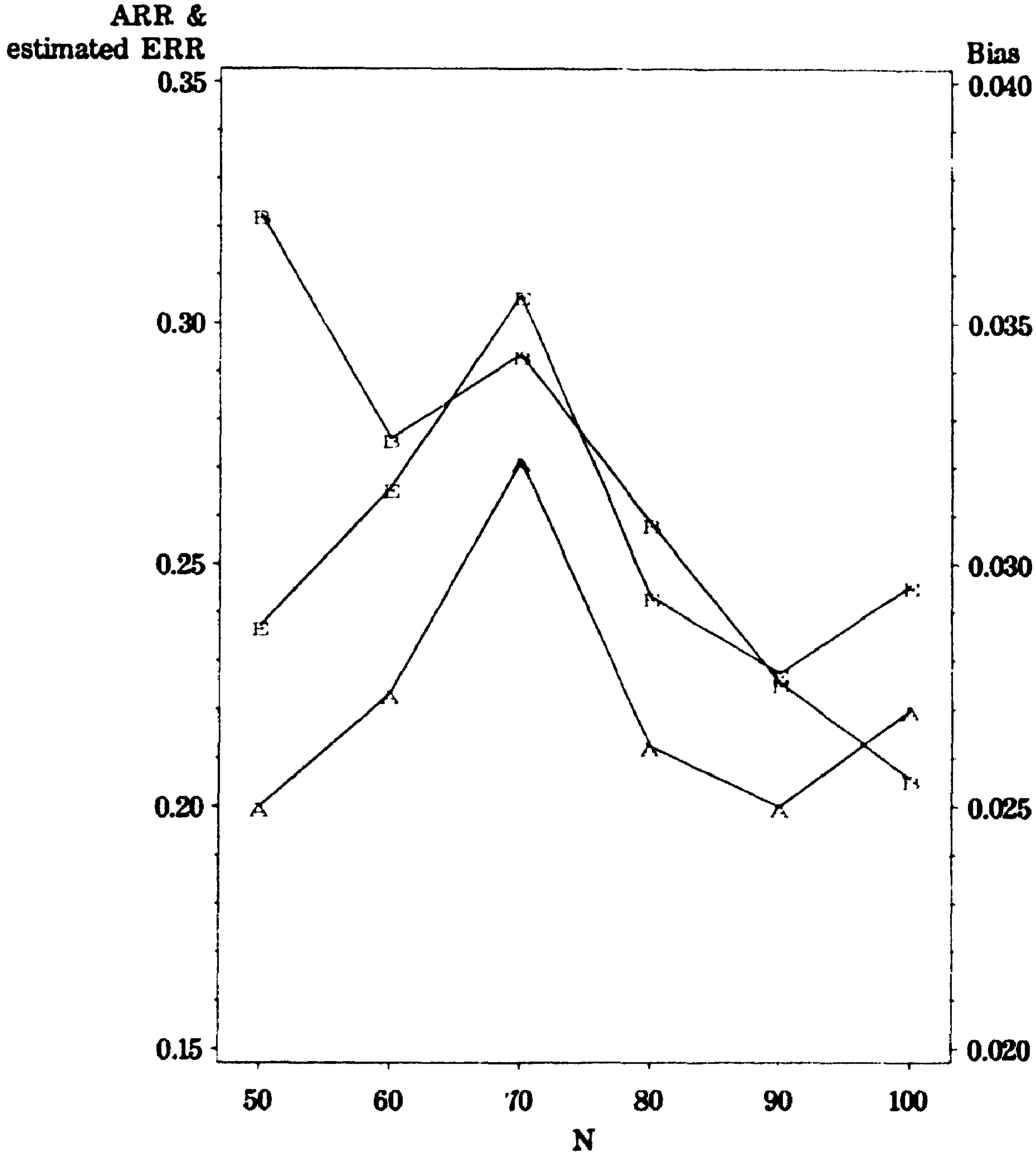


Figure 6.1.2

ARR, Bias, and estimated ERR for different sample sizes of N in the Low Birth Weight Data



–A–: ARR –E–: estimated ERR –B–: Bias

6.2 Example two

The second data set, the Intensive Care Unit Data, consists of N = 80 ($n_0 = n_1 = 40$) and P = 10 ($P_c = 3$ and $P_b = 7$), which is the part of the data in appendix 2 in Hosmer (1989). The original data were collected to develop a logistic model to predict the probability of survival to hospital discharge of the ICU patients at Baystate Medical Center in Springfield, Massachusetts, during 1983. A number of publications have been published which have focused on various facets of the problem (Teres et al., 1982; Lemeshow et al., 1985; Lemeshow et al., 1987; Teres et al., 1987; Lemeshow et al., 1988). The details of the data may be found there or in other quoted references. The variables considered in the second data set are given below.

The variables in the Intensive Care Unit Data

Description	Code	Variable
1. Subject number	Number	SUB
2. Vital status	0 = Lived, 1 = Died	STA
3. Age	Years	AGE
4. Sex	0 = Male, 1 = Female	SEX
5. Service at admission	0 = Medical, 1 = Surgical	SER
6. Cancer part of present problem	0 = No, 1 = Yes	CAN
7. History of chronic renal failure	0 = No, 1 = Yes	CRN
8. Systolic blood pressure at admission	mm Hg	SYS
9. Heart rate at admission	Beats per minute	HRA
10. Type of admission	0 = Elective, 1 = Emergency	TYP
11. PH from initial blood gases	0 = 7.25, 1 = < 7.25	PH
12. Level of Consciousness at admission	0 = No coma or stupor 1 = Deep stupor or coma	LOC

STA is the response variable. AGE, SYS, and HRA are continuous variables, and SEX, SER, CAN, CRN, TYP, PH, and LOC are binary variables.

Table 6.2.1 gives ARR, Bias, and $\hat{ERR}$ at each step of forward stepwise logistic regression with the Intensive Care Unit Data. The corresponding graph of ARR, Bias, and $\hat{ERR}$ is presented in figure 6.2.1. Figure 6.1.1 shows that as step number increases ARR does not always decrease. ARR is U-shaped which implies that inclusion of uninformative variables into the model reduce the predictive ability of the model. There is an increase in Bias as step number increases which agrees with both results of the simulation data and of the first data set. It is observed that $\hat{ERR}$ has a minimum with a subset of 6 variables rather than with the full model of 10 variables. This result again supports the previous proposition.

ARR, Bias, and $\hat{ERR}$ for different sample sizes in the Intensive Care Unit Data are presented in table 6.2.2. The corresponding graph of ARR, Bias, and $\hat{ERR}$ is displayed in figure 6.2.2. The shape of figure 6.2.2 is very similar that of figure 5.3.1.5 in the multivariate normal case. There is a decrease in Bias as N increases. However, there are no clear directional trends in both ARR and $\hat{ERR}$ as N increases.

Table 6.2.3 show the order of the variables selected by the seven selection criteria in forward stepwise logistic regression with the Intensive Care Unit Data. It is found that one group of {LR, LS} selection criteria selected the variables in the one order, and that the other group of {WD, SC, PH, LK, SW} selection criteria selected the variables in another order. The order of variables selected by the first group of selection criteria was TYP, LOC, AGE, CAN, CRN, HRA, SER, SEX, PH, and SYS, whereas TYP, LOC, AGE, CAN, HRA, CRN, SER, SEX, PH, and SYS was the order of variables for the second group of selection criteria. It is noticed that only the fifth and sixth variables are in different orders in two groups of selection criteria.

Table 6.2.4 shows the size of subset for the six stopping criteria with the LR and SC selection criteria in the Intensive Care Unit Data. Note that only the two selection criteria, LR and SC, as the representative in each group of selection criteria were included in the table. For both LR and SC selection criteria, the size of subset for the six stopping criteria may be written in descending order as $E_m > C_{pm} > \chi^2_{(0.15)} = \chi^2_{(0.20)} = AIC_m > SCH_m$. The same remarks can be made as those in the first example because of the same results of two examples.

Referring to table 6.2.4 again, for both LR and SC selection criteria, $\hat{ERR}$ for the six stopping criteria may be written in ascending order as $E_m < C_{pm} < \chi^2_{(0.15)} = \chi^2_{(0.20)} = AIC_m < SCH_m$. Given that the corresponding result in the multivariate normal case was $E_m < \{AIC_m < \chi^2_{(0.20)} < C_{pm}\}$, $\{C_{pm} < SCH_m\}$, and that in the multivariate binary case $E_m < \{AIC_m < \chi^2_{(0.15)}\} < \{SCH_m < C_{pm}\}$, there is one major discrepancy between the result of the example data and the results of the simulation data; C_{pm} gave smaller $\hat{ERR}$ than $\chi^2_{(0.15)}$, $\chi^2_{(0.20)}$, and AIC_m in the example data.

Table 6.2.1
ARR, Bias, and $\hat{ERR}$ at each step of forward stepwise logistic regression with the Intensive Care Unit Data

[a]Step number	Variable entered	ARR	Bias	$\hat{ERR}$
1	TYP	0.28750	0.00177	0.28927
2	LOC	0.27500	0.01413	0.28913
3	AGE	0.26250	0.01092	0.27342
4	CAN	0.26250	0.01464	0.27714
5	CRN	0.23750	0.01701	0.25451
6	HRA	0.17500	0.02080	0.19580
7	SER	0.20000	0.02292	0.22292
8	SEX	0.22500	0.02762	0.25262
9	PH	0.21250	0.02927	0.24177
10	SYS	0.21250	0.03185	0.24435

[a]The LR selection criterion was used for the entry of the variables.

Table 6.2.2
ARR, Bias, and $\hat{ERR}$ for different sample sizes of N in the Intensive Care Unit Data

Subject number	N	ARR	Bias	$\hat{ERR}$
1-40, 41-80	80	0.21250	0.03185	0.24435
1-35, 41-75	70	0.22857	0.03683	0.26540
1-30, 41-70	60	0.23333	0.04091	0.27424
1-25, 41-65	50	0.22000	0.04930	0.26930
1-20, 41-60	40	0.17500	0.05774	0.23274

Table 6.2.3
[a]Order of variables selected by the seven selection criteria
in forward stepwise logistic regression with the Intensive Care Unit Data

Variable	Selection criteria						
	LR	LS	WD	SC	PH	LK	SW
AGE	3	3	3	3	3	3	3
SEX	8	8	8	8	8	8	8
SER	7	7	7	7	7	7	7
CAN	4	4	4	4	4	4	4
CRN	5	5	6	6	6	6	6
SYS	10	10	10	10	10	10	10
HRA	6	6	5	5	5	5	5
TYP	1	1	1	1	1	1	1
PH	9	9	9	9	9	9	9
LOC	2	2	2	2	2	2	2

[a]The number indicates the order of entry for the variable.

Table 6.2.4
Size of subset and $\hat{ERR}$ for the six stopping criteria
with the LR and SC selection criteria in the Intensive Care Unit Data

Selection criteria	Stopping criteria	Size	$\hat{ERR}$
LR	$\chi^2_{(0.15)}$	4	0.27650
	$\chi^2_{(0.20)}$	4	0.27650
	E_m	6	0.19506
	C_{pm}	5	0.25381
	AIC_m	4	0.27650
	SCH_m	2	0.28343
SC	$\chi^2_{(0.15)}$	4	0.27650
	$\chi^2_{(0.20)}$	4	0.27650
	E_m	6	0.19506
	C_{pm}	5	0.25558
	AIC_m	4	0.27650
	SCH_m	2	0.28343

Figure 6.2.1

ARR, Bias, and estimated ERR at each step of forward stepwise logistic regression with the Intensive Care Unit Data

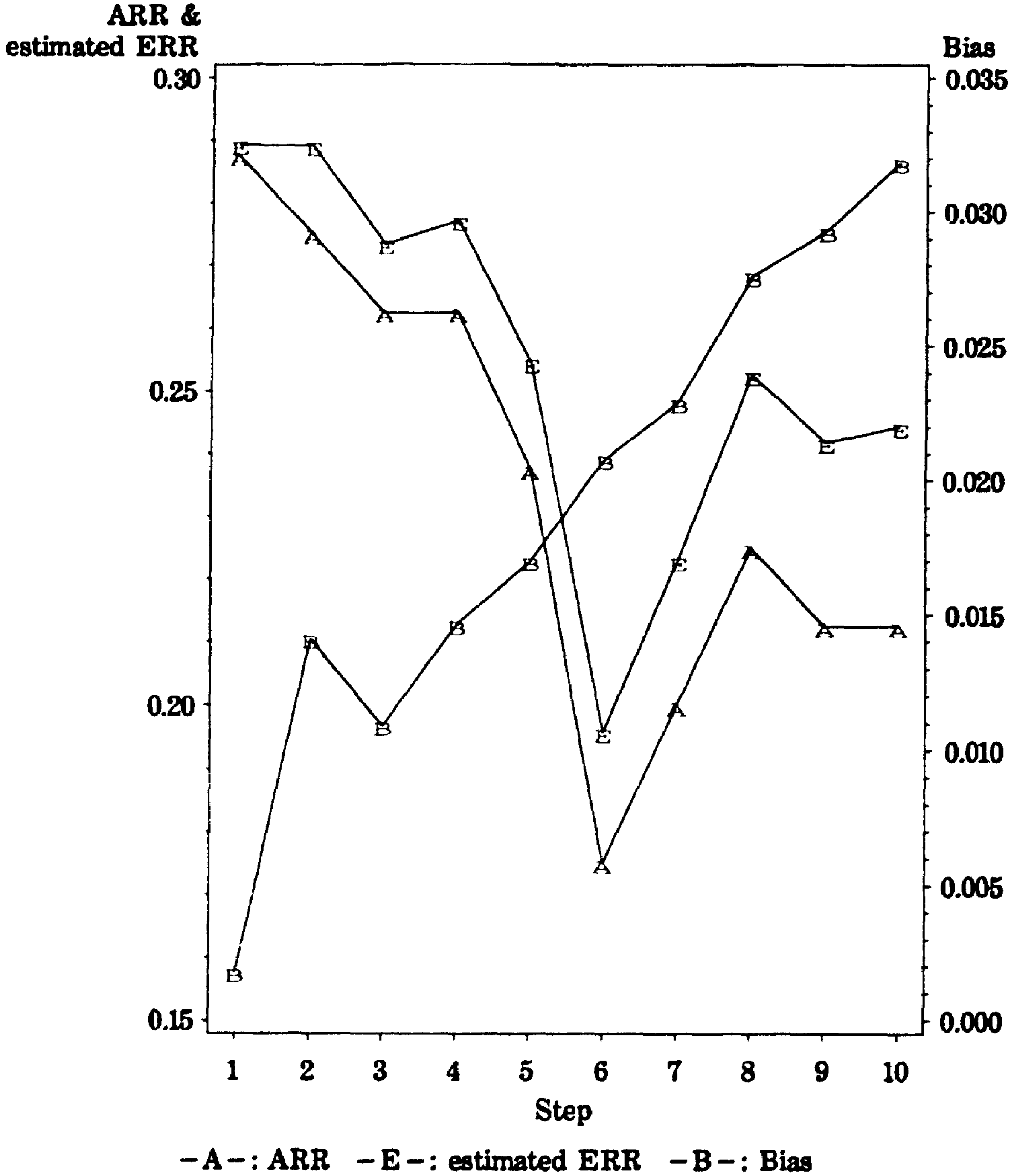
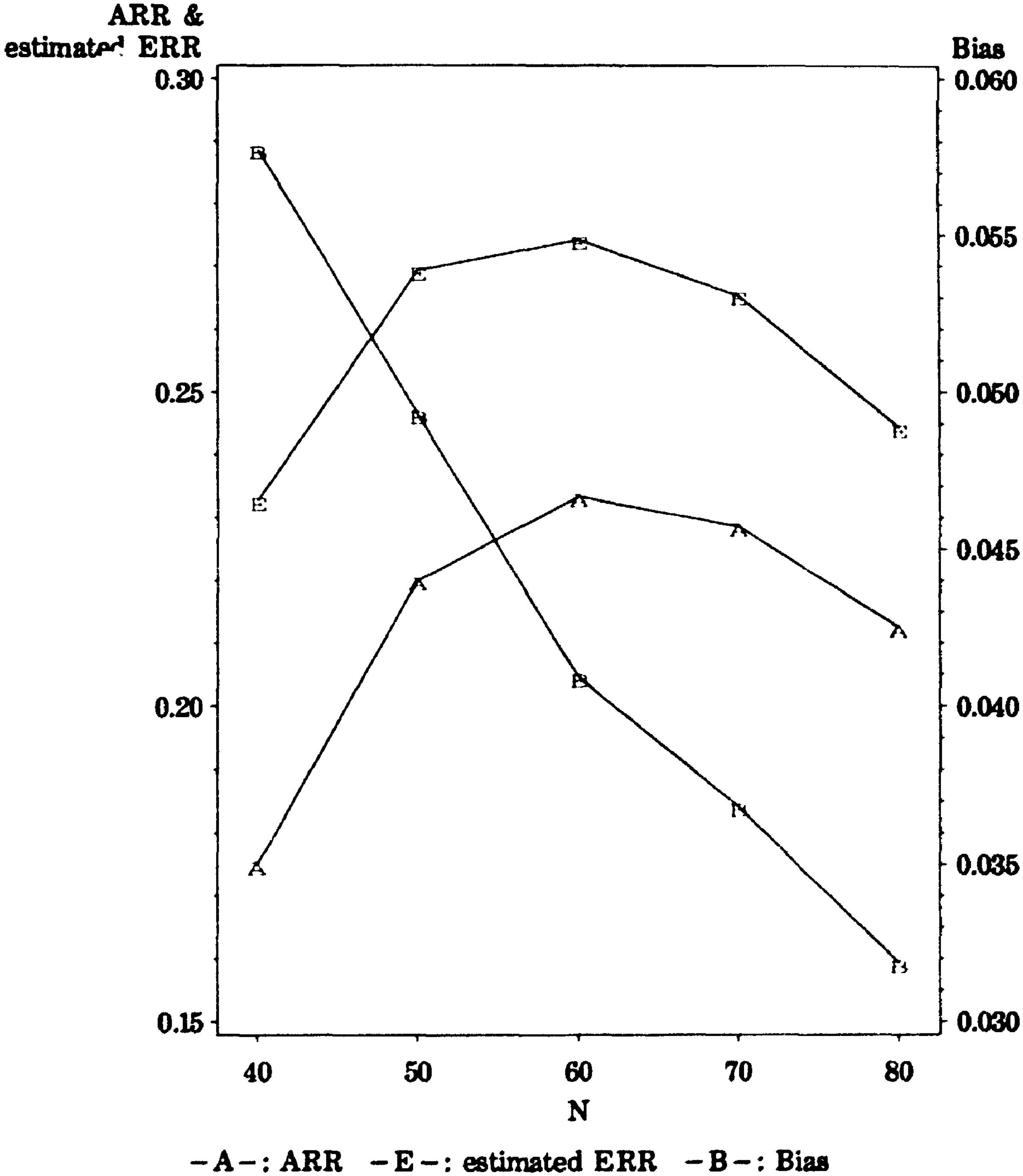


Figure 6.2.2

ARR, Bias, and estimated ERR for different sample sizes of N in the Intensive Care Unit Data



6.3 Example three

The third data set, the Walking Study Data, comes from a study conducted from mid-1987 to mid-1988 in London, Ontario. The purpose of this study was to investigate the relationship between physical activity and health among order persons. The details of data collection have been reported in the literature (Koval et al., 1992). The third data set consists of N = 200 ($n_0 = n_1 = 100$) and P = 12 ($P_c = 10$ and $P_b = 2$). The variables included in the third data set are given below.

The variables in the Walking Study Data

Description	Variable
1. Subject number	SUB
2. Maximum oxygen intake (0 = low, 1 = high)	O2
3. Age in years	AGE
4. Demispan in cm	DEMISPAN
5. Sitting knee height in cm	SITKNEE
6. Height in cm	HEIGHT
7. Weight in kg	WEIGHT
8. Maximum heart rate at end of test	MAXHR
9. Body mass index	BMI
10. Systolic blood pressure	SYSBP
11. Diastolic blood pressure	DIASBP
12. Resting heart rate	RESTHR
13. Sex (0 = Female, 1 = Male)	SEX
14. Activity compared past (0 = low, 1 = high)	ACTIVITY

O2 is the response variable. AGE, DEMISPAN, SITKNEE, HEIGHT, WEIGHT, MAXHR, BMI, SYSBP, DIASBP, and RESTHR are continuous variables, and SEX and ACTIVITY are binary variables.

Table 6.3.1 gives ARR, Bias, and $\hat{ERR}$ at each step of forward stepwise logistic regression with the Walking Study Data. The corresponding graph of ARR, Bias, and $\hat{ERR}$ is presented in figure 6.3.1. Figure 6.3.1 shows that as step number increases, there is a decrease in ARR and an increase in Bias. It is observed that $\hat{ERR}$ has a minimum at step 9. This means that the minimum of $\hat{ERR}$ is obtained with a subset of 9 variables rather than with the full model of 12 variables. The similar results were found in the two previous examples; a subset model gave a smaller $\hat{ERR}$ than the full model.

Table 6.3.2 gives ARR, Bias, and $\hat{ERR}$ for different sample sizes in the Walking Study Data and the corresponding graph is presented in figure 6.3.2. These results show the effect of N, where N is the sample size, on ARR, Bias, and $\hat{ERR}$. There is a decrease in Bias as N increases which agrees with the findings in both two previous example data and simulation data. However, there are no clear directional trends in both ARR and $\hat{ERR}$ as N increases. The same remarks can be made as those in the two previous example data because of the similar results.

The order of the variables selected by the seven selection criteria in forward stepwise logistic regression with the Walking Study Data is given in table 6.3.3. It is found that only the SW selection criterion was different from the other six selection criteria in selecting the variables. The order of the variables selected by the LR and the other five selection criteria was AGE, RESTHR, DIASBP, SYSBP, MAXHR, DEMISPAN, SITKNEE, HEIGHT, ACTIVITY, WEIGHT, BMI, and SEX, whereas AGE, RESTHR, BMI, DIASBP, SYSBP, MAXHR, SEX, SITKNEE, DEMISPAN, ACTIVITY, HEIGHT, and WEIGHT was the order for the SW selection criterion. It is noticed that only the first two variables were in the same order for the SW selection criteria, compared with the LR and the other five selection criteria.

Table 6.3.4 shows the size of subset for the six stopping criteria with the LR and SW selection criteria in the Walking Study Data. Note that only the two selection criteria, LR and SW, were included in the table because the results of the other five selection criteria are the same as those of the LR selection criterion. For both LR and SW selection criteria, the descending order of the size of subset for the six stopping criteria is $E_m > C_{pm} > \chi^2_{(0.15)} = \chi^2_{(0.20)} = AIC_m > SCH_m$. This result is the same as in the two previous examples.

Referring to table 6.3.4 again, for both LR and SW selection criteria the ascending order of $\hat{ERR}$ for the six stopping criteria is $E_m < SCH_m < \chi^2_{(0.15)} = \chi^2_{(0.20)} = AIC_m < C_{pm}$. Given that the corresponding result in the multivariate normal case was $E_m < \{AIC_m < \chi^2_{(0.20)} < C_{pm}\}$, $\{C_{pm} < SCH_m\}$, and that in the multivariate binary case $E_m < \{AIC_m < \chi^2_{(0.15)}\} < \{SCH_m < C_{pm}\}$, there is one major discrepancy between the result of the example data and the results of the simulation data; SCH_m gave a smaller $\hat{ERR}$ than $\chi^2_{(0.15)}$, $\chi^2_{(0.20)}$, AIC_m, and C_{pm} in the example data.

Table 6.3.1
ARR, Bias, and ER̂R at each step of forward stepwise logistic regression with the Walking Study Data

[a]Step number	Variable entered	ARR	Bias	ER̂R
1	AGE	0.40000	0.00448	0.40448
2	RESTHR	0.41000	0.00767	0.41767
3	DIASBP	0.41000	0.01060	0.42060
4	SYSBP	0.36500	0.01353	0.37853
5	MAXHR	0.40500	0.01639	0.42139
6	DEMISPAN	0.39500	0.01850	0.41350
7	SITKNEE	0.36500	0.01958	0.38458
8	HEIGHT	0.37000	0.02102	0.39102
9	ACTIVITY	0.34500	0.02318	0.36818
10	WEIGHT	0.35000	0.02530	0.37530
11	BMI	0.35500	0.02689	0.38189
12	SEX	0.36000	0.02866	0.38866

[a]The LR selection criterion was used for the entry of the variables.

Table 6.3.2
ARR, Bias, and ER̂R for different sample sizes of N in the Walking Study Data

Subject number	N	ARR	Bias	ER̂R
1-100, 101-200	200	0.36000	0.02866	0.38866
1-90, 101-190	180	0.38333	0.03002	0.41335
1-80, 101-180	160	0.36250	0.03130	0.39380
1-70, 101-170	140	0.37857	0.03473	0.41330
1-60, 101-160	120	0.37500	0.04057	0.41557
1-50, 101-150	100	0.36000	0.04572	0.40572

Table 6.3.3
[a]Order of variables selected by the seven selection criteria in forward stepwise logistic regression with the Walking Study Data

Variable	Selection criteria						
	LR	LS	WD	SC	PH	LK	SW
AGE	1	1	1	1	1	1	1
DEMISPAN	6	6	6	6	6	6	9
SITKNEE	7	7	7	7	7	7	8
HEIGHT	8	8	8	8	8	8	11
WEIGHT	10	10	10	10	10	10	12
MAXHR	5	5	5	5	5	5	6
BMI	11	11	11	11	11	11	3
SYSBP	4	4	4	4	4	4	5
DIASBP	3	3	3	3	3	3	4
RESTHR	2	2	2	2	2	2	2
SEX	12	12	12	12	12	12	7
ACT	9	9	9	9	9	9	10

[a]The number indicates the order of entry for the variable.

Table 6.3.4
Size of subset and $\hat{ERR}$ for the six stopping criteria
with the LR and SW selection criteria in the Walking Study Data

Selection criteria	Stopping criteria	Size	$\hat{ERR}$
LR	$\chi^2_{(0.15)}$	2	0.41724
	$\chi^2_{(0.20)}$	2	0.41724
	E_m	9	0.36794
	C_{pm}	3	0.41959
	AIC_m	2	0.41724
	SCH_m	1	0.40506
SW	$\chi^2_{(0.15)}$	2	0.41724
	$\chi^2_{(0.20)}$	2	0.41724
	E_m	7	0.37397
	C_{pm}	3	0.42975
	AIC_m	2	0.41724
	SCH_m	1	0.40506

Figure 6.3.1

ARR, Bias, and estimated ERR at each step of forward stepwise logistic regression with the Walking Study Data

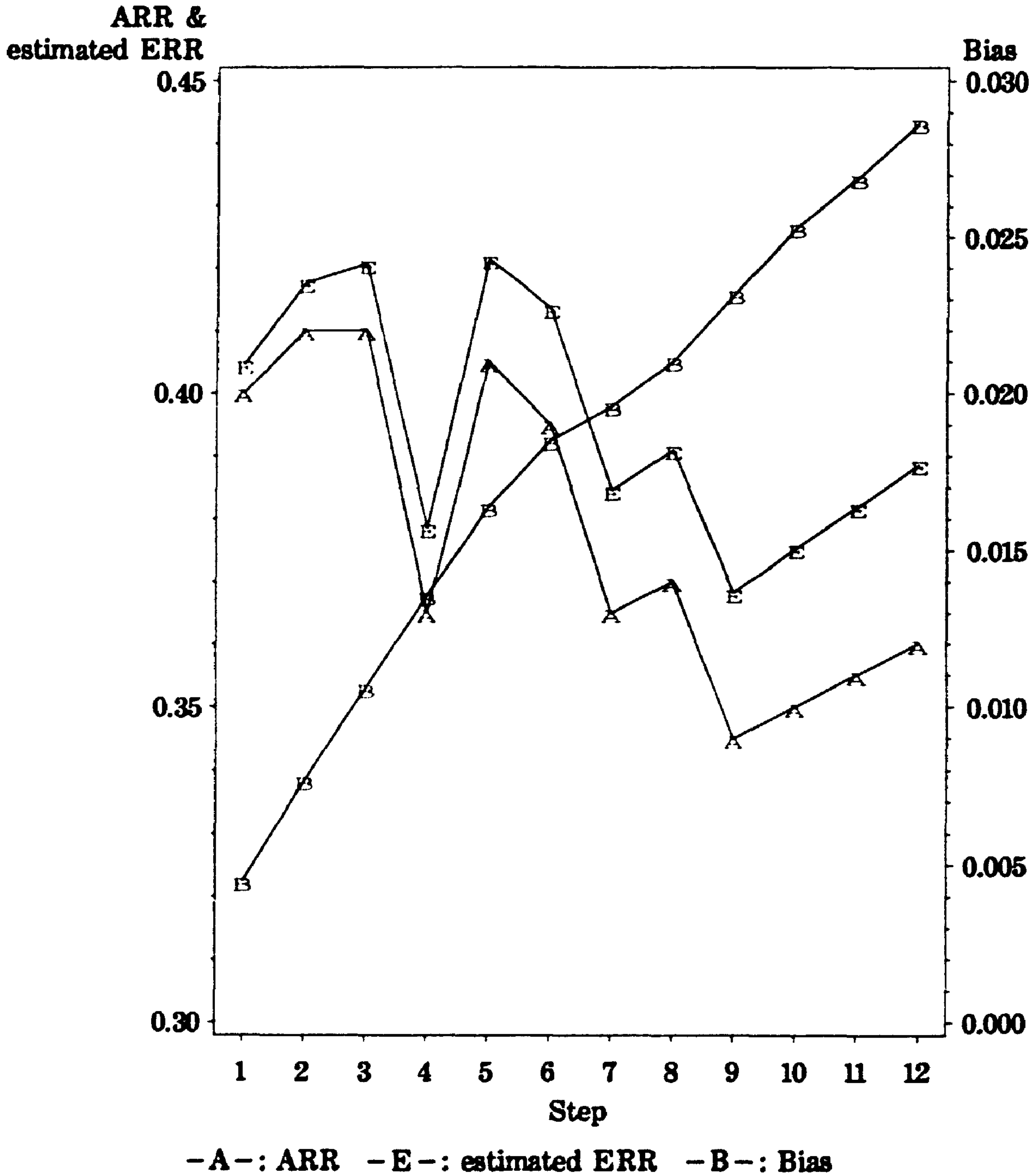
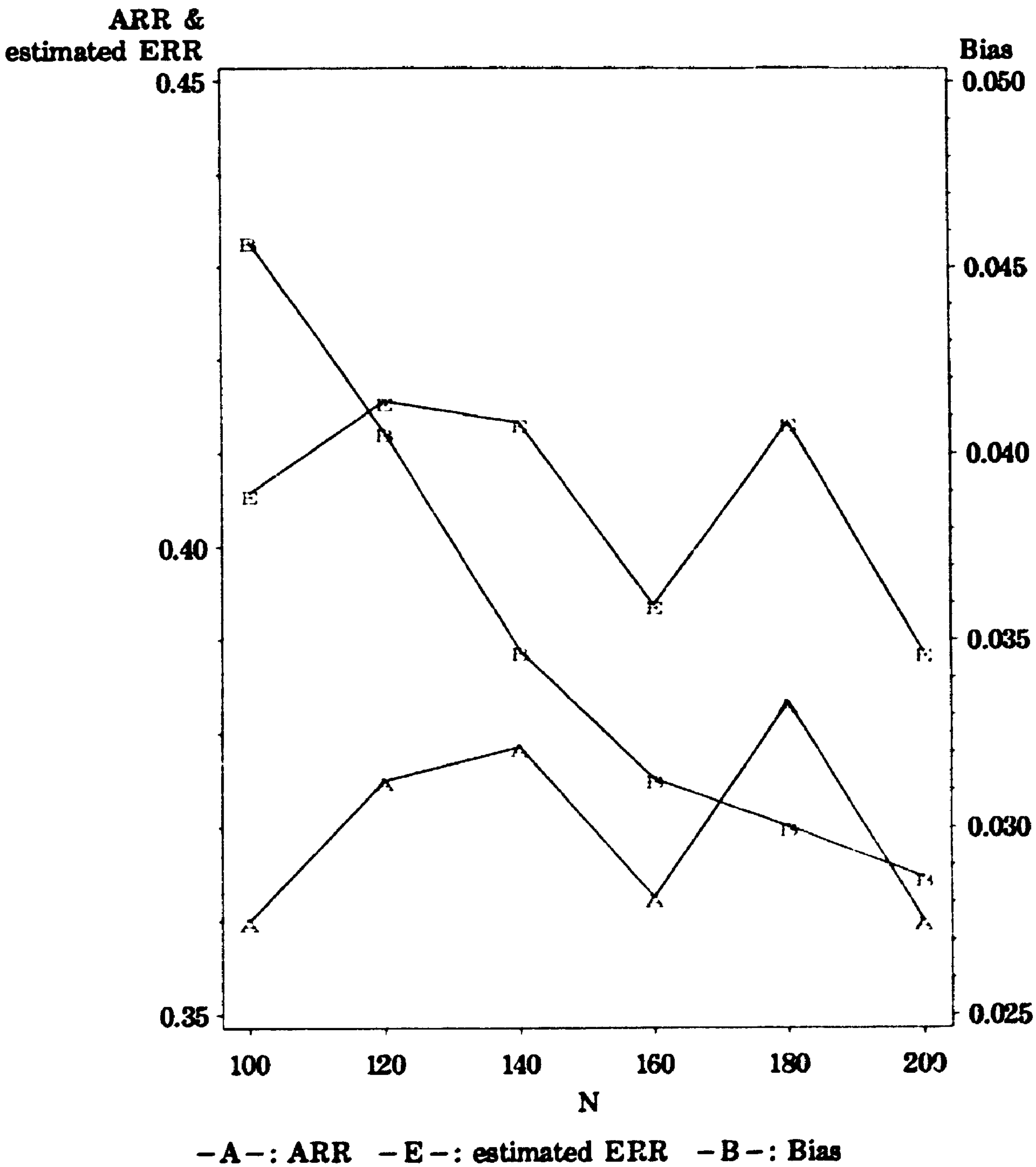


Figure 6.3.2

ARR, Bias, and estimated ERR for different sample sizes of N in the Walking Study Data



6.4 Discussion and areas of further study

The logistic model generally has two main purposes, namely description and prediction. In the case of description, a general strategy for selecting the predictor variables to retain in a logistic model is formulated with the goal of obtaining a valid estimate of the effect of a specified risk variable on a particular disease variable, while controlling for other confounding variables.

Another important purpose of the logistic model is prediction. In the case of prediction, the valid estimate of an individual regression coefficient in the logistic model is not of the primary interest. The selection of predictor variables which yield a good prediction is of the primary interest.

Over the last several years the logistic model has been used with increasing frequency for the predictive purpose in medical and epidemiologic studies (e.g., Pozen et al. 1984; Lemeshow et al. 1985, 1987, and 1988; Tierney et al. 1985; Teres et al. 1987; Baumgartner et al. 1992; Ferraris et al. 1992; Algren et al. 1993; and Horbar et al. 1993). With the recent efforts in developing predictive models in medical and epidemiologic research, there has been little known about what factors are important for predictive models.

In this thesis we have investigated five topics which are related to the prediction in the logistic model. Monte Carlo simulations were used to investigate them in both multivariate normal variables and multivariate binary variables.

In section 5.1 we studied the determination of the best α level for the $\chi^2_{(\alpha)}$ stopping criterion in conjunction with forward stepwise logistic regression. Levels of α can vary from $\alpha = 0$ (no predictors in the logistic model) to $\alpha = 1$ (all predictors in the logistic model); levels of α determine the size of the subset model. The usual conventional level for α has been 0.05. However, previously it was not known whether $\alpha = 0.05$ was the best level in the logistic model for the purpose of prediction.

The 'best' choice of α varies between 0.05 and 0.40 in both multivariate normal and multivariate binary cases. In the multivariate normal case, the choice of α depends upon the factor P which is the number of predictor variables in the data; a larger α level should be specified as the number of predictor variables increases. In the multivariate binary case, the choice of α level depends upon the factor $(p_1 - p_0)$ which is the difference between the two means of binary variables in population 1 and population 0; a smaller α level should be specified as the two populations become more distinct. These results imply that uniform specification of the 'best' α level for the standard $\chi^2_{(\alpha)}$ stopping criterion cannot be made. However, if a recommendation had to be made, we would recommend that $0.15 \leq \alpha \leq 0.20$ be used in forward stepwise logistic regression for the purpose of prediction.

In section 5.2 we have compared the six selection criteria LS, WD, SC, PH, LK, and SW with the LR selection criterion with respect to the order of variables in forward stepwise logistic regression. It has been shown that the LS selection criterion always selected the variables in the same order as the LR selection criterion in both multivariate normal and multivariate binary cases. This means that the LS selection criterion was an excellent approximation to the LR selection criterion over a region around $\boldsymbol{\beta}$ in each of all sampling situations in this study. Further study is needed to investigate the adequacy of the approximation of the LS selection criterion to the LR selection criterion in other situations. Since the LS selection criterion also needs the MLE for the next variable to be selected, as far as the forward selection procedure is concerned, the LS selection criterion has little gain over the LR selection criterion in terms of computation.

It has been shown that the LK selection criterion is almost equivalent to the SC selection criterion in selecting the variables; the LK selection criterion may be used as an alternative to the SC selection criterion becaused it requires less computation. Since the distribution of the LK selection criterion is not known and was assumed to follow a χ^2 distribution, further work is necessary to investigate the adequacy of the χ^2 distribution of the LK selection criterion.

Since the WD selection criterion had a large chance (27 percent and 6 percent in the multivariate normal and multivariate binary cases, respectively) of selecting a different variable at the first step in the forward selection procedure, we advise that the WD selection criterion should be avoided, whenever possible, as an alternative to the LR selection criterion. This result confirms the strange behavior of the Wald statistic illustrated by Hauck and Donner (1977).

The SW criterion has been obtained by a stepwise algorithm which adapted the sweep operator (Beaton, 1964). Since it does not need to recompute $\boldsymbol{\beta}$ which must be computed by the other methods, this criterion is the fastest one among all other criteria. This criterion is potentially useful in situations with a very large number of variables, say a survey with hundreds of variables. A reasonable procedure is to reduce the number of variables by picking important variables using the SW, then those variables would be examined by using more computationally intensive computer procedures.

It is interesting to see that the factors Δ^2 and $(p_1 - p_0)$ had the largest effect on the order of variables in the multivariate normal and multivariate binary cases, respectively; the larger the value of the factors Δ^2 and $(p_1 - p_0)$, the greater the difference in values between the LR selection criterion and the other selection criteria. This implies that there is a higher chance of selecting a different variable as Δ^2 and $(p_1 - p_0)$ increase.

In section 5.3 we have investigated the effects of the factors of predictor variables on ARR, Bias, and $\hat{\text{ERR}}$. The structures of predictor variables depend on the five factors P, V, Δ^2, D, and N in the multivariate normal case and the four factors P, p_0, $(p_1 - p_0)$, and N in the multivariate binary case.

We have shown that the factor Δ^2 had the largest effect on ARR and $E\hat{R}R$, whereas the factor P had the largest effect on Bias in the multivariate normal case, and that the factor $(p_1 - p_0)$ had the largest effect on ARR, Bias, and $E\hat{R}R$ in the mul'ivariate binary case.

The following interpretations may made based on these results.

1. Better prediction can be made with the variables which characterize the populations more distintly.
2. We would not include any variables which have little information on the difference between two populations. This means that the final model should be based on a small number of predictor variables which have good discriminating power.
3. In contrast to widely accepted notion, the larger sample size is not necessarily beneficial for the prediction purpose. Many uninformative observations may reduce the predictive ability of the model. The prediction largely depends on the quality of the data, not on the size of the data. A better prediction can be made on the basis of a smaller data set which has more informative observations than a larger data set which has fewer informative observations.
4. The multicollinearity among predictor variables has little effect on ARR, Bias, and $E\hat{R}R$.

In section 5.4 we compared the size of subset models for the five stopping criteria with the LR selection criterion. The descending order for the overall size of subset models for the five stopping criteria was $E_m > AIC_m > \chi^2_{(0.20)} > C_{pm} > SCH_m$ in the multivariate normal case, and $AIC_m > \chi^2_{(0.15)} > E_m > C_{pm} > SCH_m$ in the multivariate binary case. In both cases, the AIC_m and $\chi^2_{(\alpha)}$ criteria with $\alpha = 0.15$ or 0.20 tend to have the same size of subset models. It is interesting to see the change in the ranking of the E_m stopping criterion between the multivariate normal and multivariate binary cases.

In section 5.5 we compared the performance of the seven selection criteria and the five stopping criteria in term of $E\hat{R}R$. It was shown that there were no differences in $E\hat{R}R$ among the seven selection criteria. However, it was shown that there were statistically significant differences in $E\hat{R}R$ between the E_m stopping criterion and the other stopping criteria including the standard stopping criterion $\chi^2_{(\alpha)}$ with $\alpha = 0.15$ or 0.20. The E_m stopping criterion had the smallest $E\hat{R}R$ in both the multivariate normal and multivariate binary cases.

There is a need for further study of selection and stopping criteria for the prediction purpose in forward stepwise logistic regression. Until now the model derived from forward stepwise procedure has been used for both the description and prediction purposes. All of the selection criteria employed in statistical computer packages such as GLIM, BMDP, SAS, and SPSS in forward stepwise logistic regression are based on the significance or p-value of the predictor variables. This means that the procedures of model-building in GLIM, BMDP, SAS, and SPSS are compatible with the descriptive purpose of the model, but not compatible with the predictive purpose of the model. In other words, the variables selected for the one purpose are used in the other purpose. This practice is a major contradiction. Hence it is clear that alternative selection and stopping criteria should be developed and evaluated for prediction in the logistic model.

A selection criterion based on the error rate of prediction would be appropriate; given the variables previously selected the selection criterion would select a variable that gives the smallest error rate of prediction among non-selected variables. A stopping criterion based on probabilistic arguments would be particularly valuable; it is surely desirable to be able to claim the the addition of the next variable does not provide a significant reduction in the error rate of prediction. The ideal procedure for the prediction purpose would be then the combination of these selection and stopping criteria.

The computational requirement for such a procedure is very extensive. However, it is now feasible to run such a procedure with the help of the high speed computers currently available.

The resolution of these important issues will make the logistic model more useful for prediction purposes in medical and epidemiologic studies.

APPENDIX A

SAS/IML program for generating multivariate normal variables

```
filename david '/usr/accts/david/simulate/crun.dat';
libname lee '/usr/accts/david/simulate/';

data lee.rundat;
 infile david;
 input p v deltasq d n;

proc iml;
 use lee.rundat;
 read all var _all_ into runpar;
 firstset=1;
 newrun=1;

start runrep;
 do r=1 to 48;
  do s=1 to 20;
   param=runpar[r,];
   p=param[,1];
   v=param[,2];
   deltasq=param[,3];
   d=param[,4];
   n=param[,5];
   seed_v=shape(s,p,1);
   if (s=1 & newrun=1) then simcon1;
   else run simcon2;
   newrun=0;
  end;
  newrun=1;
 end;
finish runrep;
```

```
start simcon1;
 run initmn;
 run multinor;
 run data;
finish simcon1;

start simcon2;
 run multinor;
 run data;
finish simcon2;

start initmn;
 lambda_v=shape(0,p,1);
 mustar_v=shape(0,p,1);
 i=1;
 if v=1 then a=0.9;
 else a=0.9*p*(1-v)/(1-v**p);
 do while (i<=p);
  lambda_s=a*v**(i-1)+0.1;
  lambda_v[i]=lambda_s;
  i=i+1;
 end;
 diag_m=diag(lambda_v);
 orth_m=orpol(p:1,p-1);
 sigma_m=orth_m*diag_m*orth_m`;
 i=1;
 if d=1 then b=deltasq/p;
 else b=deltasq**2*(1-d)/(1-d**p);
 do while (i<=p);
  mustar_s=sqrt(b*d**(i-1));
  mustar_v[i]=mustar_s;
  i=i+1;
 end;
 chol_m=root(sigma_m);
 mu_v=chol_m*mustar_v;
finish initmn;
```

```
start multinor;
 normal_v=normal(seed_v);
 y1_x_m=mu_v+chol_m*normal_v;
 i=1;
 do while (i<n);
  normal_v=normal(seed_v);
  y1_x_m=y1_x_m||mu_v+chol_m*normal_v;
  i=i+1;
 end;
 y1_x_t_m=y1_x_m`;
 one=j(n,1,1);
 y1_x_t_m=one||y1_x_t_m;
 normal_v=normal(seed_v);
 y0_x_m=chol_m*normal_v;
 i=1;
 do while (i<n);
  normal_v=normal(seed_v);
  y0_x_m=y0_x_m||chol_m*normal_v;
  i=i+1;
 end;
 y0_x_t_m=y0_x_m`;
 zero=j(n,1,0);
 y0_x_t_m=zero||y0_x_t_m;
finish multinor;

start data;
 data=y0_x_t_m//y1_x_t_m;
finish data;

run runrep;
quit;
endsas;
```

APPENDIX B

SAS/IML program for generating multivariate binary variables

```
filename david '/usr/accts/david/dsim/drun.dat';
libname lee '/usr/accts/david/dsim/';

data lee.rundat;
 infile david;
 input p n pj p1j rho rho1;

proc iml;
 use lee.rundat;
 read all var _all_ into runpar;
 firstset=1;

start runrep;
 do r=1 to 81;
  do s=1 to 20;
   param=runpar[r,];
   p=param[,1];
   n=param[,2];
   p0j=param[,3];
   p1j=param[,4];
   rho0=param[,5];
   rho1=param[,6];
   run=r;
   rep=s;
   if p=10 then do;
    i=1;
    do while (i<=n);
     pj=p0j;
     rho=rho0;
     run x_p10;
     if i=1 then data0=x1||x2||x3||x4||x5||x6||x7||x8||x9||x10;
     else data0=data0//(x1||x2||x3||x4||x5||x6||x7||x8||x9||x10);
```

```
  pj=p1j;
  rho=rho1;
  run x_p10;
  if i=1 then data1=x1||x2||x3||x4||x5||x6||x7||x8||x9||x10;
  else data1=data1//(x1||x2||x3||x4||x5||x6||x7||x8||x9||x10);
  i=i+1;
 end;
 p0j_bar=data0[+,]/n;
 p1j_bar=data1[+,]/n;
 zero=j(n,1,0);
 one=j(n,1,1);
 data0=zero||data0;
 data1=one||data1;
 data=data0//data1;
end;
if p=15 then do;
 i=1;
 do while (i<=n);
  pj=p0j;
  rho=rho0;
  run x_p10;
  run x_p15;
  if i=1 then data0=x1||x2||x3||x4||x5||x6||x7||x8||
                  x9||x10||x11||x12||x13||x14||x15;
  else data0=data0//(x1||x2||x3||x4||x5||x6||x7||x8||x9||x10||
                  x11||x12||x13||x14||x15);
  pj=p1j;
  rho=rho1;
  run x_p10;
  run x_p15;
  if i=1 then data1=x1||x2||x3||x4||x5||x6||x7||x8||
                  x9||x10||x11||x12||x13||x14||x15;
  else data1=data1//(x1||x2||x3||x4||x5||x6||x7||x8||x9||x10||
                  x11||x12||x13||x14||x15);
  i=i+1;
 end;
 p0j_bar=data0[+,]/n;
 p1j_bar=data1[+,]/n;
 zero=j(n,1,0);
```

```
      one=j(n,1,1);
      data0=zero||data0;
      data1=one||data1;
      data=data0//data1;
    end;
    if p=20 then do;
      i=1;
      do while (i<=n);
        pj=p0j;
        rho=rho0;
        run x_p10;
        run x_p15;
        run x_p20;
        if i=1 then data0=x1||x2||x3||x4||x5||x6||x7||x8||x9||x10||
                          x11||x12||x13||x14||x15||x16||x17||x18||x19||x20;
        else data0=data0//(x1||x2||x3||x4||x5||x6||x7||x8||x9||x10||
                          x11||x12||x13||x14||x15||x16||x17||x18||x19||x20);
        pj=p1j;
        rho=rho1;
        run x_p10;
        run x_p15;
        run x_p20;
        if i=1 then data1=x1||x2||x3||x4||x5||x6||x7||x8||x9||x10||
                          x11||x12||x13||x14||x15||x16||x17||x18||x19||x20;
        else data1=data1//(x1||x2||x3||x4||x5||x6||x7||x8||x9||x10||
                          x11||x12||x13||x14||x15||x16||x17||x18||x19||x20);
        i=i+1;
      end;
      p0j_bar=data0[+,]/n;
      p1j_bar=data1[+,]/n;
      zero=j(n,1,0);
      one=j(n,1,1);
      data0=zero||data0;
      data1=one||data1;
      data=data0//data1;
    end;
  end;
 end;
finish runrep;
```

```
start x_p10;
 x1=1;
 z1=(x1-pj)/sqrt(pj*(1-pj));
 fx1=pj;
 u=uniform(1*r*s);
 if (u>fx1) then do;
  x1=0;
  z1=(x1-pj)/sqrt(pj*(1-pj));
 end;
 x2=1;
 z2=(x2-pj)/sqrt(pj*(1-pj));
 u=uniform(2*r*s);
 cfa=(1+rho*z1*z2);
 fx2=pj*cfa;
 if (u>fx2) then do;
  x2=0;
  z2=(x2-pj)/sqrt(pj*(1-pj));
 end;
 x3=1;
 z3=(x3-pj)/sqrt(pj*(1-pj));
 u=uniform(3*r*s);
 cfb=(1+rho*(z1*z2+z1*z3+z2*z3));
 fx3=pj*cfb/cfa;
 if (u>fx3) then do;
  x3=0;
  z3=(x3-pj)/sqrt(pj*(1-pj));
 end;
 x4=1;
 z4=(x4-pj)/sqrt(pj*(1-pj));
 u=uniform(4*r*s);
 cfa=(1+rho*(z1*z2+z1*z3+z1*z4+z2*z3+z2*z4+z3*z4));
 fx4=pj*cfa/cfb;
 if (u>fx4) then do;
  x4=0;
  z4=(x4-pj)/sqrt(pj*(1-pj));
 end;
 x5=1;
 z5=(x5-pj)/sqrt(pj*(1-pj));
 u=uniform(5*r*s);
```

```
cfb=(1+rho*(z1*z2+z1*z3+z1*z4+z1*z5+z2*z3+z2*z4+z2*z5+z3*z4+
     z3*z5+z4*z5));
fx5=pj*cfb/cfa;
if (u>fx5) then do;
 x5=0;
 z5=(x5-pj)/sqrt(pj*(1-pj));
end;
x6=1;
z6=(x6-pj)/sqrt(pj*(1-pj));
u=uniform(6*r*s);
cfa=(1+rho*(z1*z2+z1*z3+z1*z4+z1*z5+z1*z6+z2*z3+z2*z4+z2*z5+
     z2*z6+z3*z4+z3*z5+z3*z6+z4*z5+z4*z6+z5*z6));
fx6=pj*cfa/cfb;
if (u>fx6) then do;
 x6=0;
 z6=(x6-pj)/sqrt(pj*(1-pj));
end;
x7=1;
z7=(x7-pj)/sqrt(pj*(1-pj));
u=uniform(7*r*s);
cfb=(1+rho*(z1*z2+z1*z3+z1*z4+z1*z5+z1*z6+z1*z7+z2*z3+z2*z4+
     z2*z5+z2*z6+z2*z7+z3*z4+z3*z5+z3*z6+z3*z7+z4*z5+z4*z6+
     z4*z7+z5*z6+z5*z7+z6*z7));
fx7=pj*cfb/cfa;
if (u>fx7) then do;
 x7=0;
 z7=(x7-pj)/sqrt(pj*(1-pj));
end;
x8=1;
z8=(x8-pj)/sqrt(pj*(1-pj));
u=uniform(8*r*s);
cfa=(1+rho*(z1*z2+z1*z3+z1*z4+z1*z5+z1*z6+z1*z7+z1*z8+z2*z3+
     z2*z4+z2*z5+z2*z6+z2*z7+z2*z8+z3*z4+z3*z5+z3*z6+z3*z7+
     z3*z8+z4*z5+z4*z6+z4*z7+z4*z8+z5*z6+z5*z7+z5*z8+z6*z7+
     z6*z8+z7*z8));
fx8=pj*cfa/cfb;
if (u>fx8) then do;
 x8=0;
 z8=(x8-pj)/sqrt(pj*(1-pj));
```

```
 end;
 x9=1;
 z9=(x9-pj)/sqrt(pj*(1-pj));
 u=uniform(9*r*s);
 cfb=(1+rho*(z1*z2+z1*z3+z1*z4+z1*z5+z1*z6+z1*z7+z1*z8+z1*z9+
      z2*z3+z2*z4+z2*z5+z2*z6+z2*z7+z2*z8+z2*z9+z3*z4+z3*z5+
      z3*z6+z3*z7+z3*z8+z3*z9+z4*z5+z4*z6+z4*z7+z4*z8+z4*z9+
      z5*z6+z5*z7+z5*z8+z5*z9+z6*z7+z6*z8+z6*z9+z7*z8+z7*z9+
      z8*z9));
 fx9=pj*cfb/cfa;
 if (u>fx9) then do;
  x9=0;
  z9=(x9-pj)/sqrt(pj*(1-pj));
 end;
 x10=1;
 z10=(x10-pj)/sqrt(pj*(1-pj));
 u=uniform(10*r*s);
 cfa=(1+rho*(z1*z2+z1*z3+z1*z4+z1*z5+z1*z6+z1*z7+z1*z8+z1*z9+
      z1*z10+z2*z3+z2*z4+z2*z5+z2*z6+z2*z7+z2*z8+z2*z9+z2*z10+
      z3*z4+z3*z5+z3*z6+z3*z7+z3*z8+z3*z9+z3*z10+z4*z5+z4*z6+
      z4*z7+z4*z8+z4*z9+z4*z10+z5*z6+z5*z7+z5*z8+z5*z9+z5*z10+
      z6*z7+z6*z8+z6*z9+z6*z10+z7*z8+z7*z9+z7*z10+z8*z9+z8*z10+
      z9*z10));
 fx10=pj*cfa/cfb;
 if (u>fx10) then do;
  x10=0;
  z10=(x10-pj)/sqrt(pj*(1-pj));
 end;
finish x_p10;

start x_p15;
 x11=1;
 z11=(x11-pj)/sqrt(pj*(1-pj));
 u=uniform(11*r*s);
 cfb=(1+rho*(z1*z2+z1*z3+z1*z4+z1*z5+z1*z6+z1*z7+z1*z8+z1*z9+
      z1*z10+z1*z11+z2*z3+z2*z4+z2*z5+z2*z6+z2*z7+z2*z8+z2*z9+
      z2*z10+z2*z11+z3*z4+z3*z5+z3*z6+z3*z7+z3*z8+z3*z9+z3*z10+
      z3*z11+z4*z5+z4*z6+z4*z7+z4*z8+z4*z9+z4*z10+z4*z11+z5*z6+
      z5*z7+z5*z8+z5*z9+z5*z10+z5*z11+z6*z7+z6*z8+z6*z9+z6*z10+
```

```
     z6*z11+z7*z8+z7*z9+z7*z10+z7*z11+z8*z9+z8*z10+z8*z11+
     z9*z10+z9*z11+z10*z11));
fx11=pj*cfb/cfa;
if (u>fx11) then do;
 x11=0;
 z11=(x11-pj)/sqrt(pj*(1-pj));
end;
x12=1;
z12=(x12-pj)/sqrt(pj*(1-pj));
u=uniform(12*r*s);
cfa=(1+rho*(z1*z2+z1*z3+z1*z4+z1*z5+z1*z6+z1*z7+z1*z8+z1*z9+
     z1*z10+z1*z11+z1*z12+z2*z3+z2*z4+z2*z5+z2*z6+z2*z7+z2*z8+
     z2*z9+z2*z10+z2*z11+z2*z12+z3*z4+z3*z5+z3*z6+z3*z7+z3*z8+
     z3*z9+z3*z10+z3*z11+z3*z12+z4*z5+z4*z6+z4*z7+z4*z8+z4*z9+
     z4*z10+z4*z11+z4*z12+z5*z6+z5*z7+z5*z8+z5*z9+z5*z10+
     z5*z11+z5*z12+z6*z7+z6*z8+z6*z9+z6*z10+z6*z11+z6*z12+
     z7*z8+z7*z9+z7*z10+z7*z11+z7*z12+z8*z9+z8*z10+z8*z11+
     z8*z12+z9*z10+z9*z11+z9*z12+z10*z11+z10*z12+z11*z12));
fx12=pj*cfa/cfb;
if (u>fx12) then do;
 x12=0;
 z12=(x12-pj)/sqrt(pj*(1-pj));
end;
x13=1;
z13=(x13-pj)/sqrt(pj*(1-pj));
u=uniform(13*r*s);
cfb=(1+rho*(z1*z2+z1*z3+z1*z4+z1*z5+z1*z6+z1*z7+z1*z8+z1*z9+
     z1*z10+z1*z11+z1*z12+z1*z13+z2*z3+z2*z4+z2*z5+z2*z6+
     z2*z7+z2*z8+z2*z9+z2*z10+z2*z11+z2*z12+z2*z13+z3*z4+
     z3*z5+z3*z6+z3*z7+z3*z8+z3*z9+z3*z10+z3*z11+z3*z12+
     z3*z13+z4*z5+z4*z6+z4*z7+z4*z8+z4*z9+z4*z10+z4*z11+
     z4*z12+z4*z13+z5*z6+z5*z7+z5*z8+z5*z9+z5*z10+z5*z11+
     z5*z12+z5*z13+z6*z7+z6*z8+z6*z9+z6*z10+z6*z11+z6*z12+
     z6*z13+z7*z8+z7*z9+z7*z10+z7*z11+z7*z12+z7*z13+z8*z9+
     z8*z10+z8*z11+z8*z12+z8*z13+z9*z10+z9*z11+z9*z12+z9*z13+
     z10*z11+z10*z12+z10*z13+z11*z12+z11*z13+z12*z13));
fx13=pj*cfb/cfa;
if (u>fx13) then do;
 x13=0;
```

```
 z13=(x13-pj)/sqrt(pj*(1-pj));
end;
x14=1;
z14=(x14-pj)/sqrt(pj*(1-pj));
u=uniform(14*r*s);
cfa=(1+rho*(z1*z2+z1*z3+z1*z4+z1*z5+z1*z6+z1*z7+z1*z8+z1*z9+
     z1*z10+z1*z11+z1*z12+z1*z13+z1*z14+z2*z3+z2*z4+z2*z5+
     z2*z6+z2*z7+z2*z8+z2*z9+z2*z10+z2*z11+z2*z12+z2*z13+
     z2*z14+z3*z4+z3*z5+z3*z6+z3*z7+z3*z8+z3*z9+z3*z10+
     z3*z11+z3*z12+z3*z13+z3*z14+z4*z5+z4*z6+z4*z7+z4*z8+
     z4*z9+z4*z10+z4*z11+z4*z12+z4*z13+z4*z14+z5*z6+z5*z7+
     z5*z8+z5*z9+z5*z10+z5*z11+z5*z12+z5*z13+z5*z14+z6*z7+
     z6*z8+z6*z9+z6*z10+z6*z11+z6*z12+z6*z13+z6*z14+z7*z8+
     z7*z9+z7*z10+z7*z11+z7*z12+z7*z13+z7*z14+z8*z9+z8*z10+
     z8*z11+z8*z12+z8*z13+z8*z14+z9*z10+z9*z11+z9*z12+z9*z13+
     z9*z14+z10*z11+z10*z12+z10*z13+z10*z14+z11*z12+z11*z13+
     z11*z14+z12*z13+z12*z14+z13*z14));
fx14=pj*cfa/cfb;
if (u>fx14) then do;
 x14=0;
 z14=(x14-pj)/sqrt(pj*(1-pj));
end;
x15=1;
z15=(x15-pj)/sqrt(pj*(1-pj));
u=uniform(15*r*s);
cfb=(1+rho*(z1*z2+z1*z3+z1*z4+z1*z5+z1*z6+z1*z7+z1*z8+z1*z9+
     z1*z10+z1*z11+z1*z12+z1*z13+z1*z14+z1*z15+z2*z3+z2*z4+
     z2*z5+z2*z6+z2*z7+z2*z8+z2*z9+z2*z10+z2*z11+z2*z12+
     z2*z13+z2*z14+z2*z15+z3*z4+z3*z5+z3*z6+z3*z7+z3*z8+
     z3*z9+z3*z10+z3*z11+z3*z12+z3*z13+z3*z14+z3*z15+z4*z5+
     z4*z6+z4*z7+z4*z8+z4*z9+z4*z10+z4*z11+z4*z12+z4*z13+
     z4*z14+z4*z15+z5*z6+z5*z7+z5*z8+z5*z9+z5*z10+z5*z11+
     z5*z12+z5*z13+z5*z14+z5*z15+z6*z7+z6*z8+z6*z9+z6*z10+
     z6*z11+z6*z12+z6*z13+z6*z14+z6*z15+z7*z8+z7*z9+z7*z10+
     z7*z11+z7*z12+z7*z13+z7*z14+z7*z15+z8*z9+z8*z10+z8*z11+
     z8*z12+z8*z13+z8*z14+z8*z15+z9*z10+z9*z11+z9*z12+z9*z13+
     z9*z14+z9*z15+z10*z11+z10*z12+z10*z13+z10*z14+z10*z15+
     z11*z12+z11*z13+z11*z14+z11*z15+z12*z13+z12*z14+z12*z15+
     z13*z14+z13*z15+z14*z15));
```

```
 fx15=pj*cfb/cfa;
 if (u>fx15) then do;
  x15=0;
  z15=(x15-pj)/sqrt(pj*(1-pj));
 end;
finish x_p15;

start x_p20;
 x16=1;
 z16=(x16-pj)/sqrt(pj*(1-pj));
 u=uniform(16*r*s);
 cfa=(1+rho*(z1*z2+z1*z3+z1*z4+z1*z5+z1*z6+z1*z7+z1*z8+z1*z9+
      z1*z10+z1*z11+z1*z12+z1*z13+z1*z14+z1*z15+z1*z16+z2*z3+
      z2*z4+z2*z5+z2*z6+z2*z7+z2*z8+z2*z9+z2*z10+z2*z11+
      z2*z12+z2*z13+z2*z14+z2*z15+z2*z16+z3*z4+z3*z5+z3*z6+
      z3*z7+z3*z8+z3*z9+z3*z10+z3*z11+z3*z12+z3*z13+z3*z14+
      z3*z15+z3*z16+z4*z5+z4*z6+z4*z7+z4*z8+z4*z9+z4*z10+
      z4*z11+z4*z12+z4*z13+z4*z14+z4*z15+z4*z16+z5*z6+z5*z7+
      z5*z8+z5*z9+z5*z10+z5*z11+z5*z12+z5*z13+z5*z14+z5*z15+
      z5*z16+z6*z7+z6*z8+z6*z9+z6*z10+z6*z11+z6*z12+z6*z13+
      z6*z14+z6*z15+z6*z16+z7*z8+z7*z9+z7*z10+z7*z11+z7*z12+
      z7*z13+z7*z14+z7*z15+z7*z16+z8*z9+z8*z10+z8*z11+z8*z12+
      z8*z13+z8*z14+z8*z15+z8*z16+z9*z10+z9*z11+z9*z12+z9*z13+
      z9*z14+z9*z15+z9*z16+z10*z11+z10*z12+z10*z13+z10*z14+
      z10*z15+z10*z16+z11*z12+z11*z13+z11*z14+z11*z15+z11*z16+
      z12*z13+z12*z14+z12*z15+z12*z16+z13*z14+z13*z15+z13*z16+
      z14*z15+z14*z16+z15*z16));
 fx16=pj*cfa/cfb;
 if (u>fx16) then do;
  x16=0;
  z16=(x16-pj)/sqrt(pj*(1-pj));
 end;
 x17=1;
 z17=(x17-pj)/sqrt(pj*(1-pj));
 u=uniform(17*r*s);
 cfb=(1+rho*(z1*z2+z1*z3+z1*z4+z1*z5+z1*z6+z1*z7+z1*z8+z1*z9+
      z1*z10+z1*z11+z1*z12+z1*z13+z1*z14+z1*z15+z1*z16+z1*z17+
      z2*z3+z2*z4+z2*z5+z2*z6+z2*z7+z2*z8+z2*z9+z2*z10+z2*z11+
      z2*z12+z2*z13+z2*z14+z2*z15+z2*z16+z2*z17+z3*z4+z3*z5+
```

```
     z3*z6+z3*z7+z3*z8+z3*z9+z3*z10+z3*z11+z3*z12+z3*z13+
     z3*z14+z3*z15+z3*z16+z3*z17+z4*z5+z4*z6+z4*z7+z4*z8+
     z4*z9+z4*z10+z4*z11+z4*z12+z4*z13+z4*z14+z4*z15+z4*z16+
     z4*z17+z5*z6+z5*z7+z5*z8+z5*z9+z5*z10+z5*z11+z5*z12+
     z5*z13+z5*z14+z5*z15+z5*z16+z5*z17+z6*z7+z6*z8+z6*z9+
     z6*z10+z6*z11+z6*z12+z6*z13+z6*z14+z6*z15+z6*z16+z6*z17+
     z7*z8+z7*z9+z7*z10+z7*z11+z7*z12+z7*z13+z7*z14+z7*z15+
     z7*z16+z7*z17+z8*z9+z8*z10+z8*z11+z8*z12+z8*z13+z8*z14+
     z8*z15+z8*z16+z8*z17+z9*z10+z9*z11+z9*z12+z9*z13+z9*z14+
     z9*z15+z9*z16+z9*z17+z10*z11+z10*z12+z10*z13+z10*z14+
     z10*z15+z10*z16+z10*z17+z11*z12+z11*z13+z11*z14+z11*z15+
     z11*z16+z11*z17+z12*z13+z12*z14+z12*z15+z12*z16+z12*z17+
     z13*z14+z13*z15+z13*z16+z13*z17+z14*z15+z14*z16+z14*z17+
     z15*z16+z15*z17+z16*z17));
fx17=pj*cfb/cfa;
if (u>fx17) then do;
  x17=0;
  z17=(x17-pj)/sqrt(pj*(1-pj));
end;
x18=1;
z18=(x18-pj)/sqrt(pj*(1-pj));
u=uniform(18*r*s);
cfa=(1+rho*(z1*z2+z1*z3+z1*z4+z1*z5+z1*z6+z1*z7+z1*z8+z1*z9+
     z1*z10+z1*z11+z1*z12+z1*z13+z1*z14+z1*z15+z1*z16+z1*z17+
     z1*z18+z2*z3+z2*z4+z2*z5+z2*z6+z2*z7+z2*z8+z2*z9+z2*z10+
     z2*z11+z2*z12+z2*z13+z2*z14+z2*z15+z2*z16+z2*z17+z2*z18+
     z3*z4+z3*z5+z3*z6+z3*z7+z3*z8+z3*z9+z3*z10+z3*z11+z3*z12+
     z3*z13+z3*z14+z3*z15+z3*z16+z3*z17+z3*z18+z4*z5+z4*z6+
     z4*z7+z4*z8+z4*z9+z4*z10+z4*z11+z4*z12+z4*z13+z4*z14+
     z4*z15+z4*z16+z4*z17+z4*z18+z5*z6+z5*z7+z5*z8+z5*z9+
     z5*z10+z5*z11+z5*z12+z5*z13+z5*z14+z5*z15+z5*z16+z5*z17+
     z5*z18+z6*z7+z6*z8+z6*z9+z6*z10+z6*z11+z6*z12+z6*z13+
     z6*z14+z6*z15+z6*z16+z6*z17+z6*z18+z7*z8+z7*z9+z7*z10+
     z7*z11+z7*z12+z7*z13+z7*z14+z7*z15+z7*z16+z7*z17+z7*z18+
     z8*z9+z8*z10+z8*z11+z8*z12+z8*z13+z8*z14+z8*z15+z8*z16+
     z8*z17+z8*z18+z9*z10+z9*z11+z9*z12+z9*z13+z9*z14+z9*z15+
     z9*z16+z9*z17+z9*z18+z10*z11+z10*z12+z10*z13+z10*z14+
     z10*z15+z10*z16+z10*z17+z10*z18+z11*z12+z11*z13+z11*z14+
     z11*z15+z11*z16+z11*z17+z11*z18+z12*z13+z12*z14+z12*z15+
```

```
     z12*z16+z12*z17+z12*z18+z13*z14+z13*z15+z13*z16+z13*z17+
     z13*z18+z14*z15+z14*z16+z14*z17+z14*z18+z15*z16+z15*z17+
     z15*z18+z16*z17+z16*z18+z17*z18));
fx18=pj*cfa/cfb;
if (u>fx18) then do;
 x18=0;
 z18=(x18-pj)/sqrt(pj*(1-pj));
end;
x19=1;
z19=(x19-pj)/sqrt(pj*(1-pj));
u=uniform(19*r*s);
cfb=(1+rho*(z1*z2+z1*z3+z1*z4+z1*z5+z1*z6+z1*z7+z1*z8+z1*z9+
     z1*z10+z1*z11+z1*z12+z1*z13+z1*z14+z1*z15+z1*z16+z1*z17+
     z1*z18+z1*z19+z2*z3+z2*z4+z2*z5+z2*z6+z2*z7+z2*z8+z2*z9+
     z2*z10+z2*z11+z2*z12+z2*z13+z2*z14+z2*z15+z2*z16+z2*z17+
     z2*z18+z2*z19+z3*z4+z3*z5+z3*z6+z3*z7+z3*z8+z3*z9+
     z3*z10+z3*z11+z3*z12+z3*z13+z3*z14+z3*z15+z3*z16+z3*z17+
     z3*z18+z3*z19+z4*z5+z4*z6+z4*z7+z4*z8+z4*z9+z4*z10+
     z4*z11+z4*z12+z4*z13+z4*z14+z4*z15+z4*z16+z4*z17+z4*z18+
     z4*z19+z5*z6+z5*z7+z5*z8+z5*z9+z5*z10+z5*z11+z5*z12+
     z5*z13+z5*z14+z5*z15+z5*z16+z5*z17+z5*z18+z5*z19+z6*z7+
     z6*z8+z6*z9+z6*z10+z6*z11+z6*z12+z6*z13+z6*z14+z6*z15+
     z6*z16+z6*z17+z6*z18+z6*z19+z7*z8+z7*z9+z7*z10+z7*z11+
     z7*z12+z7*z13+z7*z14+z7*z15+z7*z16+z7*z17+z7*z18+z7*z19+
     z8*z9+z8*z10+z8*z11+z8*z12+z8*z13+z8*z14+z8*z15+z8*z16+
     z8*z17+z8*z18+z8*z19+z9*z10+z9*z11+z9*z12+z9*z13+z9*z14+
     z9*z15+z9*z16+z9*z17+z9*z18+z9*z19+z10*z11+z10*z12+
     z10*z13+z10*z14+z10*z15+z10*z16+z10*z17+z10*z18+z10*z19+
     z11*z12+z11*z13+z11*z14+z11*z15+z11*z16+z11*z17+z11*z18+
     z11*z19+z12*z13+z12*z14+z12*z15+z12*z16+z12*z17+z12*z18+
     z12*z19+z13*z14+z13*z15+z13*z16+z13*z17+z13*z18+z13*z19+
     z14*z15+z14*z16+z14*z17+z14*z18+z14*z19+z15*z16+z15*z17+
     z15*z18+z15*z19+z16*z17+z16*z18+z16*z19+z17*z18+z17*z19+
     z18*z19));
fx19=pj*cfb/cfa;
if (u>fx19) then do;
 x19=0;
 z19=(x19-pj)/sqrt(pj*(1-pj));
end;
```

```
 x20=1;
 z20=(x20-pj)/sqrt(pj*(1-pj));
 u=uniform(20*r*s);
 cfa=(1+rho*(z1*z2+z1*z3+z1*z4+z1*z5+z1*z6+z1*z7+z1*z8+z1*z9+
       z1*z10+z1*z11+z1*z12+z1*z13+z1*z14+z1*z15+z1*z16+z1*z17+
       z1*z18+z1*z19+z1*z20+z2*z3+z2*z4+z2*z5+z2*z6+z2*z7+
       z2*z8+z2*z9+z2*z10+z2*z11+z2*z12+z2*z13+z2*z14+z2*z15+
       z2*z16+z2*z17+z2*z18+z2*z19+z2*z20+z3*z4+z3*z5+z3*z6+
       z3*z7+z3*z8+z3*z9+z3*z10+z3*z11+z3*z12+z3*z13+z3*z14+
       z3*z15+z3*z16+z3*z17+z3*z18+z3*z19+z3*z20+z4*z5+z4*z6+
       z4*z7+z4*z8+z4*z9+z4*z10+z4*z11+z4*z12+z4*z13+z4*z14+
       z4*z15+z4*z16+z4*z17+z4*z18+z4*z19+z4*z20+z5*z6+z5*z7+
       z5*z8+z5*z9+z5*z10+z5*z11+z5*z12+z5*z13+z5*z14+z5*z15+
       z5*z16+z5*z17+z5*z18+z5*z19+z5*z20+z6*z7+z6*z8+z6*z9+
       z6*z10+z6*z11+z6*z12+z6*z13+z6*z14+z6*z15+z6*z16+z6*z17+
       z6*z18+z6*z19+z6*z20+z7*z8+z7*z9+z7*z10+z7*z11+z7*z12+
       z7*z13+z7*z14+z7*z15+z7*z16+z7*z17+z7*z18+z7*z19+z7*z20+
       z8*z9+z8*z10+z8*z11+z8*z12+z8*z13+z8*z14+z8*z15+z8*z16+
       z8*z17+z8*z18+z8*z19+z8*z20+z9*z10+z9*z11+z9*z12+z9*z13+
       z9*z14+z9*z15+z9*z16+z9*z17+z9*z18+z9*z19+z9*z20+z10*z11+
       z10*z12+z10*z13+z10*z14+z10*z15+z10*z16+z10*z17+z10*z18+
       z10*z19+z10*z20+z11*z12+z11*z13+z11*z14+z11*z15+z11*z16+
       z11*z17+z11*z18+z11*z19+z11*z20+z12*z13+z12*z14+z12*z15+
       z12*z16+z12*z17+z12*z18+z12*z19+z12*z20+z13*z14+z13*z15+
       z13*z16+z13*z17+z13*z18+z13*z19+z13*z20+z14*z15+z14*z16+
       z14*z17+z14*z18+z14*z19+z14*z20+z15*z16+z15*z17+z15*z18+
       z15*z19+z15*z20+z16*z17+z16*z18+z16*z19+z16*z20+z17*z18+
       z17*z19+z17*z20+z18*z19+z18*z20+z19*z20));
 fx20=pj*cfa/cfb;
 if (u>fx20) then do;
  x20=0;
  z20=(x20-pj)/sqrt(pj*(1-pj));
 end;
finish x_p20;

run runrep;
run;
endsas;
```

APPENDIX C

SAS/IML program for sections 9.1 and 9.3

```
libname lee '/usr/accts/david/simulate/';

start selecrit;
 run initsele;
 run lintcept;
 run fullx;
 run selelr;

 run initsele;
 run lintcept;
 run selels;

 run initsw;
 run lintcept;
 run selesw;

 run initsele;
 run lintcept;
 run selesc;

 run initsele;
 run lintcept;
 run seleph;

 run initsele;
 run lintcept;
 run selelk;

 run initsele;
 run lintcept;
 run selewd;
finish selecrit;
```

```
start initsele;
 run=r;
 rep=s;
 y=data[,1];
 k=ncol(data);
 cindx=2:k;
 x=repeat(1,nrow(data),1)||(data[,cindx]);
 in=repeat(0,k,1);
 arr=repeat(9,k,1);
 bias=repeat(9,k,1);
 err=repeat(9,k,1);
 l=1;
 fx=x;
 px=x[,1];
 in[1,]=1;
 signal05=1; signal10=1; signal15=1; signal20=1; signal25=1; signal30=1;
 signal35=1; signal40=1; signal45=1; signal50=1; signal55=1; signal60=1;
 signal65=1; signal70=1; signal75=1; signal80=1; signal85=1; signal90=1;
finish initsele;

start initsw;
 run=r;
 rep=s;
 y=data[,1];
 k=ncol(data);
 cindx=2:k;
 x=repeat(1,nrow(data),1)||(data[,cindx]);
 in=repeat(0,k,1);
 arr=repeat(9,k,1);
 bias=repeat(9,k,1);
 err=repeat(9,k,1);
 l=1;
 fx=x;
 run fullx;
 fpi=choose(fpi=0,0.0000001,fpi);
 g=log(fpi/(1-fpi));
 x2=fx[,2:k];
 subx=x[,1];
 px=subx;
```

```
 sweepx=(subx#fvi)`*subx;
 x1tx2=(subx#fvi)`*x2;
 x1tg=(subx#fvi)`*g;
 x2tx1=x2`*(subx#fvi);
 x2tx2=(x2#fvi)`*x2;
 x2tg=(x2#fvi)`*g;
 gtx1=g`*(subx#fvi);
 gtx2=g`*(x2#fvi);
 gtg=g`*g;
 sweepm=(sweepx||x1tx2||x1tg)//(x2tx1||x2tx2||x2tg)//(gtx1||gtx2||gtg);
 swp=sweep(sweepm,l);
 ssr0=swp[k+1,k+1];
 in[1,]=1;
 signal05=1; signal10=1; signal15=1; signal20=1; signal25=1; signal30=1;
 signal35=1; signal40=1; signal45=1; signal50=1; signal55=1; signal60=1;
 signal65=1; signal70=1; signal75=1; signal80=1; signal85=1; signal90=1;
finish initsw;

start lintcept;
 pi0=sum((y=1)#y)/nrow(fx);
 lintcept=sum((y=1)#log(pi0)+(y=0)#log(1-pi0));
 loglik0=lintcept;
 arr[1,]=.;
 bias[1,]=.;
 err[1,]=.;
finish lintcept;

start fullx;
 fb=repeat(0,ncol(fx),1);
 foldb=fb+1;
 do iter=1 to 10 while (max(abs(fb-foldb))>1e-2);
  foldb=fb;
  fz=fx*fb;
  fpi=1/(1+exp(-fz));
  fpi=choose(fpi=0,0.0000001,fpi);
  fpi=choose(fpi=1,0.9999999,fpi);
  fvi=fpi#(1-fpi);
  fpiqix=fvi#fx;
  fw=1/fvi;
```

```
    fxpxi=inv(fpiqix`*(fw#fpiqix));
    fb=fb+fxpxi*(fpiqix`*(fw#(y-fpi)));
  end;
finish fullx;

start selelr;
  selecrit='1';
  flag=1;
  do while (flag=1);
    piest=repeat(0,nrow(data),k-1);
    loglik=repeat(-99999,k-1,1);
    index=loc(^in);
    i=1;
    do while (i<=(k-1));
      ci=index[,i];
      cx=fx[,ci];
      subx=px||cx;
      x=subx;
      run logitest;
      z=x*b;
      pi=1/(1+exp(-z));
      pi=choose(pi=0,0.0000001,pi);
      pi=choose(pi=1,0.9999999,pi);
      piest[,i]=pi;
      loglik[i,]=sum((y=1)#log(pi)+(y=0)#log(1-pi));
      i=i+1;
    end;
    run stopcrit;
  end;
finish selelr;

start selels;
  selecrit='2';
  flag=1;
  do while (flag=1);
    piest=repeat(0,nrow(data),k-1);
    loglik=repeat(-99999,k-1,1);
    loglikls=repeat(-99999,k-1,1);
    index=loc(^in);
```

```
  i=1;
  do while (i<=(k-l));
   ci=index[,i];
   cx=fx[,ci];
   subx=px||cx;
   x=subx;
   run logitest;
   z=x*b;
   pi=1/(1+exp(-z));
   pi=choose(pi=0,0.0000001,pi);
   pi=choose(pi=1,0.9999999,pi);
   piest[,i]=pi;
   loglik[i,]=sum((y=1)#log(pi)+(y=0)#log(1-pi));
   bls=b[1:l,]-xpxi[1:l,l+1]*(inv(xpxi[l+1,l+1]))*b[l+1,];
   zls=px*bls;
   pils=1/(1+exp(-zls));
   pils=choose(pils=0,0.0000001,pils);
   pils=choose(pils=1,0.9999999,pils);
   logliklsl[i,]=sum((y=1)#log(pils)+(y=0)#log(1-pils));
   i=i+1;
  end;
  run stopcrit;
 end;
finish selels;

start selesw;
 selecrit='3';
 flag=1;
 full=0;
 do while (flag=1);
  ssr=repeat(-99999,k-l,1);
  index=loc(^in);
  i=1;
  do while (i<=(k-l) & full=0);
   ci=index[,i];
   cx=fx[,ci];
   t=index#(index^=ci);
   if t=0 then run stopcrit;
   else do;
```

```
      nzindx=loc(t>0);
      index2=index[,nzindx];
      x2=fx[,index2];
      subx=px||cx;
      sweepx=(subx#fvi)`*subx;
      x1tx2=(subx#fvi)`*x2;
      x1tg=(subx#fvi)`*g;
      x2tx1=x2`*(subx#fvi);
      x2tx2=(x2#fvi)`*x2;
      x2tg=(x2#fvi)`*g;
      gtx1=g`*(subx#fvi);
      gtx2=g`*(x2#fvi);
      gtg=g`*g;
      sweepm=(sweepx||x1tx2||x1tg)//(x2tx1||x2tx2||x2tg)//(gtx1||gtx2||gtg);
      swp=sweep(sweepm,1:l+1);
      ssr[i,]=swp[k+1,k+1];
      i=i+1;
     end;
   end;
   if full=0 then run stopcrit;
  end;
finish selesw;

start selesc;
  selecrit='4';
  flag=1;
  do while (flag=1);
   piest=repeat(0,nrow(data),k-1);
   score=repeat(-99999,k-1,1);
   loglik=repeat(-99999,k-1,1);
   index=loc(^in);
   x=px;
   scx=px;
   run logitest;
   sb=b;
   i=1;
   do while (i<=(k-1));
    ci=index[,i];
    cx=fx[,ci];
```

```
      subx=px||cx;
      run scorest;
      x=subx;
      run logitest;
      z=x*b;
      pi=1/(1+exp(-z));
      pi=choose(pi=0,0.0000001,pi);
      pi=choose(pi=1,0.9999999,pi);
      piest[,i]=pi;
      loglik[i,]=sum((y=1)#log(pi)+(y=0)#log(1-pi));
      i=i+1;
    end;
    run stopcrit;
  end;
finish selesc;

start seleph;
  selecrit='5';
  flag=1;
  do while (flag=1);
    piest=repeat(0,nrow(data),k-1);
    scoreph=repeat(-99999,k-1,1);
    loglik=repeat(-99999,k-1,1);
    index=loc(^in);
    x=px;
    scx=px;
    run logitest;
    sb=b;
    i=1;
    do while (i<=(k-1));
      ci=index[,i];
      cx=fx[,ci];
      subx=px||cx;
      run scoreph;
      x=subx;
      run logitest;
      z=x*b;
      pi=1/(1+exp(-z));
      pi=choose(pi=0,0.0000001,pi);
```

```
    pi=choose(pi=1,0.9999999,pi);
    piest[,i]=pi;
    loglik[i,]=sum((y=1)#log(pi)+(y=0)#log(1-pi));
    i=i+1;
   end;
   run stopcrit;
  end;
finish seleph;

start selelk;
 selecrit='6';
 flag=1;
 do while (flag=1);
  piest=repeat(0,nrow(data),k-1);
  scorelk=repeat(-99999,k-1,1);
  loglik=repeat(-99999,k-1,1);
  index=loc(^in);
  x=px;
  sex=px;
  run logitest;
  sb=b;
  i=1;
  do while (i<=(k-1));
   ci=index[,i];
   cx=fx[,ci];
   subx=px||cx;
   run scorelk;
   x=subx;
   run logitest;
   z=x*b;
   pi=1/(1+exp(-z));
   pi=choose(pi=0,0.0000001,pi);
   pi=choose(pi=1,0.9999999,pi);
   piest[,i]=pi;
   loglik[i,]=sum((y=1)#log(pi)+(y=0)#log(1-pi));
   i=i+1;
  end;
  run stopcrit;
 end;
```

```
finish selelk;

start selewd;
 selecrit='7';
 flag=1;
 do while (flag=1);
  piest=repeat(.,nrow(data),k-l);
  piest=repeat(0,nrow(data),k-l);
  wald=repeat(-99999,k-l,1);
  index=loc(^in);
  i=1;
  do while (i<=(k-l));
   ci=index[,i];
   cx=fx[,ci];
   subx=px||cx;
   x=subx;
   run logitest;
   z=x*b;
   pi=1/(1+exp(-z));
   pi=choose(pi=0,0.0000001,pi);
   pi=choose(pi=1,0.9999999,pi);
   piest[,i]=pi;
   wald[i,]=(b[l+1,])`*(inv(xpxi[l+1,l+1]))*(b[l+1,]);
   i=i+1;
  end;
  run stopcrit;
 end;
finish selewd;

start logitest;
 b=repeat(0,ncol(x),1);
 oldb=b+1;
 do iter=1 to 10 while (max(abs(b-oldb))>1e-2);
  oldb=b;
  z=x*b;
  pi=1/(1+exp(-z));
  pi=choose(pi=0,0.0000001,pi);
  pi=choose(pi=1,0.9999999,pi);
  piqix=pi#(1-pi)#x;
```

```
    pvi=pi#(1-pi);
    w=1/(pi#(1-pi));
    xpxi=inv(piqix`*(w#piqix));
    b=b+xpxi*(piqix`*(w#(y-pi)));
  end;
finish logitest;

start scorest;
  z=scx*sb;
  pi=1/(1+exp(-z));
  pi=choose(pi=0,0.0000001,pi);
  pi=choose(pi=1,0.9999999,pi);
  piqisubx=pi#(1-pi)#subx;
  u=subx`*(y-pi);
  invi=inv(subx`*piqisubx);
  score[i,]=u`*invi*u;
finish scorest;

start scoreph;
  z=scx*sb;
  pi=1/(1+exp(-z));
  pi=choose(pi=0,0.0000001,pi);
  pi=choose(pi=1,0.9999999,pi);
  piqisubx=pi#(1-pi)#subx;
  u=cx`*(y-pi);
  invi=inv(subx`*piqisubx);
  cc=ncol(subx);
  invicc=invi[cc,cc];
  scoreph[i,]=u*invicc*u;
finish scoreph;

start scorelk;
  z=scx*sb;
  pi=1/(1+exp(-z));
  pi=choose(pi=0,0.0000001,pi);
  pi=choose(pi=1,0.9999999,pi);
  piqisubx=pi#(1-pi)#subx;
  u=subx`*(y-pi);
  invi=inv(subx`*piqisubx);
```

```
 invid=vecdiag(invi`
 scorelk[i,]=(u#invid)`*u;
finish scorelk;

start stopcrit;
 if selecrit='1' then do;
  imax=loglik[<:>,];
  log_lr=(2#abs((loglik[imax,]-loglik0)));
  sprob=1-probchi(log_lr,1);
  loglik0=loglik[imax,];
  ii=index[,imax];
  pihat=piest[,imax];
 end;

 if selecrit='2' then do;
  imax=loglik[<:>,];
  ls_chisq=(2#abs((loglik[imax,]-loglikls[imax,])));
  sprob=1-probchi(ls_chisq,1);
  ii=index[,imax];
  pihat=piest[,imax];
 end;

 if selecrit='3' then do;
  if t=0 then do;
   pihat=fpi;
   ii=index;
   full=1;
   flag=0;
   loglik1=sum((y=1)#log(fpi)+(y=0)#log(1-fpi));
   log_lr=(2#abs((loglik1-loglik0)));
   sprob=1-probchi(log_lr,1);
  end;
  else do;
   imin=ssr[>:<,];
   ssr1=ssr[imin,];
   ii=index[,imin];
   x=px||fx[,ii];
   run logitest;
   pihat=pi;
```

```
    loglik1=sum((y=1)#log(pi)+(y=0)#log(1-pi));
    log_lr=(2#abs((loglik1-loglik0)));
    sprob=1-probchi(log_lr,1);
    loglik0=loglik1;
  end;
end;

if selecrit='4' then do;
  imax=score[<:>,];
  score_ch=score[imax,];
  sprob=1-probchi(score_ch,1);
  ii=index[,imax];
  pihat=piest[,imax];
end;

if selecrit='5' then do;
  imax=scoreph[<:>,];
  score_ch=scoreph[imax,];
  sprob=1-probchi(score_ch,1);
  ii=index[,imax];
  pihat=piest[,imax];
end;

if selecrit='6' then do;
  imax=scorelk[<:>,];
  score_ch=scorelk[imax,];
  sprob=1-probchi(score_ch,1);
  ii=index[,imax];
  pihat=piest[,imax];
end;

if selecrit='7' then do;
  imax=wald[<:>,];
  wald_chi=wald[imax,];
  sprob=1-probchi(wald_chi,1);
  ii=index[,imax];
  pihat=piest[,imax];
end;
```

```
in[ii,]=^in[ii,];
l=sum(in);
if signal05=0 then goto skip10;
if (sprob>0.05 | l=k) then do;
 stopcrit='05'; run arralpha; run erralpha; signal05=0; goto skippx;
end;
else do;
 pialpha=pihat; subalpha=px||fx[,ii]; q=l-1; goto skippx;
end;
skip10:if signal10=0 then goto skip15;
if (sprob>0.10 | l=k) then do;
 stopcrit='10'; run arralpha; run erralpha; signal10=0; goto skippx;
end;
else do;
 pialpha=pihat; subalpha=px||fx[,ii]; q=l-1; goto skippx;
end;
skip15:if signal15=0 then goto skip20;
if (sprob>0.15 | l=k) then do;
 stopcrit='15'; run arralpha; run erralpha; signal15=0; goto skippx;
end;
else do;
 pialpha=pihat; subalpha=px||fx[,ii]; q=:   , goto skippx;
end;
skip20:if signal20=0 then goto skip25;
if (sprob>0.20 | l=k) then do;
 stopcrit='20'; run arralpha; run erralpha; signal20=0; goto skippx;
end;
else do;
 pialpha=pihat; subalpha=px||fx[,ii]; q=l-1; goto skippx;
end;
skip25:if signal25=0 then goto skip30;
if (sprob>0.25 | l=k) then do;
 stopcrit='25'; run arralpha; run erralpha; signal25=0; goto skippx;
end;
else do;
 pialpha=pihat; subalpha=px||fx[,ii]; q=l-1; goto skippx;
end;
skip30:if signal30=0 then goto skip35;
if (sprob>0.30 | l=k) then do;
```

```
  stopcrit='30', run arralpha; run erralpha; signal30=0; goto skippx;
end;
else do;
  pialpha=pihat; subalpha=px||fx[,ii]; q=l-1; goto skippx;
end;
skip35:if signal35=0 then goto skip40;
if (sprob>0.35 | l=k) then do;
  stopcrit='35'; run arralpha; run erralpha; signal35=0; goto skippx;
end;
else do;
  pialpha=pihat; subalpha=px||fx[,ii]; q=l-1; goto skippx;
end;
skip40:if signal40=0 then goto skip45;
if (sprob>0.40 | l=k) then do;
  stopcrit='40'; run arralpha; run erralpha; signal40=0; goto skippx;
end;
else do;
  pialpha=pihat; subalpha=px||fx[,ii]; q=l-1; goto skippx;
end;
skip45:if signal45=0 then goto skip50;
if (sprob>0.45 | l=k) then do;
  stopcrit='45'; run arralpha; run erralpha; signal45=0; goto skippx;
end;
else do;
  pialpha=pihat; subalpha=px||fx[,ii]; q=l-1; goto skippx;
end;
skip50:if signal50=0 then goto skip55;
if (sprob>0.50 | l=k) then do;
  stopcrit='50'; run arralpha; run erralpha; signal50=0; goto skippx;
end;
else do;
  pialpha=pihat; subalpha=px||fx[,ii]; q=l-1; goto skippx;
end;
skip55:if signal55=0 then goto skip60;
if (sprob>0.55 | l=k) then do;
  stopcrit='55'; run arralpha; run erralpha; signal55=0; goto skippx;
end;
else do;
  pialpha=pihat; subalpha=px||fx[,ii]; q=l-1; goto skippx;
```

```
end;
skip60:if signal60=0 then goto skip65;
if (sprob>0.60 | l=k) then do;
 stopcrit='60'; run arralpha; run erralpha; signal60=0; goto skippx;
end;
else do;
 pialpha=pihat; subalpha=px||fx[,ii]; q=l-1; goto skippx;
end;
skip65:if signal65=0 then goto skip70;
if (sprob>0.65 | l=k) then do;
 stopcrit='65'; run arralpha; run erralpha; signal65=0; goto skippx;
end;
else do;
 pialpha=pihat; subalpha=px||fx[,ii]; q=l-1; goto skippx;
end;
skip70:if signal70=0 then goto skip75;
if (sprob>0.70 | l=k) then do;
 stopcrit='70'; run arralpha; run erralpha; signal70=0; goto skippx;
end;
else do;
 pialpha=pihat; subalpha=px||fx[,ii]; q=l-1; goto skippx;
end;
skip75:if signal75=0 then goto skip80;
if (sprob>0.75 | l=k) then do;
 stopcrit='75'; run arralpha; run erralpha; signal75=0; goto skippx;
end;
else do;
 pialpha=pihat; subalpha=px||fx[,ii]; q=l-1; goto skippx;
end;
skip80:if signal80=0 then goto skip85;
if (sprob>0.80 | l=k) then do;
 stopcrit='80'; run arralpha; run erralpha; signal80=0; goto skippx;
end;
else do;
 pialpha=pihat; subalpha=px||fx[,ii]; q=l-1; goto skippx;
end;
skip85:if signal85=0 then goto skip90;
if (sprob>0.85 | l=k) then do;
 stopcrit='85'; run arralpha; run erralpha; signal85=0; goto skippx;
```

```
  end;
  else do;
   pialpha=pihat; subalpha=px||fx[,ii]; q=l-1; goto skippx;
  end;
  skip90:if signal90=0 then goto skip95;
  if (sprob>0.90 | l=k) then do;
   stopcrit='90'; run arralpha; run erralpha; signal90=0; goto skippx;
  end;
  else do;
   pialpha=pihat; subalpha=px||fx[,ii]; q=l-1; goto skippx;
  end;
  skip95:if (sprob>0.95 | l=k) then do;
   stopcrit='95'; run arralpha; run erralpha;
  end;
  else do;
   pialpha=pihat; subalpha=px||fx[,ii]; q=l-1; goto skippx;
  end;
  skippx:px=px||fx[,ii];
finish stopcrit;

start arralpha;
  if l=k then do; flag=0; pialpha=pihat; end;
  y1=repeat(1,nrow(data),1);
  yhat=(pialpha>=0.5)#y1;
  yy=y+yhat;
  marr=choose(yy=2,0,yy);
  arralpha=sum(marr)/nrow(data);
  perarr=arralpha;
finish arralpha;

start erralpha;
  if l=k then do;
   flag=0;
   cihat=-fx*fb;
   sqrtdi=sqrt(vecdiag(fx*fxpxi*fx'));
   z=1/sqrt(2*3.1415927)*exp(-1/2*(cihat/sqrtdi)#(cihat/sqrtdi));
   biasalph=2/(2*n)*sum(fvi#z#sqrtdi);
   erralpha=arralpha+biasalph;
   pere.: =erralpha;
```

```
if stopcrit='05' then do;
 stopcrit='05'; link saveout; stopcrit='10'; link saveout;
 stopcrit='15'; link saveout; stopcrit='20'; link saveout;
 stopcrit='25'; link saveout; stopcrit='30'; link saveout;
 stopcrit='35'; link saveout; stopcrit='40'; link saveout;
 stopcrit='45'; link saveout; stopcrit='50'; link saveout;
 stopcrit='55'; link saveout; stopcrit='60'; link saveout;
 stopcrit='65'; link saveout; stopcrit='70'; link saveout;
 stopcrit='75'; link saveout; stopcrit='80'; link saveout;
 stopcrit='85'; link saveout; stopcrit='90'; link saveout;
 stopcrit='95'; link saveout; goto done; end;
if stopcrit='10' then do;
 stopcrit='10'; link saveout; stopcrit='15'; link saveout;
 stopcrit='20'; link saveout; stopcrit='25'; link saveout;
 stopcrit='30'; link saveout; stopcrit='35'; link saveout;
 stopcrit='40'; link saveout; stopcrit='45'; link saveout;
 stopcrit='50'; link saveout; stopcrit='55'; link saveout;
 stopcrit='60'; link saveout; stopcrit='65'; link saveout;
 stopcrit='70'; link saveout; stopcrit='75'; link saveout;
 stopcrit='80'; link saveout; stopcrit='85'; link saveout;
 stopcrit='90'; link saveout; stopcrit='95'; link saveout;
 goto done; end;
if stopcrit='15' then do;
 stopcrit='15'; link saveout; stopcrit='20'; link saveout;
 stopcrit='25'; link saveout; stopcrit='30'; link saveout;
 stopcrit='35'; link saveout; stopcrit='40'; link saveout;
 stopcrit='45'; link saveout; stopcrit='50'; link saveout;
 stopcrit='55'; link saveout; stopcrit='60'; link saveout;
 stopcrit='65'; link saveout; stopcrit='70'; link saveout;
 stopcrit='75'; link saveout; stopcrit='80'; link saveout;
 stopcrit='85'; link saveout; stopcrit='90'; link saveout;
 stopcrit='95'; link saveout; goto done; end;
if stopcrit='20' then do;
 stopcrit='20'; link saveout; stopcrit='25'; link saveout;
 stopcrit='30'; link saveout; stopcrit='35'; link saveout;
 stopcrit='40'; link saveout; stopcrit='45'; link saveout;
 stopcrit='50'; link saveout; stopcrit='55'; link saveout;
 stopcrit='60'; link saveout; stopcrit='65'; link saveout;
 stopcrit='70'; link saveout; stopcrit='75'; link saveout;
```

```
  stopcrit='80'; link saveout; stopcrit='85'; link saveout;
  stopcrit='90'; link saveout; stopcrit='95'; link saveout;
  goto done; end;
if stopcrit='25' then do;
  stopcrit='25'; link saveout; stopcrit='30'; link saveout;
  stopcrit='35'; link saveout; stopcrit='40'; link saveout;
  stopcrit='45'; link saveout; stopcrit='50'; link saveout;
  stopcrit='55'; link saveout; stopcrit='60'; link saveout;
  stopcrit='65'; link saveout; stopcrit='70'; link saveout;
  stopcrit='75'; link saveout; stopcrit='80'; link saveout;
  stopcrit='85'; link saveout; stopcrit='90'; link saveout;
  stopcrit='95'; link saveout; goto done; end;
if stopcrit='30' then do;
  stopcrit='30'; link saveout; stopcrit='35'; link saveout;
  stopcrit='40'; link saveout; stopcrit='45'; link saveout;
  stopcrit='50'; link saveout; stopcrit='55'; link saveout;
  stopcrit='60'; link saveout; stopcrit='65'; link saveout;
  stopcrit='70'; link saveout; stopcrit='75'; link saveout;
  stopcrit='80'; link saveout; stopcrit='85'; link saveout;
  stopcrit='90'; link saveout; stopcrit='95'; link saveout;
  goto done; end;
if stopcrit='35' then do;
  stopcrit='35'; link saveout; stopcrit='40'; link saveout;
  stopcrit='45'; link saveout; stopcrit='50'; link saveout;
  stopcrit='55'; link saveout; stopcrit='60'; link saveout;
  stopcrit='65'; link saveout; stopcrit='70'; link saveout;
  stopcrit='75'; link saveout; stopcrit='80'; link saveout;
  stopcrit='85'; link saveout; stopcrit='90'; link saveout;
  stopcrit='95'; link saveout; goto done; end;
if stopcrit='40' then do;
  stopcrit='40'; link saveout; stopcrit='45'; link saveout;
  stopcrit='50'; link saveout; stopcrit='55'; link saveout;
  stopcrit='60'; link saveout; stopcrit='65'; link saveout;
  stopcrit='70'; link saveout; stopcrit='75'; link saveout;
  stopcrit='80'; link saveout; stopcrit='85'; link saveout;
  stopcrit='90'; link saveout; stopcrit='95'; link saveout;
  goto done; end;
if stopcrit='45' then do;
  stopcrit='45'; link saveout; stopcrit='50'; link saveout;
```

```
  stopcrit='55'; link saveout; stopcrit='60'; link saveout;
  stopcrit='65'; link saveout; stopcrit='70'; link saveout;
  stopcrit='75'; link saveout; stopcrit='80'; link saveout;
  stopcrit='85'; link saveout; stopcrit='90'; link saveout;
  stopcrit='95'; link saveout; goto done; end;
if stopcrit='50' then do;
  stopcrit='50'; link saveout; stopcrit='55'; link saveout;
  stopcrit='60'; link saveout; stopcrit='65'; link saveout;
  stopcrit='70'; link saveout; stopcrit='75'; link saveout;
  stopcrit='80'; link saveout; stopcrit='85'; link saveout;
  stopcrit='90'; link saveout; stopcrit='95'; link saveout;
  goto done; end;
if stopcrit='55' then do;
  stopcrit='55'; link saveout; stopcrit='60'; link saveout;
  stopcrit='65'; link saveout; stopcrit='70'; link saveout;
  stopcrit='75'; link saveout; stopcrit='80'; link saveout;
  stopcrit='85'; link saveout; stopcrit='90'; link saveout;
  stopcrit='95'; link saveout; goto done; end;
if stopcrit='60' then do;
  stopcrit='60'; link saveout; stopcrit='65'; link saveout;
  stopcrit='70'; link saveout; stopcrit='75'; link saveout;
  stopcrit='80'; link saveout; stopcrit='85'; link saveout;
  stopcrit='90'; link saveout; stopcrit='95'; link saveout;
  goto done; end;
if stopcrit='65' then do;
  stopcrit='65'; link saveout;
  stopcrit='70'; link saveout; stopcrit='75'; link saveout;
  stopcrit='80'; link saveout; stopcrit='85'; link saveout;
  stopcrit='90'; link saveout; stopcrit='95'; link saveout;
  goto done; end;
if stopcrit='70' then do;
  stopcrit='70'; link saveout; stopcrit='75'; link saveout;
  stopcrit='80'; link saveout; stopcrit='85'; link saveout;
  stopcrit='90'; link saveout; stopcrit='95'; link saveout;
  goto done; end;
if stopcrit='75' then do;
  stopcrit='75'; link saveout; stopcrit='80'; link saveout;
  stopcrit='85'; link saveout; stopcrit='90'; link saveout;
  stopcrit='95'; link saveout; goto done; end;
```

```
  if stopcrit='80' then do;
    stopcrit='80'; link saveout; stopcrit='85'; link saveout;
    stopcrit='90'; link saveout; stopcrit='95'; link saveout;
    goto done; end;
  if stopcrit='85' then do;
    stopcrit='85'; link saveout; stopcrit='90'; link saveout;
    stopcrit='95'; link saveout; goto done; end;
  if stopcrit='90' then do;
    stopcrit='90'; link saveout; stopcrit='95'; link saveout;
    goto done; end;
  if stopcrit='95' then do;
    stopcrit='95'; link saveout; goto done; end;
  saveout:
  edit lee.cp1 var {run rep selecrit stopcrit perarr pererr};
  append;
  close lee.cp1;
  return;
end;
else do;
  vi0=pialpha#(1-pialpha);
  cihat=-fx*fb;
  pxpxi=inv((subalpha#vi0)`*subalpha);
  xpxi0=inv((subalpha#fvi)`*subalpha);
  ixpxi0=inv(xpxi0);
  temp1=subalpha*pxpxi;
  temp2=pxpxi*subalpha`;
  temp3=temp1*ixpxi0;
  temp4=temp3*temp2;
  free temp3;
  sqrtcdi=sqrt(vecdiag(temp4));
  free temp4;
  di0=vecdiag(temp1*subalpha`);
  z=1/sqrt(2*3.1415927)*exp(-1/2*(cihat/sqrtcdi)#(cihat/sqrtcdi));
  biasalph=2/(2*n)*sum(fvi#z#(di0/sqrtcdi));
  erralpha=arralpha+biasalph;
  pererr=erralpha;
  if firstset=1 then do;
    create lee.cp1 var {run rep selecrit stopcrit perarr pererr;
    append;
```

```
    close lee.cp1;
    firstset=0;
  end;
  else do;
    edit lee.cp1 var {run rep selecrit stopcrit perarr pererr};
    append;
    close lee.cp1;
  end;
  free temp1 temp2 vi0 cihat pxpxi xpxi0 ixpxi0 sqrtedi di0 z;
 end;
 done:
finish erralpha;

run runrep;
quit;
endsas;
```

APPENDIX D

SAS/IML program for sections 9.2, 9.4, and 9.5

```
libname lee '/usr/accts/david/simulate/';

start selecrit;
 run initss;
 run lintcept;
 run fullx;
 run selelr;

 run initsele;
 run lintcept;
 run selels;

 run initsw;
 run lintcept;
 run selesw;

 run initsele;
 run lintcept;
 run selesc;

 run initsele;
 run lintcept;
 run seleph;

 run initsele;
 run lintcept;
 run selelk;

 run initsele;
 run lintcept;
 run selewd;

 run cmord;
finish selecrit;
```

```
start initsele;
 run=r;
 rep=s;
 y=data[,1];
 k=ncol(data);
 cindx=2:k;
 x=repeat(1,nrow(data),1)||(data[,cindx]);
 in=repeat(0,k,1);
 order=repeat(0,k,1);
 arr=repeat(9,k,1);
 bias=repeat(9,k,1);
 err=repeat(9,k,1);
 cp=repeat(9,k,1);
 aic=repeat(9,k,1);
 schwartz=repeat(9,k,1);
 l=1;
 fx=x;
 px=x[,1];
 in[1,]=1;
 signal=1;
finish initsele;

start initsw;
 run=r;
 rep=s;
 y=data[,1];
 k=ncol(data);
 cindx=2:k;
 x=repeat(1,nrow(data),1)||(data[,cindx]);
 in=repeat(0,k,1);
 order=repeat(0,k,1);
 arr=repeat(9,k,1);
 bias=repeat(9,k,1);
 err=repeat(9,k,1);
 cp=repeat(9,k,1);
 aic=repeat(9,k,1);
 schwartz=repeat(9,k,1);
 l=1;
 fx=x;
```

```
 run fullx;
 fpi=choose(fpi=0,0.0000001,fpi);
 g=log(fpi/(1-fpi));
 x2=fx[,2:k];
 subx=x[,1];
 px=subx;
 sweepx=(subx#fvi)`*subx;
 x1tx2=(subx#fvi)`*x2;
 x1tg=(subx#fvi)`*g;
 x2tx1=x2`*(subx#fvi);
 x2tx2=(x2#fvi)`*x2;
 x2tg=(x2#fvi)`*g;
 gtx1=g`*(subx#fvi);
 gtx2=g`*(x2#fvi);
 gtg=g`*g;
 sweepm=(sweepx||x1tx2||x1tg)//(x2tx1||x2tx2||x2tg)//(gtx1||gtx2||gtg);
 swp=sweep(sweepm,1);
 ssr0=swp[k+1,k+1];
 in[1,]=1;
 signal=1;
finish initsw;

start lintcept;
 pi0=sum((y=1)#y)/nrow(fx);
 lintcept=sum((y=1)#log(pi0)+(y=0)#log(1-pi0));
 loglik0=lintcept;
 arr[1,]=.;
 bias[1,]=.;
 err[1,]=.;
 cp[1,]=.;
 aic[1,]=.;
 schwartz[1,]=.;
finish lintcept;

start fullx;
 fb=repeat(0,ncol(fx),1);
 foldb=fb+1;
 do iter=1 to 10 while (max(abs(fb-foldb))>1e-2);
  foldb=fb;
```

```
  fz=fx*fb;
  fpi=1/(1+exp(-fz));
  fpi=choose(fpi=0,0.0000001,fpi);
  fpi=choose(fpi=1,0.9999999,fpi);
  fvi=fpi#(1-fpi);
  fpiqix=fvi#fx;
  fw=1/fvi;
  fxpxi=inv(fpiqix`*(fw#fpiqix));
  fb=fb+fxpxi*(fpiqix`*(fw#(y-fpi)));
 end;
finish fullx;

start selelr;
 selecrit='1';
 flag=1;
 do while (flag=1);
  piest=repeat(0,nrow(data),k-1);
  loglik=repeat(-99999,k-1,1);
  wald=repeat(-99999,k-1,1);
  index=loc(^in);
  i=1;
  do while (i<=(k-1));
   ci=index[,i];
   cx=fx[,ci];
   subx=px||cx;
   x=subx;
   run logitest;
   z=x*b;
   pi=1/(1+exp(-z));
   pi=choose(pi=0,0.0000001,pi);
   pi=choose(pi=1,0.9999999,pi);
   piest[,i]=pi;
   loglik[i,]=sum((y=1)#log(pi)+(y=0)#log(1-pi));
   wald[i,]=(b[l+1,])`*(inv(xpxi[l+1,l+1]))*(b[l+1,]);
   i=i+1;
  end;
  run stopcrit;
 end;
finish selelr;
```

```
start selels;
 selecrit='2';
 flag=1;
 do while (flag=1);
  piest=repeat(0,nrow(data),k-1);
  loglik=repeat(-99999,k-1,1);
  loglikls=repeat(-99999,k-1,1);
  wald=repeat(-99999,k-1,1);
  index=loc(^in);
  i=1;
  do while (i<=(k-1));
   ci=index[,i];
   cx=fx[,ci];
   subx=px||cx;
   x=subx;
   run logitest;
   z=x*b;
   pi=1/(1+exp(-z));
   pi=choose(pi=0,0.0000001,pi);
   pi=choose(pi=1,0.9999999,pi);
   piest[,i]=pi;
   loglik[i,]=sum((y=1)#log(pi)+(y=0)#log(1-pi));
   wald[i,]=(b[l+1,])`*(inv(xpxi[l+1,l+1]))*(b[l+1,]);
   bls=b[1:l,]-xpxi[1:l,l+1]*(inv(xpxi[l+1,l+1]))*b[l+1,];
   zls=px*bls;
   pils=1/(1+exp(-zls));
   pils=choose(pils=0,0.0000001,pils);
   pils=choose(pils=1,0.9999999,pils);
   loglikls[i,]=sum((y=1)#log(pils)+(y=0)#log(1-pils));
   i=i+1;
  end;
  run stopcrit;
 end;
finish selels;

start selesw;
 selecrit='3';
 flag=1;
 full=0;
```

```
do while (flag=1);
 ssr=repeat(-99999,k-l,1);
 index=loc(^in);
 i=1;
 do while (i<=(k-l) & full=0);
  ci=index[,i];
  cx=fx[,ci];
  t=index#(index^=ci);
  if t=0 then run xvxgfull;
  else do;
   nzindx=loc(t>0);
   index2=index[,nzindx];
   x2=fx[,index2];
   subx=px||cx;
   sweepx=(subx#fvi)`*subx;
   x1tx2=(subx#fvi)`*x2;
   x1tg=(subx#fvi)`*g;
   x2tx1=x2`*(subx#fvi);
   x2tx2=(x2#fvi)`*x2;
   x2tg=(x2#fvi)`*g;
   gtx1=g`*(subx#fvi);
   gtx2=g`*(x2#fvi);
   gtg=g`*g;
   sweepm=(sweepx||x1tx2||x1tg)//(x2tx1||x2tx2||x2tg)//(gtx1||gtx2||gtg);
   swp=sweep(sweepm,1:l+1);
   ssr[i,]=swp[k+1,k+1];
   i=i+1;
  end;
 end;
 if full=0 then run stopcrit;
end;
finish selesw;

start selesc;
 selecrit='4';
 flag=1;
 do while (flag=1);
  piest=repeat(0,nrow(data),k-l);
  score=repeat(-99999,k-l,1);
```

```
  loglik=repeat(-99999,k-1,1);
  wald=repeat(-99999,k-1,1);
  index=loc(^in);
  x=px;
  scx=px;
  run logitest;
  sb=b;
  i=1;
  do while (i<=(k-1));
   ci=index[,i];
   cx=fx[,ci];
   subx=px||cx;
   run scorest;
   x=subx;
   run logitest;
   z=x*b;
   pi=1/(1+exp(-z));
   pi=choose(pi=0,0.0000001,pi);
   pi=choose(pi=1,0.9999999,pi);
   piest[,i]=pi;
   loglik[i,]=sum((y=1)#log(pi)+(y=0)#log(1-pi));
   wald[i,]=(b[l+1,])`*(inv(xpxi[l+1,l+1]))*(b[l+1,]);
   i=i+1;
  end;
  run stopcrit;
 end;
finish selesc;

start seleph;
 selcrit='5';
 flag=1;
 do while (flag=1);
  piest=repeat(0,nrow(data),k-1);
  scoreph=repeat(-99999,k-1,1);
  loglik=repeat(-99999,k-1,1);
  wald=repeat(-99999,k-1,1);
  index=loc(^in);
  x=px;
  scx=px;
```

```
 run logitest;
 sb=b;
 i=1;
 do while (i<=(k-1));
  ci=index[,i];
  cx=fx[,ci];
  subx=px||cx;
  run scoreph;
  x=subx;
  run logitest;
  z=x*b;
  pi=1/(1+exp(-z));
  pi=choose(pi=0,0.0000001,pi);
  pi=choose(pi=1,0.9999999,pi);
  piest[,i]=pi;
  loglik[i,]=sum((y=1)#log(pi)+(y=0)#log(1-pi));
  wald[i,]=(b[l+1,])`*(inv(xpxi[l+1,l+1]))*(b[l+1,]);
  i=i+1;
 end;
 run stopcrit;
end;
finish seleph;

start selelk;
 selecrit='6';
 flag=1;
 do while (flag=1);
  piest=repeat(0,nrow(data),k-l);
  scorelk=repeat(-99999,k-l,1);
  loglik=repeat(-99999,k-l,1);
  wald=repeat(-99999,k-l,1);
  index=loc(^in);
  x=px;
  scx=px;
  run logitest;
  sb=b;
  i=1;
  do while (i<=(k-l));
   ci=index[,i];
```

```
   cx=fx[,ci];
   subx=px||cx;
   run scorelk;
   x=subx;
   run logitest;
   z=x*b;
   pi=1/(1+exp(-z));
   pi=choose(pi=0,0.0000001,pi);
   pi=choose(pi=1,0.9999999,pi);
   piest[,i]=pi;
   loglik[i,]=sum((y=1)#log(pi)+(y=0)#log(1-pi));
   wald[i,]=(b[l+1,])`*(inv(xpxi[l+1,l+1]))*(b[l+1,]);
   i=i+1;
  end;
  run stopcrit;
 end;
finish selelk;

start selewd;
 selecrit='7';
 flag=1;
 do while (flag=1);
  piest=repeat(0,nrow(data),k-l);
  loglik=repeat(-99999,k-l,1);
  wald=repeat(-99999,k-l,1);
  index=loc(^in);
  i=1;
  do while (i<=(k-l));
   ci=index[,i];
   cx=fx[,ci];
   subx=px||cx;
   x=subx;
   run logitest;
   z=x*b;
   pi=1/(1+exp(-z));
   pi=choose(pi=0,0.0000001,pi);
   pi=choose(pi=1,0.9999999,pi);
   piest[,i]=pi;
   loglik[i,]=sum((y=1)#log(pi)+(y=0)#log(1-pi));
```

```
   wald[i,]=(b[l+1,])`*(inv(xpxi[l+1,l+1]))*(b[l+1,]);
   i=i+1;
  end;
  run stopcrit;
 end;
finish selewd;

start logitest;
 b=repeat(0,ncol(x),1);
 oldb=b+1;
 do iter=1 to 10 while (max(abs(b-oldb))>1e-2);
  oldb=b;
  z=x*b;
  pi=1/(1+exp(-z));
  pi=choose(pi=0,0.0000001,pi);
  pi=choose(pi=1,0.9999999,pi);
  piqix=pi#(1-pi)#x;
  pvi=pi#(1-pi);
  w=1/(pi#(1-pi));
  xpxi=inv(piqix`*(w#piqix));
  b=b+xpxi*(piqix`*(w#(y-pi)));
 end;
finish logitest;

start xvxgfull;
 x=fx;
 run logitest;
 pihat=pi;
 loglik1=sum((y=1)#log(pi)+(y=0)#log(1-pi));
 wald0=(b[l+1,])`*(inv(xpxi[l+1,l+1]))*(b[l+1,]);
 ii=index;
 order[index,]=k-1;
 if signal=1 then do;
  pialpha=pi;
  q=k-1;
  run arralpha;
  l=k;
  run erralpha;
 end;
```

```
 run arr;
 run err;
 run cp;
 aic[ii,]=-2*loglik1+2*(l+1);
 schwartz[ii,]=-2*loglik1+2*(l+1)*log(2*n);
 run errmin;
 run cpmin;
 run aicmin;
 run schmin;
 order3=order;
 full=1;
finish xvxgfull;

start scorest;
 z=scx*sb;
 pi=1/(1+exp(-z));
 pi=choose(pi=0,0.0000001,pi);
 pi=choose(pi=1,0.9999999,pi);
 piqisubx=pi#(1-pi)#subx;
 u=subx`*(y-pi);
 invi=inv(subx`*piqisubx);
 score[i,]=u`*invi*u;
finish scorest;

start scoreph;
 z=scx*sb;
 pi=1/(1+exp(-z));
 pi=choose(pi=0,0.0000001,pi);
 pi=choose(pi=1,0.9999999,pi);
 piqisubx=pi#(1-pi)#subx;
 u=cx`*(y-pi);
 invi=inv(subx`*piqisubx);
 cc=ncol(subx);
 invicc=invi[cc,cc];
 scoreph[i,]=u*invicc*u;
finish scoreph;

start scorelk;
 z=scx*sb;
```

```
 pi=1/(1+exp(-z));
 pi=choose(pi=0,0.0000001,pi);
 pi=choose(pi=1,0.9999999,pi);
 piqisubx=pi#(1-pi)#subx;
 u=subx`*(y-pi);
 invi=inv(subx`*piqisubx);
 invid=vecdiag(invi);
 scorelk[i,]=(u#invid)`*u;
finish scorelk;

start stopcrit;
 if selecrit='1' then do;
  imax=loglik[<:>,];
  log_lr=(2#abs((loglik[imax,]-loglik0)));
  sprob=1-probchi(log_lr,1);
  loglik0=loglik[imax,];
  wald0=wald[imax,];
  ii=index[,imax];
  pihat=piest[,imax];
 end;

 if selecrit='2' then do;
  imax=loglik[<:>,];
  ls_chisq=(2#abs((loglik[imax,]-loglikls[imax,])));
  sprob=1-probchi(ls_chisq,1);
  wald0=wald[imax,];
  ii=index[,imax];
  pihat=piest[,imax];
 end;

 if selecrit='3' then do;
  imin=ssr[>:<,];
  ssr1=ssr[imin,];
  ii=index[,imin];
  x=px||fx[,ii];
  run logitest;
  pihat=pi;
  loglik1=sum((y=1)#log(pi)+(y=0)#log(1-pi));
  wald0=(b[l+1,])`*(inv(xpxi[l+1,l+1]))*(b[l+1,]);
```

```
 log_lr=(2#abs((loglik1-loglik0)));
 sprob=1-probchi(log_lr,1);
 loglik0=loglik1;
end;

if selecrit='4' then do;
 imax=score[<:>,];
 score_ch=score[imax,];
 sprob=1-probchi(score_ch,1);
 wald0=wald[imax,];
 ii=index[,imax];
 pihat=piest[,imax];
end;

if selecrit='5' then do;
 imax=scoreph[<:>,];
 score_ch=scoreph[imax,];
 sprob=1-probchi(score_ch,1);
 wald0=wald[imax,];
 ii=index[,imax];
 pihat=piest[,imax];
end;

if selecrit='6' then do;
 imax=scorelk[<:>,];
 score_ch=scorelk[imax,];
 sprob=1-probchi(score_ch,1);
 wald0=wald[imax,];
 ii=index[,imax];
 pihat=piest[,imax];
end;

if selecrit='7' then do;
 imax=wald[<:>,];
 wald_chi=wald[imax,];
 sprob=1-probchi(wald_chi,1);
 wald0=wald_chi;
 ii=index[,imax];
 pihat=piest[,imax];
```

```
end;

if (selecrit='1'|selecrit='2'|selecrit='4'|selecrit='5'|
   selecrit='6'|selecrit='7') then do;
 run arr;
 run err;
 run cp;
 run aic;
 run schwartz;
end;
else do;
 run arr;
 run err;
 run cp;
 aic[ii,]=-2*loglik1+2*(l+1);
 schwartz[ii,]=-2*loglik1+2*(l+1)*log(2*n);
end;
in[ii,]=^in[ii,];
order[ii,]=l;
l=sum(in);
if (sprob>0.20 & signal=1) then do;
 run arralpha;
 run erralpha;
 signal=0;
end;
else do;
 pialpha=pihat;
 subalpha=px||fx[,ii];
 q=l-1;
end;
if (l=k) & (selecrit='1'|selecrit='2'|selecrit='4'|selecrit='5'|
  selecrit='6'|selecrit='7') then do;
 if signal=1 then do;
  pialpha=fpi;
  q=k-1;
  run arralpha;
  run erralpha;
 end;
 if selecrit='1' then order1=order;
```

```
  if selecrit='2' then order2=order;
  if selecrit='4' then order4=order;
  if selecrit='5' then order5=order;
  if selecrit='6' then order6=order;
  if selecrit='7' then order7=order;
  run errmin;
  run cpmin;
  run aicmin;
  run schmin;
 end;
 else do;
  px=px||fx[,ii];
 end;
finish stopcrit;

start arr;
 y1=repeat(1,nrow(data),1);
 y1=repeat(1,nrow(data),1);
 yhat=(pihat>=0.5)#y1;
 yy=y+yhat;
 marr=choose(yy=2,0,yy);
 arr[ii,]=sum(marr)/nrow(data);
finish arr;

start err;
 if l=(k-1) then do;
  q=k-1;
  cihat=-fx*fb;
  sqrtdi=sqrt(vecdiag(fx*fxpxi*fx`));
  z=1/sqrt(2*3.1415927)*exp(-1/2*(cihat/sqrtdi)#(cihat/sqrtdi));
  bias[ii,]=2/(2*n)*sum(fvi#z#sqrtdi);
  err[ii,]=arr[ii,]+bias[ii,];
 end;
 else do;
  subx=px||fx[,ii];
  vi0=pihat#(1-pihat);
  cihat=-fx*fb;
  pxpxi=inv((subx#vi0)`*subx);
  xpxi0=inv((subx#fvi)`*subx);
```

```
 ixpxi()=inv(xpxi());
 spxpxi=subx*pxpxi;
 pxpxist=pxpxi*subx`;
 temp1=spxpxi*ixpxi();
 temp2=temp1*pxpxist;
 sqrtcdi=sqrt(vecdiag(temp2));
 temp3=spxpxi*subx`;
 di()=vecdiag(temp3);
 z=1/sqrt(2*3.1415927)*exp(-1/2*(cihat/sqrtcdi)#(cihat/sqrtcdi));
 bias[ii,]=2/(2*n)*sum(fvi#z#(di()/sqrtcdi));
 err[ii,]=arr[ii,]+bias[ii,];
 free temp3 vi() cihat pxpxi xpxi() ixpxi() spxpxi pxpxist sqrtcdi di() z;
end;
finish err;

start cp;
 ss=(y-fpi)#(y-fpi);
 gof_chis=sum(ss/fvi);
 cp[ii,]=(gof_chis+wald())*(2*n-p-1)/gof_chis+2*(l+1)-2*n;
finish cp;

start aic;
 aic[ii,]=-2*loglik[imax,]+2*(l+1);
finish aic;

start schwartz;
 schwartz[ii,]=-2*loglik[imax,]+2*(l+1)*log(2*n);
finish schwartz;

start arralpha;
 y1=repeat(1,nrow(data),1);
 yhat=(pialpha>=0.5)#y1;
 yy=y+yhat;
 marr=choose(yy=2,0,yy);
 arralpha=sum(marr)/nrow(data);
finish arralpha;

start erralpha;
 stopcrit='1';
```

```
if l=k then do;
 q=k-1,
 cihat=-fx*fb;
 sqrtdi=sqrt(vecdiag(fx*fxpxi*fx`));
 z=1/sqrt(2*3.1415927)*exp(-1/2*(cihat/sqrtdi)#(cihat/sqrtdi));
 biasalph=2/(2*n)*sum(fvi#z#sqrtdi);
 erralpha=arralpha+biasalph;
 pererr=erralpha;
 numq=q;
end;
else do;
 vi0=pialpha#(1-pialpha);
 cihat=-fx*fb;
 pxpxi=inv((subalpha#vi0)`*subalpha);
 xpxi0=inv((subalpha#fvi)`*subalpha);
 ixpxi0=inv(xpxi0);
 temp1=subalpha*pxpxi;
 temp2 -pxpxi*subalpha`;
 temp3=temp1*ixpxi0;
 temp4=temp3*temp2;
 sqrtcdi=sqrt(vecdiag(temp4));
 di0=vecdiag(temp1*subalpha`);
 z=1/sqrt(2*3.1415927)*exp(-1/2*(cinat/sqrtcdi)#(cihat/sqrtcdi));
 biasalph=2/(2*n)*sum(fvi#z#(di0/sqrtcdi));
 erralpha=arralpha+biasalph;
 pererr=erralpha;
 numq=q;
 free temp1 temp2 vi0 cihat pxpxi xpxi0 ixpxi0 sqrtcdi di0 z;
end;
if firstset=1 then do;
 create lee.cm var {run rep selecrit stopcrit pererr numq};
 append;
 close lee.cm;
 firstset=0;
end;
else do;
 edit lee.cm var {run rep selecrit stopcrit pererr numq};
 append;
 close lee.cm;
```

```
 end;
finish erralpha;

start errmin;
 stopcrit='2';
 errmindx=err[>:<,];
 biasem=bias[errmindx];
 errem=err[errmindx,];
 qem=order[errmindx,];
 pererr=errem;
 numq=qem;
 edit lee.cm var {run rep selecrit stopcrit pererr numq};
 append;
 close lee.cm;
finish errmin;

start cpmin;
 stopcrit='3';
 cpmin=min(cp);
 cpmindx=cp[>:<,];
 biascpm=bias[cpmindx,];
 errcpm=err[cpmindx,];
 qcpm=order[cpmindx,];
 pererr=errcpm;
 numq=qcpm;
 edit lee.cm var {run rep selecrit stopcrit pererr numq};
 append;
 close lee.cm;
finish cpmin;

start aicmin;
 stopcrit='4';
 aicmin=min(aic);
 aicmindx=aic[>:<,];
 biasaicm=bias[aicmindx,];
 erraicm=err[aicmindx,];
 qaicm=order[aicmindx,];
 pererr=erraicm;
 numq=qaicm;
```

```
 edit lee.cm var {run rep selecrit stopcrit pererr numq};
 append;
 close lee.cm;
finish aicmin;

start schmin;
 stopcrit='5';
 schmin=min(schwartz);
 schmindx=schwartz[>:<,];
 biasschm=bias[schmindx,];
 errschm=err[schmindx,];
 qschm=order[schmindx,];
 pererr=errschm;
 numq=qschm;
 edit lee.cm var {run rep selecrit stopcrit pererr numq};
 append;
 close lee.cm;
 flag=0;
finish schmin;

start cmord;
 if run=1 & rep=1 then do;
  create lee.cmord var {runv repv order1 order2 order3
                        order4 order5 order6 order7};
  append;
  close lee.cmord;
 end;
 else do;
  edit lee.cmord var {runv repv order1 order2 order3
                      order4 order5 order6 order7};
  append;
  close lee.cmord;
 end;
finish cmord;

run selecrit;
quit;
endsas;
```

APPENDIX E

Low Birth Weight Data

1	2	3	4	5	6	7	8	9	10
1	0	22	131	1	0	0	0	0	1
2	0	32	170	1	0	0	0	0	0
3	0	30	110	0	0	0	0	0	0
4	0	20	127	0	0	0	0	0	0
5	0	23	123	0	0	0	0	0	0
6	0	17	120	0	1	0	0	0	0
7	0	19	105	0	0	0	0	0	0
8	0	23	130	1	0	0	0	0	0
9	0	36	175	1	0	0	0	0	0
10	0	22	125	1	0	0	0	0	1
11	0	24	133	1	0	0	0	0	0
12	0	21	134	0	0	0	0	0	2
13	0	19	235	1	1	0	1	0	0
14	0	25	95	1	1	3	0	1	0
15	0	16	135	1	1	0	0	0	0
16	0	29	135	1	0	0	0	0	1
17	0	29	154	1	0	0	0	0	1
18	0	19	147	1	1	0	0	0	0
19	0	19	147	1	1	0	0	0	0
20	0	30	137	1	0	0	0	0	1
21	0	24	110	1	0	0	0	0	1
22	0	19	184	1	1	0	1	0	0
23	0	24	110	0	0	1	0	0	0
24	0	23	110	1	0	0	0	0	1
25	0	20	120	0	0	0	0	0	0
26	0	25	241	0	0	0	1	0	0
27	0	30	112	1	0	0	0	0	1
28	0	22	169	1	0	0	0	0	0
29	0	18	120	1	1	0	0	0	2
30	0	16	170	0	0	0	0	0	4
31	0	32	186	1	0	0	0	0	2
32	0	18	120	0	0	0	0	0	1
33	0	29	130	1	1	0	0	0	2
34	0	33	117	1	0	0	0	1	1
35	0	20	170	1	1	0	0	0	0
36	0	28	134	0	0	0	0	0	1
37	0	14	135	1	0	0	0	0	0
38	0	28	130	0	0	0	0	0	0
39	0	25	120	1	0	0	0	0	2

40	0	16	95	0	0	0	0	0	1
41	0	20	158	1	0	0	0	0	1
42	0	26	160	0	0	0	0	0	0
43	0	21	115	1	0	0	0	0	1
44	0	22	129	1	0	0	0	0	0
45	0	25	130	1	0	0	0	0	2
46	0	31	120	1	0	0	0	0	2
47	0	35	170	1	0	1	0	0	1
48	0	19	120	1	1	0	0	0	0
49	0	24	116	1	0	0	0	0	1
50	0	45	123	1	0	0	0	0	1
51	1	28	120	0	1	1	0	1	0
52	1	29	130	1	0	0	0	1	2
53	1	34	187	0	1	0	1	0	0
54	1	25	105	0	0	1	1	0	0
55	1	25	85	0	0	0	0	1	0
56	1	27	150	0	0	0	0	0	0
57	1	23	97	0	0	0	0	1	1
58	1	24	128	0	0	1	0	0	1
59	1	24	132	0	0	0	1	0	0
60	1	21	165	1	1	0	1	0	1
61	1	32	105	1	1	0	0	0	0
62	1	19	91	1	1	2	0	1	0
63	1	25	115	0	0	0	0	0	1
64	1	16	130	0	0	0	0	0	1
65	1	25	92	1	1	0	0	0	0
66	1	20	150	1	1	0	0	0	2
67	1	21	200	0	0	0	0	1	2
68	1	24	155	1	1	1	0	0	0
69	1	21	103	0	0	0	0	0	0
70	1	20	125	0	0	0	0	1	0
71	1	25	89	0	0	2	0	0	1
72	1	19	102	1	0	0	0	0	2
73	1	19	112	1	1	0	0	1	0
74	1	26	117	1	1	1	0	0	0
75	1	24	138	1	0	0	0	0	0
76	1	17	130	0	1	1	0	1	0
77	1	20	120	0	1	0	0	0	3
78	1	22	130	1	1	1	0	1	1
79	1	27	130	0	0	0	0	1	0
80	1	20	80	0	1	0	0	1	0
81	1	17	110	1	1	0	0	0	0
82	1	25	105	0	0	1	0	0	1
83	1	20	109	0	0	0	0	0	0
84	1	18	148	0	0	0	0	0	0

85	1	18	110	0	1	1	0	0	0
86	1	20	121	1	1	1	0	1	0
87	1	21	100	0	0	1	0	0	4
88	1	26	96	0	0	0	0	0	0
89	1	31	102	1	1	1	0	0	1
90	1	15	110	1	0	0	0	0	0
91	1	23	187	0	1	0	0	0	1
92	1	20	122	0	1	0	0	[illegible]	0
93	1	24	105	0	1	0	0	0	0
94	1	15	115	0	0	0	0	1	0
95	1	23	120	0	0	0	0	0	0
96	1	30	142	1	1	1	0	0	0
97	1	22	130	1	1	0	0	0	1
98	1	17	120	1	1	0	0	0	3
99	1	23	110	1	1	1	0	0	0
100	1	17	120	0	0	0	0	0	2

APPENDIX F

Intensive Care Unit Data

1	2	3	4	5	6	7	8	9	10	11	12
1	0	27	1	0	0	0	142	88	1	0	0
2	0	59	0	0	0	0	112	80	1	0	0
3	0	77	0	1	0	0	100	70	0	0	0
4	0	54	0	0	0	0	142	103	1	0	0
5	0	87	1	1	0	0	110	154	1	0	0
6	0	69	0	0	0	0	110	132	1	0	0
7	0	63	0	1	0	0	104	66	0	0	0
8	0	30	1	0	0	0	144	110	1	0	0
9	0	35	0	0	0	0	108	60	1	0	0
10	0	70	1	1	1	0	138	103	0	0	0
11	0	55	1	1	0	0	188	86	0	0	0
12	0	48	0	1	1	0	162	100	0	0	0
13	0	66	1	1	0	0	160	80	0	0	0
14	0	61	1	0	0	1	174	99	1	1	0
15	0	66	0	0	0	0	206	90	1	0	0
16	0	52	0	1	0	0	150	71	0	0	0
17	0	55	0	1	0	0	140	116	0	0	0
18	0	59	0	0	0	0	48	39	1	0	1
19	0	63	0	0	0	0	132	128	1	0	0
20	0	72	0	1	0	0	120	80	0	0	0
21	0	60	0	0	0	0	114	110	1	0	0
22	0	78	0	1	0	0	180	75	0	0	0
23	0	16	1	0	0	0	104	111	1	0	0
24	0	62	0	1	0	1	200	120	0	0	0
25	0	61	0	0	0	0	110	120	1	0	0
26	0	35	0	0	0	0	150	98	1	0	0
27	0	74	1	1	0	0	170	92	0	0	0
28	0	68	0	1	0	0	158	96	0	0	0
29	0	69	1	1	0	0	132	60	1	0	0
30	0	51	0	0	0	0	110	99	1	0	0
31	0	55	0	1	0	0	128	92	0	0	0
32	0	64	1	1	0	0	158	90	1	0	0
33	0	88	1	1	0	0	140	88	1	0	0
34	0	23	1	1	0	0	112	64	1	0	0
35	0	73	1	1	1	0	134	60	0	0	0
36	0	53	0	1	0	0	110	70	0	0	0
37	0	74	0	1	0	0	174	86	0	0	0
38	0	68	0	1	0	0	142	89	0	0	0
39	0	66	1	0	0	0	170	95	1	0	0

40	0	60	0	1	1	0	110	92	0	0	0
41	1	87	1	1	0	0	80	96	1	1	0
42	1	76	1	1	0	0	128	90	1	0	0
43	1	78	0	0	0	0	130	132	1	0	0
44	1	63	0	0	0	1	112	106	1	0	0
45	1	19	0	1	0	0	140	76	1	0	0
46	1	67	1	0	0	0	62	145	1	0	0
47	1	53	1	0	0	0	148	128	1	1	0
48	1	92	0	0	0	0	124	80	1	0	0
49	1	57	0	0	0	0	110	124	1	0	1
50	1	75	1	1	1	0	130	136	0	0	0
51	1	91	0	0	0	0	64	125	1	0	0
52	1	70	0	1	0	0	168	122	0	0	1
53	1	88	0	0	0	0	141	140	1	0	0
54	1	41	0	1	0	0	140	58	1	0	1
55	1	61	0	0	0	0	140	81	1	0	0
56	1	80	0	1	0	0	100	85	1	0	0
57	1	40	0	0	0	0	86	80	1	0	0
58	1	75	0	0	0	0	90	100	1	0	1
59	1	63	1	1	0	1	36	86	1	0	1
60	1	75	1	0	1	0	190	94	1	0	0
61	1	20	0	1	0	0	148	72	1	0	0
62	1	71	0	0	0	0	142	95	1	0	0
63	1	51	1	1	0	0	134	100	1	0	1
64	1	65	0	0	0	0	66	94	1	0	1
65	1	69	1	0	0	1	170	60	1	0	0
66	1	55	0	1	0	1	122	100	1	0	0
67	1	50	1	1	1	0	120	96	1	0	0
68	1	78	0	0	0	0	110	81	1	0	0
69	1	71	1	0	0	0	70	112	1	0	1
70	1	85	1	1	0	0	136	96	1	0	0
71	1	75	0	0	0	1	130	119	1	1	0
72	1	65	1	0	0	0	104	150	1	0	1
73	1	49	0	0	0	0	140	108	1	0	0
74	1	75	1	0	0	1	150	66	1	0	1
75	1	72	1	0	0	0	90	160	1	0	0
76	1	69	0	0	0	1	80	81	1	0	1
77	1	64	0	0	1	0	80	118	1	0	0
78	1	60	0	0	0	0	56	114	1	1	0
79	1	60	0	1	0	1	130	55	1	0	1
80	1	50	1	0	0	0	256	64	1	0	1

APPENDIX G

Walking Study Data

1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0	58.6	79.0	54.2	170.5	81.1	164	27.90	150	105	60	1	1
2	0	58.6	73.5	49.5	156.2	53.2	140	21.80	112	74	100	0	1
3	0	59.3	80.0	52.1	169.5	82.2	140	28.61	142	90	60	1	1
4	0	68.6	76.5	49.2	166.0	68.9	154	25.00	130	84	76	0	1
5	0	59.0	74.5	48.2	160.0	75.0	174	29.30	130	78	78	0	1
6	0	67.6	75.5	47.9	155.5	90.2	120	37.30	147	97	84	0	1
7	0	71.1	75.0	50.5	159.5	57.8	164	22.72	124	68	72	0	1
8	0	79.8	80.0	54.1	177.0	72.5	140	23.14	142	78	84	1	1
9	0	84.7	81.5	52.3	171.2	59.2	120	20.20	180	89	66	1	1
10	0	59.6	69.5	46.8	156.5	52.1	146	21.27	130	68	72	0	1
11	0	61.2	84.0	54.0	178.0	82.1	164	25.91	116	64	56	1	1
12	0	65.0	74.0	50.9	160.5	65.8	141	25.54	156	76	72	0	1
13	0	82.6	76.0	48.4	162.0	71.4	98	27.21	145	90	54	1	1
14	0	64.7	79.5	50.1	171.0	67.7	120	23.15	116	76	92	1	1
15	0	61.4	84.0	56.8	181.5	70.5	164	21.40	148	80	90	1	1
16	0	65.5	79.5	53.2	175.0	108.1	180	35.30	134	70	66	1	1
17	0	57.6	74.5	52.2	157.0	90.7	120	36.80	120	80	66	0	1
18	0	78.4	76.5	50.9	158.0	54.6	140	21.87	140	76	72	0	1
19	0	57.8	69.0	43.4	149.5	56.2	160	25.15	142	82	72	0	1
20	0	69.2	84.0	55.2	178.5	74.4	150	23.35	150	80	78	1	1
21	0	73.2	69.0	47.3	154.0	56.4	156	23.78	156	84	78	0	1
22	0	67.9	73.5	48.1	157.0	63.2	157	25.64	110	64	60	0	0
23	0	69.7	80.5	54.2	174.0	94.0	170	31.05	134	80	78	1	0
24	0	64.9	75.0	49.9	163.0	70.6	130	26.57	120	80	72	0	1
25	0	69.9	71.0	47.9	160.0	58.1	180	22.70	150	78	84	0	1
26	0	67.5	74.0	46.4	159.5	53.6	160	21.07	130	90	60	0	1
27	0	62.2	85.0	56.1	180.0	92.2	164	28.46	144	72	72	1	0
28	0	68.9	79.0	51.7	169.5	65.4	170	22.76	130	75	66	1	1
29	0	79.5	74.0	49.1	161.0	64.7	136	24.96	148	76	72	0	0
30	0	60.3	89.0	56.0	185.2	93.0	180	27.11	160	92	66	1	1
31	0	56.4	75.5	51.0	163.0	69.3	195	26.08	120	85	60	0	1
32	0	60.5	74.5	50.0	166.0	56.2	164	20.39	140	74	84	0	1
33	0	79.8	80.0	58.3	182.5	91.4	160	27.44	160	86	74	1	1
34	0	80.7	74.5	49.0	162.0	75.3	170	28.69	130	82	60	1	1
35	0	86.1	74.5	48.2	164.0	63.4	130	23.57	158	58	72	1	1
36	0	57.2	79.0	52.4	168.5	71.1	156	25.04	154	94	60	1	1
37	0	59.6	81.0	53.8	170.5	79.9	187	27.49	120	60	67	1	1
38	0	72.8	76.0	51.3	164.0	59.6	160	22.16	126	82	66	0	0
39	0	78.5	73.0	48.3	158.0	55.0	156	22.03	124	86	72	1	1

40	0	57.8	84.0	55.8	178.5	88.6	170	27.81	118	82	64	1	1
41	0	72.3	69.5	45.5	153.0	53.3	140	22.77	114	68	60	0	0
42	0	59.3	76.0	51.2	168.0	65.5	186	23.21	114	70	60	1	1
43	0	65.8	81.0	53.6	174.5	62.1	170	20.39	135	80	72	1	1
44	0	65.0	77.4	51.8	169.0	70.3	170	24.61	142	92	66	1	1
45	0	57.5	74.5	51.5	163.2	69.8	120	26.21	120	80	72	0	1
46	0	84.7	69.0	47.6	148.5	68.5	132	31.06	154	98	72	0	1
47	0	74.9	69.5	47.6	157.0	55.3	156	22.43	170	90	90	0	1
48	0	73.5	75.5	51.8	164.0	57.4	138	21.34	160	84	84	0	1
49	0	72.7	71.0	46.7	147.5	56.8	157	26.11	143	82	78	0	1
50	0	67.0	71.0	49.9	159.5	58.6	140	23.03	152	68	78	0	1
51	0	75.8	82.2	54.0	172.0	61.9	157	20.92	110	70	78	1	1
52	0	64.3	76.5	50.8	167.0	59.2	195	21.23	165	76	84	0	0
53	0	73.4	82.0	55.6	174.0	76.6	187	25.30	140	78	84	1	1
54	0	65.4	71.5	47.3	151.0	68.6	164	30.09	122	80	65	0	0
55	0	62.5	74.0	48.6	162.0	82.6	130	31.47	150	84	72	0	1
56	0	68.8	77.5	48.6	167.0	76.7	150	27.50	158	92	48	1	1
57	0	78.5	76.0	52.5	167.0	76.2	110	27.32	128	92	66	1	1
58	0	71.1	85.0	53.7	181.5	94.8	164	28.78	130	80	60	1	1
59	0	81.8	77.5	57.2	173.5	73.4	164	24.38	140	86	84	1	1
60	0	77.0	75.5	50.6	161.0	49.8	156	19.21	160	78	72	0	1
61	0	58.1	79.7	54.1	170.5	95.4	164	32.82	124	76	64	0	1
62	0	67.6	71.5	48.8	158.0	62.2	172	24.92	120	76	60	0	1
63	0	64.0	64.5	41.7	138.0	40.7	157	21.37	130	90	74	0	1
64	0	66.8	80.5	51.6	170.0	61.6	174	21.31	150	72	84	1	1
65	0	72.8	71.2	47.2	153.5	69.4	164	29.45	130	64	84	0	1
66	0	82.2	74.0	49.1	165.5	65.1	154	23.77	140	68	72	1	1
67	0	75.7	74.0	53.9	167.5	85.3	160	30.40	150	78	72	0	1
68	0	58.5	86.5	57.4	184.0	91.6	142	27.06	130	78	78	1	1
69	0	81.4	69.5	48.1	156.0	73.8	130	30.33	140	86	72	0	1
70	0	75.2	79.8	54.6	173.0	83.5	136	27.90	146	84	64	1	1
71	0	82.7	79.5	51.1	162.5	81.4	157	30.83	140	72	72	1	1
72	0	79.1	62.0	43.4	143.0	36.9	158	18.04	132	72	90	0	1
73	0	74.1	80.5	54.3	170.5	80.9	154	27.83	142	86	52	0	1
74	0	67.9	75.0	48.2	161.5	81.5	108	31.25	140	72	68	0	1
75	0	73.7	81.7	56.7	174.5	77.5	150	25.45	152	76	72	1	1
76	0	82.6	77.0	51.7	167.0	66.4	117	23.81	148	86	84	1	1
77	0	77.0	72.8	46.8	157.5	51.6	164	20.80	148	76	72	0	1
78	0	84.5	80.5	50.9	162.0	64.7	134	24.65	140	98	76	1	1
79	0	83.1	66.5	47.8	148.0	58.1	120	26.52	138	68	66	0	1
80	0	70.6	72.0	48.5	156.5	76.6	154	31.28	154	76	64	0	1
81	0	60.8	70.0	46.2	153.0	76.0	146	32.47	148	86	76	0	1
82	0	63.4	72.5	45.7	155.5	52.4	156	21.67	120	70	60	1	1
83	0	83.1	78.0	52.1	168.5	75.3	120	26.52	170	76	64	1	1
84	0	80.1	80.7	53.4	170.0	84.7	148	29.31	150	80	60	0	1

85	0	77.1	74.0	49.1	157.0	72.2	136	29.29	152	100	72	0	1
86	0	66.6	76.5	49.7	163.2	70.1	169	26.32	146	90	72	1	1
87	0	82.4	81.5	54.1	174.0	82.5	128	27.25	170	74	72	1	1
88	0	82.3	69.2	43.5	145.0	49.4	128	23.50	124	80	82	0	1
89	0	56.8	71.5	46.8	159.0	57.4	178	22.70	164	88	72	0	1
90	0	60.1	82.0	54.3	177.0	77.5	170	24.74	140	82	72	1	1
91	0	71.8	73.5	49.3	162.5	71.3	134	27.00	140	70	66	0	1
92	0	85.9	71.0	46.5	149.5	64.0	146	28.64	120	84	90	0	1
93	0	74.3	79.0	54.1	169.0	86.3	124	30.22	140	90	90	1	1
94	0	82.2	70.2	46.3	150.5	72.2	138	31.88	160	74	72	0	1
95	0	74.0	80.0	51.2	164.2	77.5	110	28.74	142	82	72	0	1
96	0	76.4	79.5	50.7	165.5	71.0	164	25.92	154	80	72	1	1
97	0	58.9	73.5	49.9	157.5	89.8	105	36.20	110	70	72	0	1
98	0	64.0	72.5	50.3	155.5	56.4	152	23.32	150	88	76	0	1
99	0	84.8	66.5	46.1	143.0	39.0	124	19.07	168	88	72	0	1
100	0	72.5	71.2	49.1	155.0	55.2	120	22.98	114	62	60	0	1
101	1	71.9	72.0	47.4	156.5	65.0	164	26.54	164	86	56	0	1
102	1	67.3	75.5	48.8	159.0	52.5	164	20.77	135	80	76	0	0
103	1	69.6	78.5	49.1	165.5	50.0	170	18.25	165	75	67	0	0
104	1	56.7	74.0	49.5	161.5	70.5	154	27.03	127	87	65	0	1
105	1	60.0	80.0	53.7	174.5	95.4	126	31.33	162	92	62	1	1
106	1	57.2	80.0	51.6	176.5	79.2	160	25.42	153	90	70	1	1
107	1	56.4	79.0	51.7	171.0	86.7	110	29.65	140	85	70	1	1
108	1	55.9	81.5	51.3	170.2	84.2	164	29.07	154	84	63	1	1
109	1	69.9	81.0	51.5	166.0	64.6	152	23.44	152	64	78	1	1
110	1	56.4	71.5	45.8	159.0	66.6	180	26.34	105	75	58	0	0
111	1	68.6	79.0	51.1	161.0	75.7	160	29.20	140	80	72	0	1
112	1	67.4	75.4	49.0	165.3	69.5	178	25.44	155	85	80	1	1
113	1	69.8	84.0	55.1	182.0	76.5	174	23.10	150	85	60	1	1
114	1	66.6	79.5	48.7	166.5	79.1	170	28.53	130	80	80	1	1
115	1	76.4	82.5	50.3	168.5	59.7	170	21.03	150	90	80	1	1
116	1	77.2	72.5	49.7	158.5	76.0	160	30.25	160	82	66	0	0
117	1	66.5	74.5	48.1	155.5	65.6	146	27.13	172	78	66	0	1
118	1	64.7	72.5	46.3	154.0	65.7	124	27.70	148	88	72	0	1
119	1	64.1	75.0	46.1	159.0	61.1	164	24.17	108	68	72	0	0
120	1	76.7	80.5	52.4	164.0	69.6	174	25.88	148	90	76	1	1
121	1	76.9	67.0	46.1	148.5	55.5	128	25.17	110	62	64	0	1
122	1	72.3	72.5	47.6	159.0	57.8	184	22.86	138	74	86	0	1
123	1	76.1	80.0	53.5	173.5	75.6	124	25.11	142	64	78	1	1
124	1	65.5	76.0	49.4	167.0	74.1	164	26.57	130	60	86	0	1
125	1	64.0	76.6	49.0	165.5	69.6	164	25.41	160	80	84	0	1
126	1	62.3	74.5	49.4	166.0	78.3	150	28.41	158	90	78	0	1
127	1	67.4	78.5	50.2	168.5	74.5	110	26.24	128	80	60	1	1
128	1	65.4	76.0	50.3	166.0	68.1	154	24.71	156	66	66	0	1
129	1	63.1	79.0	53.5	170.5	72.8	164	25.04	138	76	72	1	1

130	1	61.8	82.5	53.4	178.0	91.0	170	28.72	132	86	72	1	1
131	1	62.4	84.0	55.7	184.0	82.5	161	24.37	115	78	72	1	1
132	1	62.4	76.5	51.7	174.5	63.8	160	20.95	110	72	72	1	1
133	1	61.8	87.0	55.0	178.5	80.1	160	25.14	120	68	78	1	1
134	1	63.1	72.5	48.9	157.5	65.4	200	26.36	160	90	78	0	1
135	1	63.0	84.5	49.5	172.0	88.8	157	30.02	132	84	72	1	0
136	1	62.8	77.1	50.4	166.0	74.0	146	26.85	176	78	72	1	1
137	1	81.4	68.0	45.4	151.5	60.6	174	26.40	150	73	78	0	1
138	1	61.1	80.5	53.5	170.0	91.8	180	31.76	132	82	56	1	1
139	1	61.7	69.0	48.7	154.5	76.0	160	31.84	132	82	78	0	1
140	1	67.8	67.6	46.7	150.0	65.8	164	29.24	168	90	78	0	0
141	1	81.9	76.0	49.0	166.5	67.3	125	24.28	112	82	84	1	1
142	1	63.6	81.0	52.3	172.0	66.3	150	22.41	108	64	60	1	1
143	1	73.2	77.0	49.0	162.2	65.3	185	24.82	140	86	66	0	1
144	1	78.8	72.0	49.7	161.5	65.6	160	25.15	132	78	60	1	1
145	1	66.9	76.5	51.8	165.0	62.2	130	22.85	118	64	66	0	1
146	1	57.1	78.0	50.8	167.5	48.3	160	17.22	120	74	66	0	1
147	1	80.1	71.0	49.1	163.0	56.4	90	21.23	120	60	66	0	1
148	1	66.7	73.5	47.2	155.5	64.0	170	26.47	132	80	86	0	1
149	1	68.3	75.5	51.3	162.5	81.1	130	30.71	136	78	70	0	1
150	1	68.0	82.5	52.6	172.0	78.4	110	26.50	158	86	50	1	1
151	1	74.5	73.0	47.2	156.0	57.0	146	23.42	148	74	72	0	0
152	1	60.1	78.5	49.6	165.5	73.0	140	26.65	136	88	50	1	1
153	1	72.9	73.0	48.2	159.0	52.9	160	20.92	142	80	84	0	1
154	1	56.6	77.5	49.8	165.0	77.9	170	28.61	122	82	72	1	1
155	1	57.1	68.5	42.9	144.0	52.8	170	25.46	110	70	66	0	1
156	1	68.2	81.5	51.0	172.5	69.0	170	23.19	142	72	84	1	1
157	1	56.1	73.5	50.1	164.2	69.8	178	25.87	155	80	60	0	1
158	1	58.9	73.0	49.2	163.0	63.0	178	23.71	140	82	68	0	1
159	1	60.3	81.0	51.9	173.5	85.8	185	28.50	114	80	76	1	1
160	1	56.7	81.5	56.2	172.5	80.4	150	27.02	140	84	66	1	1
161	1	61.7	77.5	54.4	173.0	81.2	170	27.13	110	74	72	1	1
162	1	67.5	81.0	51.5	168.0	72.4	164	25.65	120	80	60	1	1
163	1	59.8	82.0	51.8	169.0	74.6	178	26.12	140	76	54	1	1
164	1	62.2	79.5	49.0	164.0	57.3	170	21.30	130	74	72	0	1
165	1	77.8	82.5	54.1	179.5	69.8	120	21.66	190	92	60	1	1
166	1	64.8	82.5	55.5	178.0	73.2	170	23.10	115	72	66	1	1
167	1	58.3	85.0	54.6	180.0	72.7	146	22.44	132	72	60	1	1
168	1	68.1	77.0	55.2	171.5	70.5	125	23.97	114	68	72	1	0
169	1	64.0	72.0	49.2	162.0	74.8	160	28.50	130	70	66	0	1
170	1	56.9	71.0	47.8	158.0	64.6	180	25.88	110	74	60	0	1
171	1	71.9	84.5	55.4	177.0	84.2	160	26.88	162	84	66	1	1
172	1	74.9	80.5	54.4	177.0	87.1	150	27.80	160	90	60	1	1
173	1	76.2	78.5	54.3	173.5	71.2	146	23.65	130	68	66	1	1
174	1	80.4	75.5	47.7	158.5	62.3	102	24.80	154	74	66	0	0

175	1	72.5	81.5	54.3	175.0	89.8	164	29.32	150	84	60	1	1
176	1	74.4	83.5	51.4	173.0	72.9	170	24.36	128	84	54	1	1
177	1	63.4	85.5	54.8	178.5	78.2	150	24.54	140	84	60	1	1
178	1	72.7	72.5	49.3	158.0	57.0	170	22.83	158	74	78	0	1
179	1	70.8	82.0	53.6	174.0	80.9	195	26.72	150	90	88	1	1
180	1	58.9	69.0	48.0	155.2	69.3	170	28.77	114	76	60	0	0
181	1	73.5	79.0	51.8	171.5	77.3	120	26.28	140	80	60	1	1
182	1	74.9	77.5	52.9	165.5	74.0	160	27.02	140	75	84	1	1
183	1	57.6	85.2	57.1	184.0	88.6	196	26.17	120	88	72	1	1
184	1	58.8	78.0	53.2	169.5	93.3	154	32.47	120	72	60	1	1
185	1	57.3	74.5	47.4	157.5	61.2	170	24.67	98	72	72	0	1
186	1	58.7	67.7	46.0	148.5	67.4	136	30.56	152	86	60	0	1
187	1	67.7	73.5	48.3	157.2	59.1	110	23.92	162	84	66	0	0
188	1	62.8	80.7	53.4	174.5	86.0	150	28.24	140	80	66	1	1
189	1	64.4	73.0	46.8	151.0	61.9	170	27.15	142	76	72	0	1
190	1	69.7	74.5	49.4	162.0	68.8	176	26.22	150	70	72	0	1
191	1	77.1	84.0	56.6	172.2	76.7	120	25.87	146	82	54	1	1
192	1	72.6	73.0	48.9	157.2	63.6	156	25.74	142	78	72	0	1
193	1	69.6	72.0	52.8	164.5	65.4	170	24.17	136	60	78	0	1
194	1	83.0	75.0	49.1	152.0	62.5	130	27.05	142	86	66	0	1
195	1	71.9	74.5	51.1	165.0	70.9	130	26.04	140	76	78	1	1
196	1	62.3	71.5	47.8	154.0	67.5	170	28.46	140	82	60	0	1
197	1	69.1	73.5	51.4	166.0	58.1	130	21.08	148	74	60	0	1
198	1	63.0	74.0	51.2	167.0	64.3	170	23.06	130	78	78	0	1
199	1	65.0	69.5	47.0	152.5	55.2	164	23.74	135	84	70	0	1
200	1	67.7	83.4	54.3	172.0	74.0	140	25.01	152	82	72	1	1

REFERENCES

Afifi, A. A., Sack, S. T., Liu, V. Y., Weil, M. H., and Shubin, H. (1971). Accumulative prognostic index for patients with barbiturate, glutethimide and meprobamate intoxication. *The New England Journal of Medicine* **285**, 1497-1502.

Akaike, H. (1973). Information theory and an extension of the maximum lokelihood principle. 2nd *International Symposium on Information Theory*, 267-281. Budapest: Akademia Kiado.

Allen, D. M. (1971a). The prediction sum of squares as a criterion for selecting prediction variables. *Technical Report No.***23**, Department of Statistics, University of Kentucky, U.S.A..

Allen, D. M. (1971b). Mean square error of prediction as a criterion for selecting variables. *Technometrics* **13**, 469-475.

Anderson, J. A. (1972). Separate sample logistic discrimination. *Biometrika* **59**, 19-35.

Anscombe, F. J. (1967). Topics in the investigation of linear relations fitted by the method of least squares. *Journal of the Royal Statistical Society, Series B* **29**, 1-52.

Ashutosh, K., Lee, H., Mohan, C. K., Ranka, S., Mehrotra, K., and Alexander, C. (1992). Prediction criteria for successful weaning from respiratory support: Statistical and connectionist analysis. *Critical Care Medicine* **20**, 1295-1301.

Atkinson, A. C. (1980). A note on the generalized information criterion for choice of a model. *Biometrika* **67**, 413-418.

Bahadur, R. R. (1961). A representation of the joint distribution of responses to n dichotomous items. In *Studies in Item Analysis and Prediction*, H. Solomon, Ed., California: Stanford University Press.

Beaton, A. E. (1964). The use of special matrix operators in statistical calculus. *Educational Testing Service*, Princeton, New Jersey, RB-64-51.

Beaton, A. E. and Tukey, J. W. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics* **16**, 147-185.

Bendel, R. B. and Afifi, A. A. (1977). Comparison of stopping rules in forward regression. *Journal of the American Statistical Association* **72**, 46-53.

Birch, M. W. (1963). Maximum likelihood in three-way contingency tables. *Journal of the Royal Statistical Society, Series B* **25**, 220-233.

Bhansali, R. J. and Downham, D. Y. (1977). Some properties of the order of an autoregressive model selected by a generalization of Akaike's FPE criterion. *Biometrika* **64**, 547-551.

Box, G. E. P. and Behnken, D. W. (1960). Some new three level designs for the study of quantitative variables. *Technometrics* **2**, 455-475.

Box, G. E. P. and Draper, N. R. (1987). *Empirical Model-Building and Response Surfaces*. New York: Wiley.

Chow, G. C. (1981). A comparison of the information and posterior probability criteria for model selection. *Journal of Econometrics* **16**, 21-33.

Cochran, W. G. and Cox, G. M. (1957). *Experimental Designs*. 2nd edition. New York: Wiley.

Cornfield, J. (1951). A method of estimating comparative rates from clinical data; Applications to cancer of the lung, breast and cervix. *Journal of the National Cancer Institute* **11**, 1269-1275.

Cornfield, J., Gordon, T., and Smith, W. W. (1961). Quantal response curves for experimentally uncontrolled variables. *Bulletin of the International Statistical Institute* **38**, 97-115.

Cornfield, J. (1962). Joint dependence of risk of coronary heart disease on serum cholesterol and systolic blood pressure: a discriminant function analysis. *Federation Proceedings* **21**, 58-61.

Costanza, M. C. and Afifi, A. A. (1979). Comparison of stopping rules in forward stepwise discriminant analysis. *Journal of the American Statistical Association* **74**, 777-785.

Costanza, M. C. and Ashikaga, T. (1986). Monte Carlo study of forward stepwise discrimination based on small samples. *Computers and Mathematics with Applications* **12A**, 245-252.

Cox, D. R. and Snell, E. J. (1989). *Analysis of Binary Data*. 2nd edition. Chapman and Hall, New York.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. Chapman and Hall, New York.

Davis, O. L. et al (1956). *The Design and Analysis of Industrial Experiments*. Hofner, New York.

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

Dixon, W. J. (1988). BMDP *Statistical Software*. University of California Press, Berkeley.

Dobson, A. J. (1990). *Introduction to Generalized Linear Models*. Chapman and Hall, New York.

Draper, N. R. and Smith, H. (1981). *Applied Regression Analysis*, 2nd edition. New York: Wiley.

Efron, B. (1975). The efficiency of logistic regression compared to normal discriminant function analysis. *Journal of the American Statistical Association* **70**, 892-898.

Efron, B. (1986). How biased is the apparent error rate of a prediction rule?. *Journal of the American Statistical Association* **81**, 461-470.

Efroymson, M. A. (1960). Multiple regression analysis. *Mathematical Methods for Digital Computers*. Ralston, A. and Wilf, H. S. (Eds.), New York: Wiley.

Farewell, V. T. (1979). Some results on the estimation of logistic model based on retrospective data. *Biometrika* **66**, 27-32.

Ferraris, V. A. and Propp, M. E. (1992). Outcome in critical care patients: A multivariate study. *Critical Care Medicine* **20**, 967-976.

Finney, D. J. (1971). *Probit Analysis*, 3rd edition. Cambridge University Press, London.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics* **7**, 179-188.

Flack, V. F. and Chang, P. C. (1987). Frequency of selecting noise variables in subset regression analysis: A simulation study. *American Statistician* 41, 84-86.

Forsythe, A. B., Engelman, L., and Jennrich, R. (1973). Stopping rule for variable selection in multiple regression. *Journal of the American Statistical Association* **68**, 75-77.

Freeman, M. F., and Tukey, J. W. (1950). Transformation related to the angular and the square root. *Annals of Mathematical Statistics* **21**, 607-611.

Furnival, G. M. (1971). All possible regression with less computation. *Technometrics* **16**, 499-512.

Furnival, G. M., and Wilson, R. W. (1974). Regression by leaps and bounds. *Technometrics* **16**, 499-511.

Gardlund, B. (1986). Prognostic evaluation in meningococcal disease: A retrospective study of 115 cases. *Intensive Care Medicine* **12**, 302-307.

Garside, M. J. (1965). The best subset in multiple regression analysis. *Journal of the Royal Statistical Society, Series C* **14**, 196-200.

Glick, N. (1972). Sampled-based classification procedures derived from density estimators. *Journal of the American Statistical Association* **67**, 116-122.

Glick, N. (1973). Sampled-based multinomial classification. *Biometrics* **29**, 241-256.

Gong, G. (1986). Cross-validation, the jackknife, and the bootstrap: Excess error estimation in forward logistic regression. *Journal of the American Statistical Association* **81**, 108-113.

Goodnight, J. H. (1979). A tutorial on the SWEEP operator. *The American Statistician* **33**, 149-158.

Greenland, S. (1989). Modeling variable selection in epidemiologic analysis. *American Journal of Public Health* **79**, 340-349.

Griffiths, W. E. and Pope, P. J. (1987). Small sample properties of probit models. *Journal of the American Statistical Association* **82**, 929-937.

Grizzle, J. E. (1971). Multivariate logit analysis. *Biometrics* **27**, 1057-1062.

Habbema, J. D. F. and Hermans, J. (1977). Selection of variables in discriminant analysis by F-statistic and error rate. *Technometrics* **19**, 487-493.

Halinski, R. S. and Feldt, L. S. (1970). The selection of variables in multiple regression analysis. *Journal of Educational Measurement* **7**, 151-157.

Halpern, M., Blackwelder, W. C., and Verter, J. I. (1971). Estimation of the multivariate logistic risk function: A comparison of the discriminant function and maximum likelihood approaches. *Journal of Chronic Disease* **24**, 125-158.

Hans, P., Stevenaert, A., and Albert, A. (1986). Study of hypotonic polyurica after trans-sphenoidal pituitary adenomectomy. *Intensive Care Medicine* **12**, 95-99.

Hauck, W. W., and Donner, A. (1977). Wald's test as applied to hypotheses in logit analysis. *Journal of the American Statistical Association* **72**, 851-853.

Heiberger, D. J. (1978). Generation of random orthogonal matrices. Algorithm AS 127. pplied Statistics **27**, 199-206.

Hills, M. (1966). Allocation rules and their error rates. *Journal of the Royal Statistical Society, Series B* **28**, 1.

Hocking, R. R. (1972). Criterion for selection of a subset regression: Which one should be used. *Technometrics* **9**, 531-540.

Hocking, R. R. (1976). The analysis and selection of variables in linear regression. *Biometrics* **32**, 1-49.

Horbar, J. D., Onstad, L., and Wright, E. (1993). Predicting mortality risk for infants weighing 501 to 1500 grams at birth: A National Institutues of Health Neonatal Research Network Report. *Critical Care Medicine* **21**, 12-18.

Hosmer, D. W., Wang, C. Y., Lin, I. C., and Lemeshow, S. (1978). A computer program for stepwise logistic regression using maximum likelihood. *Computer Programs in Biomedicine* **8**, 121-134.

Hosmer, T., Hosmer, D. W., and Fisher, L. L. (1983). A comparison of the maximum likelihood and discriminant function estimators of the coefficients of the logistic regression model for mixed continuous and discrete variables. *Communications in Statistics* **B12**, 577-593.

Hosmer, D. W., Jovanovic, B., and Stanley, L. (1989a). Best subsets logistic regression. *Biometrics* **45**, 1265-1270.

Hosmer, D. W. and Lemeshow, S. (1989b). *Applied Logistic Regression.* New York: Wiley.

Huberty, C. J. (1975). Discriminant analysis. *Review of Educational Research* **45**, 543-593.

Kendall, M. G. and Stuart, A. (1968). *The Advanced Theory of Statistics.* Vol. 3, 2nd ed, New York: Hafner.

Kennedy, W. J. and Bancroft, T. A. (1971). Model building for prediction in regression based upon repeated significance tests. *Annals of Mathematical Statistics 42*, 1273-1284.

Kitagawa, G. (1979). On the use of AIC for the detection of outliers. *Technometrics* **21**, 193-199.

Krall, J. M., Uthoff, V. A. and Harley, J. B. (1975). A step-up procedure for selecting variables associated with survival. *Biometrics* **31**, 49-57.

Lachenbruch, P. A. (1975). *Discriminant Analysis.*, New York: Hafner.

Lawless, J. F. and Singhal, K. (1978). Efficient screening of nonnormal regression models. *Biometrics* **34**, 318-327.

Lawless, J. F. and Singhal, K. (1987a). ISMOD: An all-subsets regression program for generalized linear models. I. Statistical and computational background. *Computer Methods and Programs in Biomedicine* **24**, 117-124.

Lawless, J. F. and Singhal, K. (1987b). ISMOD: An all-subsets regression program for generalized linear models. II. Program guide and examples. *Computer Methods and Programs in Biomedicine* **24**, 125-134.

Leamer, E.E. (1979). Information criteria for choice of regression models: A comment. *Econometrica* **47**, 507-510.

Lemeshow, S., Teres, D., Pastides, H., Avrunin, J. S., and Steingrub, J. S. (1985). A method for predicting survival and mortality of ICU patients using objectively derived weights. *Critical Care Medicine* **13**, 519-526.

Lemeshow, S., Teres, D., Avrunin, J. S., and Pastides, H. (1987). A comparison of methods to predict mortality of intensive care unit patients. *Critical Care Medicine* **15**, 715-722.

Lindley, D. V. (1968). The choice of variables in multiple regression. *Journal of the Royal Statistical Society, Series B* **30**, 31-53.

Mallows, C. L. (1967). Choosing a subset regression. *Bell Telephone Laboratories* ,unpublished report.

Mallows, C. L. (1973). Some comments on C_p. *Technometrics* **15**, 661-675.

McCabe, G. P. (1975). Computations for variable selection in multiple discriminant analysis. *Technometrics* **17**, 103-109.

McCullagh, P. and Nelder, J. A. (1983). *Generalized Linear Models.* 2nd edition. New York: Chapman and Hall.

McLachlan, G. J. (1976). A criterion for selecting variables for the linear discriminant function. *Biometrics* **32**, 529-534.

Miller, A. J. (1984). Selection of subsets of regression variables. *Journal of the Royal Statistical Society, Series A* **147**, 389-425.

Morgan, J. A. and Tatar, J. R. (1973). Calculation of the residual sum of squares for all possible regressions. *Technometrics* **15**, 317-326.

Olkin, I. and Pratt, J. W. (1958). Unbiased estimation of certain correlation coefficients. *Annals of Mathematical Statistics* **29**, 201-211.

Peduzzi, P. N., Hardy, R. J., and Holford, T. R. (1980). A stepwise variable selection procedure for nonlinear regression models. *Biometrics* **36**, 511-516

Pozen, M. W., D'Agostino, R. B., Selker, H. P., Sytkowski, P. A., and Hood, W. B. (1984). A predictive instrument to improve coronary-care-unit admission practices in acute ischemic heart disease: A prospective multicenter clinical trial. *The New England Journal of Medicine* **310**, 1273-1278.

Pregibon, D. (1981). Logistic regression diagnostics. *Annals of Statistics* **9**, 705-724.

Prentice, R. L., and Pyke, R. (1979). Logistic disease incidence models and case-control studies. *Biometrika* **66**, 403-411.

Press, S. J. and Wilson S. (1978). Choosing between logistic regression and discriminant analysis. *Journal of the American Statistical Association* **73**, 699-705.

Ralston, A. (1960). *Mathematical Methods for Digital Computers*. New York: Wiley.

Rao, C. R. (1947). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proceedings of the Cambridge Philosophical Society* **44**, 50-57.

Rao, C. R. (1973). *Linear Statistical Inference and its Application* , 2nd edition. New York: Wiley.

Rock, D. A., Linn, R. L., Evans, F. R., and Patrick, C. (1970). A comparison of predictor selection techniques using Monte Carlo methods. *Educational and Psychological Measurement* **30**, 873-884.

Rosenberg, S. H., and Levy, P. S. (1972). A characterization on misspecification in the general linear regression model. *Biometrics* **28**, 1129-1132.

Rothman, D. (1968). Letter to the editor. *Technometrics* **10**, 432.

SAS/IML Software. (1990). *Usage and Reference*, Version 6, First Edition, SAS Institute Inc., Cary, NC, USA.

SAS/STAT. (1989). *User's Guide*, version 6, 4th edition, volume 2: SAS Institute Inc., Cary, NC, USA.

Sawa, T. (1978). Information criteria for discriminating among alternative regression models. *Econometrica* **46**, 1273-1292.

Schaafsma, W. and van Vark, G. N. (1979). Classification and discrimination problems with applications, part IIa. *Statistica Neerlandica* **33**, 91-126.

Schaefer, J. H., Jochimsen, F., Keller, F., Wegscheider, K., and Distler, A. (1991). Outcome prediction of acute renal failure in medical intensive care. *Intensive Care Medicine* **17**, 19-24.

Schatzoff, M., Tsao, R., and Feinberg, S. (1968). Efficient calculation of all possible regression. *Technometrics* **10**, 769-780.

Schmidt, P. (1973). Methods of choosing among alternative linear regression models. University of North Carolina, Chapel Hill, North Carolina, U.S.A..

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* **6**, 461-464.

Seigel, D. G., and Greenhouse, S. W. (1973). Multiple relative risk functions in case-control studies. *American Journal of Epidemiology* **97**, 324-331.

Shibata, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika* **63**, 117-126.

Snyder, J. V., McGuirk, M., Grenvik, A., and Stickler, D. (1981). Outcome of intensive care: An application of a predictive model. *Critical Care Medicine* **9**, 598-603.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society, Series B* **30**, 111-133.

Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *Journal of the Royal Statistical Society, Series B* **39**, 44-47.

Stone, M. (1979). Comments on model selection criteria of Akaike and Schwarz. *Journal of the Royal Statistical Society, Series B* **41**, 276-278.

Summerfield, A. and Lubin, A. (1951). A square root method of selecting a minimum set of variables in multiple regression: I. The Method. *Psychometrika* **16**, 271-284.

Tierney, W. M., Roth, B. J., Psaty, B., McHenry, R., Fitzgerald, J., Stump, D. L., Anderson, F. K., Ryder, K. W., McDonald, C. J., and Smith, D. M. (1985). Predictors of myocardial infarction in emergency room patients. *Critical Care Medicine* **13**, 526-531.

Truett, J., Cornfield, J. and Kannel, W. (1967). A multivariate analysis of the risk of coronary heart disease in Framingham. *Journal of Chronic Disease* **20**, 511-524.

Tukey, J. W. (1967). Discussion (of Anscombe (1967)). *Journal of the Royal Statistical Society, Series B* **29**, 47-48.

Viviand, X., Gouvernet, J., Granthil, C., and Francois, G. (1991). Simplication of the SAPS by selecting independent variables. *Intensive Care Medicine* **17**, 164-168.

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society* **54**, 426-482.

Walls, R. E. and Weeks, D. L. (1969). A note on the variance of a predicted response in regression. *American Statistician* **23**, 24-26.

Weiner, J. and Dunn, O. J. (1966). Elimination of variates in linear discrimination problems. *Biometrics* **22**, 268.

Welch, B. L. (1939). Note on discriminant functions. *Biometrika* **31**, 218-220.

Winer, B. J. (1971). *Statistical Principles in Experimental Design*, 2nd edition. New York: McGraw-Hill Book Co..